Springer

Handbook of Optical Networks

Mukherjee
Tomkos
Tornatore
Winzer
Zhao
Editors

Springer

Springer Handbook of Optical Networks

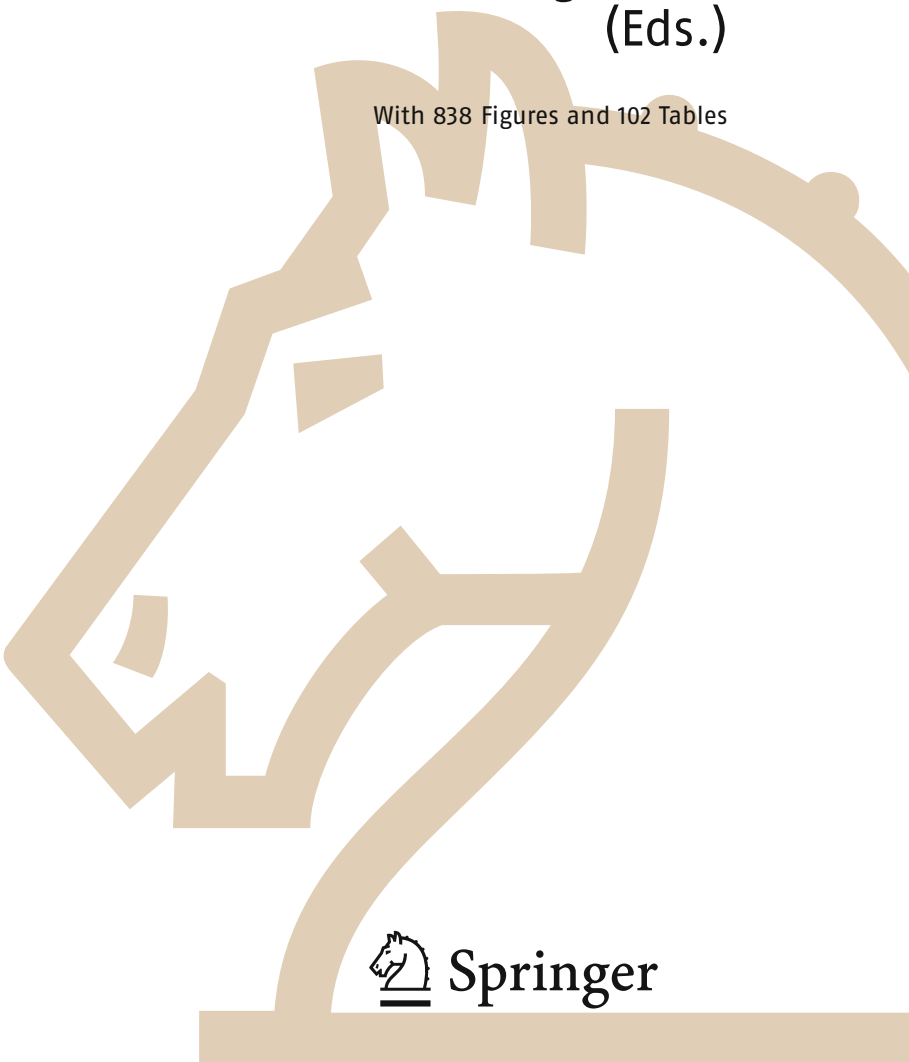
Springer Handbooks provide a concise compilation of approved key information on methods of research, general principles, and functional relationships in physical and applied sciences. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of scientific and technical information.

The volumes are designed to be useful as readable desk book to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

Springer Handbook of Optical Networks

Biswanath Mukherjee, Ioannis Tomkos,
Massimo Tornatore, Peter Winzer, Yongli Zhao
(Eds.)

With 838 Figures and 102 Tables



Springer

Editors
Biswanath Mukherjee
Dept. of Computer Science
University of California
Davis, CA, USA

Ioannis Tomkos
ECE Department
University of Patras
Patras, Greece

Massimo Tornatore
Politecnico di Milano
Milano, Italy

Peter Winzer
Nokia Bell Labs
Holmdel, NJ, USA

Yongli Zhao
Beijing University of Posts and Telecommunications
Beijing, China

ISBN 978-3-030-16249-8 e-ISBN 978-3-030-16250-4
https://doi.org/10.1007/978-3-030-16250-4

Foreword

Optical networks have moved beyond traditional applications in telecommunications to become the infrastructure of choice whenever large amounts of information need to be transmitted. The broad spectrum inherent in the use of light and the unparalleled capability to spatially pack many parallel paths into fiber cables have led to an ever-broadening range of applications. Thousands of fibers interconnect buildings full of servers in massive datacenters that underlie internet services such as search and social media. Cables with hundreds of fibers form the backbone of metropolitan access networks and interconnect wireless base stations. At the same time, traditional long-haul services—both undersea and terrestrial—demand ever more capacity. Recent years have also seen a revolution in the technology of optical networking with the advent of coherent detection, which is now ubiquitous in long-haul and metro networks and advancing into more applications as costs decline.

This new handbook on optical networks provides a broad perspective on the field, offering a survey of the fundamental technologies of optical networks: fiber, devices, and subsystems through to systems based on those technologies, and finally to the architecture and applications of the networks themselves. The editors—Peter Winzer on subsystems and technologies, Biswanath Mukherjee on core networks, Ioannis Tomkos on datacenter and supercomputer networks, and Massimo Tornatore on access and wireless networks—have assembled an impressive list of chapter authors who have made important contributions to their fields. This volume provides a valuable look at today's optical networks.

Robert W. Tkach
Director of Advanced Photonics Research
Nokia Bell Laboratories

September 2020 Robert W. Tkach

Preface

Optical communication systems form the backbone of today's communication and information society. Several billion kilometers of optical fiber are installed around the globe today—enough to wrap a string of glass as thin as a human hair around the globe more than 100 000 times. A cutting-edge optical communication system can transmit tens of terabits per second over trans-Pacific distances through a single strand of optical fiber, taking a mere 50 ms to link North America with South East Asia. Today's globally installed base of optical communication transponders is collectively capable of transmitting more than an exabit (an exabit is 1 000 petabits, 1 000 000 terabits, 1 000 000 000 gigabits, or 1 000 000 000 000 megabits) of information per second over short links (between tens of meters and a kilometer long) within a data center, tens of kilometers in mobile backhaul or fiber-to-the-home applications, hundreds of kilometers in metropolitan and regional networks, thousands of kilometers in transcontinental and submarine backbones, and even tens of thousands of kilometers or more in spaceborne satellite systems using free-space laser communications. In short, almost every bit of information we touch or consume today, whether it belongs to an Internet search, to a streamed video, or to a cellphone call, lives part of its life as an infrared photon within a gigantic global optical communications infrastructure.

It is the role of this Handbook to comprehensively describe and review the many underlying technologies that enable today's global optical communications infrastructure, as well as to explain current research trends that target continued capacity scaling and enhanced networking flexibility in support of unabated traffic growth fueled by ever-emerging new applications. Each chapter, written by world-renowned experts in its subject area, tries to paint a complete picture of that subject, from entry-level information to a snapshot of the respective state-of-the-art technologies and emerging research trends, in an effort to provide something useful for every reader—ranging from the novice who wants to get familiar with the field to the expert who wants a concise perspective on future trends.

Part A of this Handbook considers *optical subsystems for transmission and switching*, with chapters focusing on topics ranging from optical fibers and cables to optical amplifiers and switches, optical transponders and their various subsystems, as well as fiber-optic communications systems, their scalability limitations, and ways to overcome these limitations in future system designs.

Part B of this Handbook reviews *core networks*, with chapters devoted to managing the vast fiber-optic communication infrastructure at the network-wide level. Topics range from the standards required to sustain an economically viable supplier ecosystem to algorithms used to route traffic and assign infrastructure resources within optical networks, cross-layer design, and network virtualization.

Part C of this Handbook is concerned with *datacenter and supercomputer networking*, which has design requirements and solutions that differ in several ways from those of other segments of the network. Topics include reviews of industry trends and requirements as well as transponder and switching considerations specific to those applications.

Part D of this Handbook addresses *optical access and wireless networks*, and is mostly geared towards solving the last-mile problem: connecting backbone networks to end users. This may be achieved directly via fiber or visible-light free-space communications, or indirectly over a mobile wireless radio infrastructure that is heavily supported by an associated fiber-optic network. Emerging areas such as spaceborne laser communications and optical communications in avionics and autonomous vehicles round off this part of the Handbook.

The Editors gratefully acknowledge all the valuable contributions from authors and peer reviewers who took much time out of their busy schedules to write or review chapters of this Handbook. The Editors also cordially thank Judith Hinterberg and Mary James from Springer for keeping everybody aligned, on time, and happy during the process.

Biswanath Mukherjee
Ioannis Tomkos
Massimo Tornatore
Peter Winzer
Yongli Zhao

About the Editors

Biswanath Mukherjee is a Distinguished Professor Emeritus at the University of California, Davis, CA, USA. He holds a PhD from the University of Washington, Seattle (1987) and a BTech (Hons) from the Indian Institute of Technology, Kharagpur (1980). He is also a Distinguished Professor and Founding Director of the Institute for Broadband Research and Innovation (IBRI) at Soochow University in China. He has been involved in several successful optical startups, including Ennetix, a SBIR-funded company specializing in AI-powered, application-centric network analytics for optimizing the user experience. Biswanath Mukherjee has served as program chair for several OFC, IEEE INFOCOM, and IEEE Advanced Networks and Telecom Systems (ANTS) conferences, and cofounded the latter. He is a Series Editor for Springer's book series on optical networks, and has served on several journal editorial boards, including *IEEE/ACM Transactions on Networking* and *IEEE Network*. He has received multiple awards for his scholarly and educational achievements, and was the winner of the IEEE Communications Society's inaugural (2015) Outstanding Technical Achievement Award "for pioneering work on shaping the optical networking area." He was made an IEEE Fellow in 2006.

Ioannis Tomkos is a Professor of Optical Communications at the Department of Electrical and Computer Engineering at the University of Patras, Greece. His current research focuses on the use of optical communications systems for 5G/6G and datacenter networks. He has held numerous positions in industry and academia in various countries around the world (e.g., USA, Spain, Cyprus, Italy, and Greece). His research group plays a consortium-wide leading role in over 25 EU-funded research projects, and he serves as Technical Manager on 10 major EU projects. His published works have received around 10,000 citations (h-index = 47). In 2018, Dr. Tomkos was elected an IEEE Fellow "for contributions in dynamic optical networks." He is also an IET Fellow (2010) and a Fellow of the Optical Society (2012).

Massimo Tornatore is an Associate Professor in the Department of Electronics, Information, and Bioengineering at Politecnico di Milano, where he received a PhD degree in 2006. He is also an Adjunct Professor at the University of California, Davis, CA, USA and a Visiting Professor at the University of Waterloo, Canada. His research interests include performance evaluation, the optimization and design of communication networks (with an emphasis on the application of optical networking technologies), cloud computing, and the application of machine learning to network management. He has participated in several EU R&D projects as well as several other projects in the US, Canada, and Italy. Massimo Tornatore is a member of the journal editorial boards of *IEEE Communication Surveys and Tutorials*, *IEEE Communication Letters*, *Photonic Network Communications*, and *Optical Switching and Networking*. He is also an active member of the tech nical program committees of various networking conferences, and a Senior Member of the IEEE.

Peter J. Winzer received his PhD in electrical engineering from the Technical University of Vienna, Austria, in 1998. Supported by the European Space Agency, he investigated spaceborne Doppler lidar and laser communications. From 2000 to 2019, he worked at Bell Labs (NJ, USA), where he focused on fiber-optic communication systems and networks and contributed to many high-speed optical transmission records and field trials of up to 1 Tbit/s per carrier. Peter Winzer is actively involved with the IEEE Photonics Society and the Optical Society (OSA). He has served as Editor-in-Chief of the IEEE/OSA *Journal of Lightwave Technology* and was Program/General Chair of ECOC and OFC. Dr. Winzer is a highly cited researcher, a Bell Labs Fellow, an IEEE Fellow, and a Fellow of the OSA, as well as an elected member of the US National Academy of Engineering. He has received multiple awards, including the John Tyndall Award.

Yongli Zhao is a Full Professor in the State Key Laboratory of Information Photonics and Optical Communications at Beijing University of Posts and Telecommunications (BUPT). He is a Senior Member of the IEEE and an IET Member. Yongli Zhao received his PhD from BUPT in 2010. He has received support from the Youth Talent Plan of Beijing City (2013) and the National Science Fund for Excellent Young Scholars in China (2018). In the past five years, he has chaired and participated in more than 30 research projects, including the National High Technology Research and Development Program in China, the National Basic Research Program of China, and National Natural Science Foundation of China (NSFC) projects. He was a Visiting Associate Professor at UC Davis in 2016–2017. His current research focuses on software-defined optical networks, elastic optical networks, datacenter networking, machine learning in optical networks, optical network security, and quantum key distribution networking.

About the Authors

Alejandro Aguado Martín
Center for Computational Simulation
Universidad Politécnica de Madrid
Madrid, Spain
a.aguadom@fi.upm.es

Rod C. Alferness
University of California Santa Barbara
Santa Barbara, CA, USA
alferness@ucsb.edu

Cristian Antonelli
Dept. of Physical and Chemical Sciences
University of L'Aquila
67100 L'Aquila, Italy
cristian.antonelli@univaq.it

Pablo Jesus Argibay-Losada
BITS
SBI Group
Tokyo, Japan
pargibay@ieee.org

Alireza Behbahani
Turbine Engine Division, Advanced Integrated Controls and Prognostic / Diagnostic Systems Group
Air Force Research Laboratory (AFRL)
Wright-Patterson Air Force Base, OH, USA
alireza.behbahani@us.af.mil

Keren Bergman
Dept. of Electrical Engineering
Columbia University
New York, NY, USA
bergman@ee.columbia.edu

Alberto Bononi
Dipt. di Ingegneria e Architettura
Università degli Studi di Parma
Lesignano de' Bagni (Parma), Italy
alberto.bononi@unipr.it

Brad Booth
Microsoft
Snohomish, WA, USA
brbooth@microsoft.com

Gabriella Bosco
Dept. of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
gabriella.bosco@polito.it

Nicola Calabretta
Dept. of Electrical Engineering
Technical University of Eindhove
Eindhoven, The Netherlands
n.calabretta@tue.nl

Alberto Carrasco-Casado
Space Communications Laboratory
NICT – National Institute of Information and Communications Technology
Tokyo, Japan
alberto@nict.go.jp

Fabio Cavaliere
Ericsson
Pisa, Italy
fabio.cavaliere@ericsson.com

Mohit Chamania
Dept. of Research and Development
ADVA Optical Networking SE
Berlin, Germany
mchamania@advaoptical.com

Sethumadhavan Chandrasekhar
Optical Transmission Research
Nokia Bell Labs
Holmdel, NJ, USA
s.chandrasekhar@ieee.org

Gee-Kung Chang
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, USA
geekung.chang@ece.gatech.edu

Xi Chen
Optical Transmission Research
Nokia Bell Labs
Holmdel, NJ, USA
xi.v.chen@nokia-bell-labs.com

Dominique Chiaroni
Nokia Bell Labs France
Nozay, France
dominique.chiaroni@nokia-bell-labs.com

Brandon Collings
Lumentum
San Jose, CA, USA
brandon.collings@lumentum.com

Ronen Dar
Optical transmission systems
Nokia Bell Labs
Holmdel, NJ, USA
ronendar@yahoo.com

Robert D. Doverspike
Network Evolution Strategies, LLC
Tinton Falls, NJ, USA
r.d.doverspikeconsulting@gmail.com

Rudra Dutta
Dept. of Computer Science
North Carolina State University
Raleigh, NC, USA
rdutta@ncsu.edu

Jörg-Peter Elbers
Advanced Technology
ADVA Optical Networking SE
Martinsried, Germany
jelbers@adva.com

Mark Filer
Microsoft Corporation
Redmond, WA, USA

Marco Fiorentino
Hewlett Packard Labs
HPE
Palo Alto, CA, USA
marco.fiorentino@hpe.com

Pouya Fotouhi
Dept. of Electrical and Computer Engineering
University of California, Davis
Davis, CA, USA
pfotouhi@ucdavis.edu

Alexandre Graell i Amat
Dept. of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden
alexandre.graell@chalmers.se

Ashwin Gumaste
Computer Science and Engineering
Indian Institute of Technology, Bombay
Mumbai, India
ashwing@ieee.org

Hiroaki Harai
National Institute of Information and Communications Technology
184-8795 Tokyo, Japan
harai@nict.go.jp

Takemi Hasegawa
Optical Communications Laboratory
Sumitomo Electric Industries, Ltd.
Yokohama, Japan
hase@sei.co.jp

Alan Hill
Trinity College, CONNECT Telecommunications Research Centre, School of Computer Science and Statistics
University of Dublin
Dublin, Ireland
alanhill@timewave.f9.co.uk

Hiroki Ishikawa
Optical Communications Laboratory
Sumitomo Electric Industries, Ltd.
Yokohama, Japan
ishikawa-hiroki@sei.co.jp

Admela Jukan
ECE
TU Braunschweig
Braunschweig, Germany
ajukan@me.com

Jun-ichi Kani
NTT Access Network Service Systems Laboratories
NTT Corporation
Yokosuka, Japan
junichi.kani.wb@hco.ntt.co.jp

Daniel C. Kilper
Optical Sciences
University of Arizona
Tucson, AZ, USA
dkilper@optics.arizona.edu

Koteswararao Kondepu
Scuola Superiore Sant'Anna
Pisa, Italy
koteswararao.kondepu@sssup.it

David Larrabeiti
Universidad Carlos III de Madrid
Leganés, Spain
dlarra@it.uc3m.es

Xin Lin
Institute of Technology
Nakagawa Laboratories, Inc.
Tokyo, Japan
linxin@optinformation.com

Hong Liu
Google Inc.
Mountain View, CA, USA
hongliu@google.com

Victor Lopez
Telefonica
Madrid, Spain
victor.lopezalvarez@telefonica.com

Feng Lu
Wireless and Sensing Products Group
Semtech Corporation
San Jose, USA
flu@semtech.com

Andrea Marotta
DISIM
University of L'Aquila
L'Aquila, Italy
andrea.marotta@univaq.it

Ramon Mata-Calvo
Institute of Communications and Navigation
DLR – Deutsches Zentrum für Luft-und Raumfahrt
(German Aerospace Center)
Oberpfaffenhofen-Wessling, Germany
ramon.matacalvo@dlr.de

Tomokuni Matsumura
Nakagawa Laboratories, Inc.
Tokyo, Japan
tom@naka-lab.jp

David S. Millar
Mitsubishi Electric Research Laboratories
Cambridge, MA, USA
millar@merl.com

John E. Mitchell
Dept. of Electronic and Electrical Engineering
University College London
London, UK
j.mitchell@ucl.ac.uk

Biswanath Mukherjee
Dept. of Computer Science
University of California
Davis, CA, USA
bmukherjee@ucdavis.edu

Mehrdad Pakmehr
optoXense, Inc.
San Ramon, CA, USA
mp@optoxense.com

Nick Parsons
HUBER+SUHNER Polatis
Cambridge, UK
nick.parsons@hubersuhner.com

David Payne
Electrical Engineering
Trinity College Dublin
Dublin, Ireland
david.b.payne@btinternet.com

Thomas Pfeiffer
Bell Labs
Nokia
Stuttgart, Germany
thomas.pfeiffer@nokia-bell-labs.com

David Piehler
Dell Technologies
Santa Clara, CA, USA
david.piehler@dell.com

Pierluigi Poggiolini
Dept. of Electronics and Telecommunications
(DET)
Politecnico di Torino
Torino, Italy
pierluigi.poggiolini@polito.it

Josep Prat
Signal Theory and Communications Dept.
Universitat Politecnica de Catalunya (UPC)
Barcelona, Spain
jprat@tsc.upc.edu

Roberto Proietti
Dept. of Electrical and Computer Engineering
University of California, Davis
Davis, CA, USA
rproietti@ucdavis.edu

Chunming Qiao
Dept. of Computer Science & Engineering
University of Buffalo
Buffalo, NY, USA
qiao@buffalo.edu

Stojan Radic
Qualcomm Institute
University of California San Diego
La Jolla, CA, USA
sradic@ucsd.edu

George N. Rouskas
Dept. of Computer Science
North Carolina State University
Raleigh, NC, USA
rouskas@ncsu.edu

Marco Ruffini
Trinity College, CONNECT telecommunications research centre, Computer Science and Statistics
The University of Dublin
Dublin, Ireland
marco.ruffini@tcd.ie

Sébastien Rumley
Dept. of Electrical Engineering
Columbia University
New York, NY, USA
rumley@ee.columbia.edu

Roland Ryf
Advanced Photonics Research
Nokia Bell Labs
Holmdel, NJ, USA
roland.ryf@nokia-bell-labs.com

Takashi Sasaki
Innovation Core SEI, Inc.
San Jose, CA, USA
sasaki@sei-innovation.com

Seb J. Savory
Dept. of Engineering
University of Cambridge
Cambridge, UK
sjs1001@cam.ac.uk

Laurent Schmalen
Communications Engineering Lab (CEL)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
laurent.schmalen@kit.edu

Marco Secondini
Institute of Communication, Information and Perception Technologies
Sant'Anna School of Advanced Studies
Pisa, Italy
marco.secondini@sssup.it

Paolo Serena
Dipt. di Ingegneria e Architettura
Università degli Studi di Parma
Parma, Italy
paolo.serena@unipr.it

M. Ashkan Seyedi
Hewlett Packard Labs
HPE
Palo Alto, CA, USA
ashkan.seyedi@hpe.com

Jane M. Simmons
Monarch Network Architects
Holmdel, NJ, USA
jsimmons@monarchna.com

Björn Skubic
Ericsson Research
Ericsson AB
Stockholm, Sweden
bjorn.skubic@ericsson.com

William A. Stange
Universal Technologies Corp.
Dayton, OH, USA
bstange@utcdayton.com

Suresh Subramaniam
George Washington University
Washington, USA
suresh@gwu.edu

Giuseppe Talli
Photonic Systems Group
Tyndall National Institute, University College Cork
Cork, Ireland
giuseppe.talli@tyndall.ie

Ioannis Tomkos
ECE Department
University of Patras
Patras, Greece
itom@ece.upatras.gr

Massimo Tornatore
Politecnico di Milano
Milano, Italy
tornator@elet.polimi.it

Paul Townsend
Photonics
Tyndall National Institute, University College Cork
Cork, Ireland
paul.townsend@tyndall.ie

Stephen J. Trowbridge
Optical Networking Business Unit
Nokia
Boulder, CO, USA
steve.trowbridge@nokia.com

Ryohei Urata
Google Inc.
Mountain View, CA, USA
ryohei@google.com

Amin Vahdat
Google Inc.
Mountain View, CA, USA
vahdat@google.com

Luca Valcarenghi
Istituto TeCIP
Scuola Superiore Sant'Anna
Pisa, Italy
luca.valcarenghi@sssup.it

Michael Vasilyev
Dept. of Electrical Engineering
University of Texas at Arlington
Arlington, TX, USA
vasilyev@uta.edu

Doutje Veen
Bell Labs Access Lab
Nokia
Murray Hill, NJ, USA
dora.van_veen@nokia-bell-labs.com

Ricard Vilalta
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)
Castelldefels, Spain
ricard.vilalta@cttc.es

Sebastian Werner
Dept. of Electrical and Computer Engineering
University of California, Davis
Davis, CA, USA
swerner@ucdavis.edu

Peter Winzer
Nokia Bell Labs
Holmdel, NJ, USA
peter.winzer@ieee.org

Lena Wosinska
Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden
wosinska@chalmers.se

Mu Xu
Optical Center of Excellence
CableLabs
Louisville, CO, USA
m.xu@cablelabs.com

S.J. Ben Yoo
Dept. of Electrical and Computer Engineering
University of California, Davis
Davis, CA, USA
sbyoo@ucdavis.edu

Xiaosong Yu
Institute of Information Photonics and Optical Communications (IPOC)
Beijing University of Posts and Telecommunications (BUPT)
Beijing, China
xiaosongyu@bupt.edu.cn

Jie Zhang
Institute of Information Photonics and Optical Communications (IPOC)
Beijing University of Posts and Telecommunications (BUPT)
Beijing, China
lgr24@bupt.edu.cn

Yongli Zhao
Beijing University of Posts and Telecommunications
Beijing, China
yonglizhao@bupt.edu.cn

Xiang Zhou
Google Inc.
Mountain View, CA, USA
xzhou@google.com

Contents

Part D Optical Access and Wireless Networks

List of Abbreviations

3R	reamplification–reshaping–retiming
4-PAM	4-ary pulse amplitude modulation
4QD	four-quadrant detector
5G NR	5G New Radio
5GC	5G core

A

A-RoF	analog radio over fiber
AAA	authentication, authorization and accounting
ABNO	application-based network operation
ABRT	assured bandwidth restoration time
ACF	autocorrelation function
ACFS	approximate closed-form solution
ACL	access control list
ACN	access cloud network
ACO	analog coherent optics
ACT	active control technologies
ACTN	abstraction and control of transport networks
ADC	analog-to-digital converter
ADCT	adaptive diversity combing technique
ADM	add-drop multiplexer
ADSL	asymmetric digital subscriber line
AES	advanced encryption standard
AFDX	avionics full-duplex switched Ethernet
AFPM	asymmetric Fabry–Perot modulator
AGC	automatic gain control
AIMD	additive-increase-multiplicative-decrease
AIR	achievable information rate
AMCC	auxiliary management and control channel
AMP	asynchronous mapping procedure
ANN	artificial neural network
AOC	active optical cable
AoD	architecture-on-demand
AOM	acousto-optic modulator
AON	active optical network
APD	avalanche photodiode
APSK	amplitude and phase-shift keying
ARON	application-driven reconfigurable optical network
ARP	address resolution protocol
ARPANET	Advanced Research Projects Agency Network
AS-PCE	active stateful path computation element
ASE	amplified spontaneous emission
ASI	asymmetric information
ASIC	application-specific integrated circuit
ASON	automatic switched optical network
ASTN	automatic switched transport network
ATM	asynchronous transfer mode
ATMOS	asynchronous-transfer-mode optical switching
AUI	attachment unit interface
AWC	automatic wavelength controllers
AWG	arrayed waveguide grating
AWGN	additive white Gaussian noise
AWGR	arrayed waveguide grating router

B

BBG	baseband group
BBU	baseband unit
BDA	backbone destination address
BDD	bounded distance decoding
BDI	backward defect indication
BEC	binary erasure channel
BEI	backward error indication
BER	bit error rate
BFD	bidirectional forwarding detection
BFS	breadth-first search
BGA	ball grid array
BGP	border gateway protocol
BICM	bit-interleaved coded modulation
BIF	bend-insensitive fiber
BIM	baseband intensity modulation
BIP	bit-interleaved parity
BLSR	bidirectional line-switched ring
BMP	bit-synchronous mapping procedure
BMRx	burst-mode receiver
BOTDR	Brillouin optical time domain reflectometry
BP	belief propagation
BPD	balanced photodetector
BPDU	bridge protocol data unit
BPM	business process management
BPON	broadband PON
BPS	blind phase search
BPSK	binary phase-shift keying
BR	border router
BRAS	broadband remote-access server
BRF	bend-resistant fiber
BRPC	backward recursive path computation
BS	base station
BSA	bit-stream access
BSC	binary symmetric channel
BSS	business support software
BTC	block turbo code
BTS	base transceiver station
BTT	blade tip timing
BU	bandwidth update
BVT	bandwidth-variable transponder
BW	bandwidth
BXC	waveband cross-connect

C

C-MCF	coupled multicore fiber
C-RAN	centralized radio access network
CA	carrier aggregation
CA/CDA	channel aggregation/deaggregation
CaaS	computing-as-a-service
CAGR	compound annual growth rate
CAP	carrierless amplitude and phase modulation
CAPEX	capital expenditure
CaTV	cable TV
CAWG	cyclic arrayed waveguide grating
CB	coordinated beamforming
CBM	condition-based maintenance
CBR	constant bit rate
CC-MCF	coupled-core multicore fiber
CCDM	constant composition distribution matching
CCM	connectivity check message
CCR	carrier-to-crosstalk ratio
CCS	centralized CS
CCSDS	Consultative Committee for Space Data Systems
CD	chromatic dispersion
CDC	colorless, directionless and contentionless
CDF	cumulative distribution function
CDM	code-division multiplexing
CDN	content distribution network
CDR	clock and data recovery
CE	carrier-Ethernet
CEI	common electrical interfaces
CEI IA	Common Electrical I/O Implementation Agreement
CFI	canonical format indicator
CFM	connectivity fault management
CFP	40G/100G form-factor
CFP	C form-factor pluggable
CICQ	combined input-crosspoint queuing
CIOQ	combined input–output queuing
CIR	committed information rate
CLI	command line interface
CMA	constant-modulus algorithm
CMC	ceramic matrix composites
CMIS	common MIS
CMOS	complementary metal–oxide–semiconductor
CMP	chip multiprocessor
cMTC	critical machine-type communication
CO	central office
Co-DBA	cooperative dynamic bandwidth assignment
COI	channel-of-interest
CoMP	coordinated multipoint
COP	Control Orchestration Protocol
CORBA	common object request broker architecture
CORD	central office rearchitectured as data center
COTS	commercial off-the-shelf
CP	control plane
CPE	carrier-phase estimation
CPLD	complex programmable logic device
CPM	cross-phase modulation
CPO	co-packaged optics
CPRI	common public radio interface
CPS	confidential path segment
CPU	central processing unit
CQI	channel quality indicator
CR-LDP	constrained routed label distribution protocol
CRC	cyclic redundancy check
CRI	color rendering index
CRR	corner retroreflector
CRUD	create, read, update, delete
CRZ	chirped return-to-zero
CS	coordinated scheduling
CSC	customer SDN controller
CSCG	circularly symmetric complex Gaussian
CSF	cutoff-shifted fiber
CSI/CQI	channel state information channel quality indicator
CSK	color-shift keying
CSM	circuit switch manager
CSMA	carrier sense multiple access
CSP	constrained shortest path
CSRZ	carrier-suppressed RZ
CTAG	customer tag
CTS	clear to send
CU	central unit
CW	continuous wave
CWDM	coarse wavelength-division multiplexing

D

D-CPI	data-controller plane interface
D-RAN	distributed radio access network
D-RoF	digital radio over fiber
DA-RSA	distance-adaptive RSA
DAC	digital-to-analog converter
DAS	distributed antenna system
DBA	dynamic bandwidth assignment/allocation
DBORN	dual-bus optical ring network
DBP	digital backpropagation
DBR	distributed Bragg reflector
DBRu	dynamic bandwidth report upstream
DCAE	data collection analytics and events
DCC	digitized component carrier
DCF	dispersion-compensating fiber
DCI	datacenter interconnection
DCM	dispersion-compensation module
DCN	data center network
DCO	digital coherent optics
DCT	dispersion compensation technique
DCU	dispersion-compensating unit

DD	direct detection
DD-LMS	decision directed least mean squared
DDoS	distributed denial of service
DECS	distributed engine control system
DEI	drop eligibility indicator
DEMUX	demultiplexer
demux	wavelength demultiplexer
DetNet	deterministic networking
DFB	distributed feedback
DFE	decision-feedback equalizer
DFTS-DMT	discrete Fourier transform spread DMT
DGD	differential group delay
DGEF	dynamic gain equalizing filter
DiffServ	differentiated services
DL	downlink
DMD	differential mode delay
DMF	dispersion-managed fibers
DML	directly modulated laser
DMSC	discrete memoryless symmetric channel
DMT	discrete multitone
DN-MZM	double-nested Mach–Zehnder
DNN	deep neural network
DNS	domain name server
DOCSIS	data over cable service interface specification
DoD	Department of Defense
DOQ	distributed output-queued
DP	dual-polarization
DPI	deep packet inspection
DPM	demarcation point monitor
DPSK	differential phase-shift keying
DQPSK	differential quadrature phase-shift keying
DR	dynamic range
DRBS	double Rayleigh backscattered signal
DS	downstream
DSB-AM	double-sideband amplitude modulation
DSCP	differentiated services code point
DSF	dispersion-shifted fiber
DSL	digital subscriber line
DSLAM	digital subscriber line access multiplexer
DSO	digital storage oscilloscope
DSP	digital signal processing
DU	distributed unit
DVB	digital video broadcasting
DWA	dynamic wavelength assignment
DWBA	dynamic wavelength and bandwidth allocation
DWDM	dense wavelength-division multiplexing
DXC	digital cross-connect

E

E-NNI	external network-to-network interface
E-PON	Ethernet PON
E/O	electro-optic
EAM	electroabsorption modulator
EAT	electroabsorption transceiver
EB	electronic buffer
eBCH	extended BCH
ECL	external cavity laser
ECMP	equal cost multipath
eCPRI	enhanced CPRI
EDB	electrical duobinary
EDC	electronic dispersion compensation
EDFA	erbium-doped fiber amplifier
EDRS	European data relay system
EEPROM	electronically erasable programmable read-only memory
EF	exaFLOPS
EFEC	enhanced forward error correction
EGN	enhanced GN
EGT	exhaust gas temperature
EHF	extremely high frequency
EIRP	equivalent isotropically radiated power
ELF	extremely low frequency
EM	expectation-maximization
eMBB	enhanced mobile broadband
EMI	electromagnetic interference
EML	externally modulated laser
EMP	extrinsic message passing
EMS	element management system
ENOB	effective number of bits
EOM	external optical modulator
EON	elastic optical network
EPC	evolved packet core
EPON	Ethernet PON
EPS	electronic packet switching
ERO	explicit route object
eRP	enhanced RP
ERPS	Ethernet ring protection switching
ES	intercluster optical switch
ESCON	enterprise systems connection
ESS	emission sensing system
ETS	engineering test satellite
EVC	Ethernet virtual circuit
EVM	error vector magnitude
EXP	experimental

F

FADEC	full authority digital electronic controller
FASA	flexible access system architecture
FB	flattened butterfly
FBG	fiber Bragg grating
FBL	fly-by-light
FBNQ	fitting-based nonlinear quantization
FBW	fly-by-wire
FCAPS	fault identification, configuration, accounting, performance, security
FCS	frame check sequence
FDL	fiber delay line
FDM	frequency-division multiplexing
FEC	forward error correction
FER	frame error rate
FFE	feed-forward equalizer
FIB	forwarding information base
FIC	fabric interface chip

FICON	fibre connection
FIR	finite impulse response
FIT	failures in time
FLOPS	floating point operations per second
FLOW_MOD	flow table modification message
FM-HD	field modulation and heterodyne detection
FM-MFC	few-mode multicore fiber
FMDF	fiber main distribution frame
FMF	few-mode fiber
FML	frequency modulated laser
FOADM	fixed optical add/drop multiplexer
FOBS	fast optical burst selector
FOIC	FlexO interface
FOM	figure-of-merit
FOSS	fiber-optic sensing system
FOV	field of view
FP	Fabry–Pérot
FPA	focal-plane array
FPGA	field-programmable gate array
FPM	four-photon-mixing
FRLP	frequency-resolved LP
FRR	fast reroute
FSAN	full-service access network
FSC	fiber-switch capable
FSE	fast statistical estimation
FSO	free-space optics
FSOC	free-space optical communication
FSR	free spectral range
FTTH	fiber-to-the-home
FTTP	fiber to the premises
FTTx	fiber-to-the-home/building
FWA	fixed wireless access
FWI	forced wake-up indication
FWM	four-wave mixing

G

G-ACh	generic associated channel
G-LDPC	generalized LDPC
G-PON	gigabit-capable PON
GBIC	gigabit interface converter
GE-PON	gigabit Ethernet PON
GEM	G-PON encapsulation method
GEO	geostationary equatorial orbit
GF	gigaFLOPS
GFP	generic framing procedure
GigE	gigabit Ethernet
GMD	generalized minimum distance
GMI	generalized mutual information
GMP	generic mapping procedure
GMPLS	generalized multiprotocol label switching
gNB	next-generation node B
GOLD	ground/orbiter lasercomm demonstration
GPC	generalized product code
GPON	gigabit-capable PON
GPRS	general packet radio service
GSM	Global System for Mobile Communications
GTC	G-PON transmission convergence
GVD	group velocity dispersion

H

HAF	hole-assisted fiber
HALL	hierarchical all-to-all
HAP	high-altitude platform
HARQ	hybrid automatic repeat request
HC-PBGF	hollow-core PBGF
HDD	hard-decision decoding
HF	high-frequency
Hi-LIONS	hierarchical low-latency interconnect optical network switch
HICALI	high-speed communication with advanced laser instrument
HNLF	highly nonlinear fiber
HOM	high-order modulation
HOPR	hybrid optoelectronic router
HOS	hybrid optical switching
HOSA	hybrid optical switch architecture
HPC	high-performance computing
HPCG	high-performance conjugate gradient
HR-OSA	high-resolution optical spectrum analyzer
HSS	home subscriber server
HTS	high-throughput satellites
HTTP	hypertext transfer protocol

I

I-*L*PPM	inverted-*L*PPM
I-NNI	internal network-to-network interface
IA-RWA	impairment-aware routing and wavelength assignment
IaaS	Internet-as-a-service
IAD	intelligent access device
iBDD	iterative BDD
Ibpa	interblade phase angle
ICI	intercell interference
ICMP	internet control message protocol
ICR	integrated coherent receiver
ICT	information and communication technologies
ICTP	inter-channel termination protocol
ID	identification
IDA	ideal distributed amplification
IDC	inter-data-center
IE	information element
IF	intermediate frequency
IFFT	inverse fast Fourier transform
IFWM	intrachannel four-wave mixing
iGMDD-SR	iterative GMD decoding with scaled reliability
IGP	interior gateway protocol
IIR	intensity impulse response
IL	insertion loss

ILC	illumination-light communication
ILM	incoming label map
ILP	integer linear programming
IM	intensity modulation
IMS	Internet protocol multimedia subsystem
IoT	Internet of things
IP	Internet protocol
IP/MPLS	Internet protocol over multiprotocol label switching
IPC	input port interface card
iPLC	integrated planar lightwave circuit
IPoDWDM	IP over DWDM
IPR	intellectual property rights
IPSec	IP security
IS-IS-TE	intermediate system to intermediate system traffic engineering
ISG	industry specification group
ISI	intersymbol interference
ISIS-TE	intermediate system with traffic engineering
ISM	intelligent splitter monitor
ISP	Internet service provider
ITAG	intermediate service tag
ITU-T	International Telecommunication Union-Telecommunications
IXP	internet exchange point
IXPM	intrachannel cross-phase modulation

J

JDRS	Japan data relay system
JSF	Joint Strike Fighter
JSON	JavaScript object notation

L

L2SC	layer-2-switch capable
LAG	link aggregation group
LAN	local area network
LBS	location-based service
LC	line card
LCAS	link capacity adjustment scheme
LCE	laser communication equipment
LCOS	liquid crystal on silicon
LCP	local connection point
LCRD	laser communications relay demonstration
LCT	laser communications terminal
LDACS	L-band digital aeronautical communications system
LDC	linear divider/combiner
LDGD	largest differential group delay
LDP	label distribution protocol
LDPC	low-density parity check
LER	label edge router
LH	long-haul
LiFi	light-fidelity
LION	large-scale interconnect optical network
LIONS	low-latency interconnect optical network switch
LLDP	Link Layer Discovery Protocol
LLID	logical link identifier
LLN	linear lightwave network
LLR	log-likelihood ratio
LLU	local loop unbundling
LMP	link-management protocol
LMS	least mean squares
LNA	low-noise amplifier
LO	local oscillator
LOGON	locally-optimal globally-optimal Nyquist
LPF	low-pass filter
LR-PON	long-reach passive optical network
LSA	link state advertisement
LSC	lambda-switch capable
LSE	least-squares equalization
LSP	label-switched path
LSR	label-switched router
LTE	long term evolution
LTM	linktrace message
LUCE	laser-utilizing communications equipment

M

M2M	machine-to-machine
MAC	media access control
MAI	multiple access interference
MAN	metropolitan area network
MANO	management and network orchestration
MC-nodes	metro/core nodes
MCF	multicore fiber
MCLCD	Micius coherent laser communication demonstration
MCP	multiconstrained path
MCS	multicast switch
MCSB	maximal contiguous slot block
MD	mode dispersion
MD-ROADM	multidegree ROADM
MDIO	management data input/output
MDL	mode-dependent loss
MDM	mode division multiplexing
MDRU	movable and deployable resource unit
MDS	maximum distance separable
MDU	multiple dwelling unit
MEF	Metro Ethernet Forum
MEMS	microelectromechanical system
MEN	metro Ethernet network
MEP	maintenance end point
MET	multi-edge type
MFAS	multiframe alignment signal
MFD	mode-field diameter
MFH	mobile fronthaul
MGTC	minimum guaranteed transmission container content
MI	mutual information
MIB	management information base
MILP	mixed integer linear program

MIMO	multiple-input multiple-output
minEMBc	minimum effective modal bandwidth
MIP	mixed integer program
MLC	multilevel coding
MLD	multilane distribution
MMC	metal matrix composites
MMF	multimode fiber
MMT	multiple match table
mMTC	massive machine-type communication
MMUX	mode multiplexer
MMW	millimeter wave
MNH	multidomain network hypervisor
MOD	optical intensity modulator
MONET	multiple wavelength optical network
MP	minimum path
MPCP	multipoint control protocol
MPEG	Moving Picture Experts Group
MPI	multiple-path interference
MPLC	Multiplane light convertor
MPLS	multiprotocol label switching
MPO	(multifiber) connector
MPSK	*M*-level PSK constellation
MPTCP	multipath TCP
MR	maximum reach
MRR	microring resonator
MS	maximum scattering
MSA	multisource agreement
MSM	metal-semiconductor-metal
MSN	multiservice node
MSTE	minimum spanning tree with enhancement
MSTP	multiple spanning tree protocol
MT	match table
MTBF	mean time between failures
MTSO	mobile telephone switching office
MTTR	mean time to repair
MTU	maximum transmission unit
MZI	Mach–Zehnder interferometer
MZM	Mach–Zehnder modulator

N

NBI	northbound interface
NC	node controller
NCG	net coding gain
NCO	number-controlled oscillator
ND	nodal diameter
NETCONF	network configuration protocol
NetOS	network operating system
NF	noise figure
NFDM	nonlinear frequency-division multiplexing
NFS	Network File System
NFV	network function virtualization
NFVO	NFV orchestrator
NG-PON	next-generation PON
NG-RAN	next-generation radio access network
NGMI	normalized generalized mutual information
NGMN	next-generation mobile network
NIC	network interface card
NLFT	nonlinear Fourier transform
NLI	nonlinear interference
NLPN	nonlinear phase noise
NLRI	network layer reachability information
NMS	network management system
NOC	network operation center
NOMA	nonorthogonal multiplexing and multiple access
NRZ	nonreturn-to-zero
NSMS	nonintrusive stress measurement system
NSNI	nonlinear signal–noise interaction
NTE	network terminating equipment
NTP	network time protocol
NVM	nonvolatile memory
NZDSF	nonzero dispersion-shifted fiber

O

O/E	optic-electro
OA	optical amplifier
OADM	optical add-drop multiplexer
OAI	open air interface
OAM	operation, administration and maintenance
OAP	optical access point
OBI	optical beat interference
OBLC	optical burst line card
OBO	on-board optics
OBS	optical burst switching
OBSAI	Open Base Station Architecture Initiative
OBSC	optical burst switch card
OC	OpenFlow controller
OCDM	optical code division multiplexing
OCDMA	optical code division multiple access
OCh	optical channel layer
OCM	optical channel monitor
OCS	optical circuit switching
ODB	optical duobinary
ODN	optical distribution network
ODU	optical data unit
OE-PCB	optoelectronic printed circuit board
OEM	original engine manufacturers
OEO	optical–electrical–optical
OF	OpenFlow
OF-CONFIG	OpenFlow Configuration
OFDM	orthogonal frequency-division multiplexing
OFDR	optical frequency domain reflectometry
OFL	overfilled launch
OFLL	optical frequency locked loop
OFS	optical flow switching
OGS	optical ground station
OICETS	optical inter-orbit communications engineering test satellite
OLM	Optical Layer Monitoring
OLO	optical local oscillator
OLS	optical latching switch

OLT	optical line terminal
OM	output module
OMBc	overfilled modal bandwidth
OMCI	ONU management and control interface
OMM	ODN management module
OMS	optical multiplex section
ONAP	open network automation platform
ONOS	open network operating system
ONU	optical network unit
OOK	on–off keying
OPA	optical parametric amplifier
OPEX	operational expenditure
OPLL	optical phase locked loop
OPS	optical packet switching
OPU	optical channel payload unit
OQ	output-queuing
ORI	open radio equipment interface
OSFP	octal small form-factor pluggable
OSNR	optical signal-to-noise ratio
OSPF	open shortest path first
OSS	operations support system
OTDR	optical time domain reflectometry
OTL	optical trunk line
OTL	optical transport lane
OTLC	optical transport lane
OTN	optical transport network
OTS	optical transmission section
OTSiG	optical tributary signal group
OTU	optical transport unit
OVS	OpenVSwitch
OXC	optical cross-connect

P

PA	power amplification
PAM	pulse-amplitude modulation
PAM4	four-level pulse-amplitude modulation
PAPR	peak-to-average power ratio
PAS	probabilistic amplitude shaping
PBB	provider backbone bridging
PBC	polarization beam combiner
PBG	photonic band gap
PBGF	photonic band-gap fiber
PBS	polarization beam splitter
PBX	private branch exchange
PCB	printed circuit board
PCBd	physical control block downstream
PCC	path computation client
PCE	path computation element
PCEP	path computation element protocol
PCIe	peripheral component interconnect express
PCT	paired channel technology
PDCP	packet data convergence protocol
PDF	probability density function
PDG	polarization-dependent gain
PDH	plesiochronous digital hierarchy
PDL	polarization-dependent loss
PDM	polarization division multiplexing
PHY	physical layer
PIC	photonic integrated circuit
PL	photonic lantern
PLI	physical-layer impairment
PLL	phase-locked loop
PLOAM	physical-layer operations, administration and maintenance
PLOu	upstream physical layer overhead
PLZT	planar lightwave circuit
PM	polarization multiplexing
PMD	polarization mode dispersion
PMF	probability mass function
PNF	physical network function
POADM	packet optical add/drop multiplexer
POD	portable data center
POF	protocol oblivious forwarding
POL	passive optical LAN
PON	passive optical network
PON-ID	PON identifier
POS	packet-over-SONET
POTP	packet optical transport platform
PPM	pulse-position modulation
PPRN	phase and polarization-rotation noise
PQ	priority queuing
PRBS	pseudo-random binary sequence
PRC	primary reference clock
PROnet	programmable optical network
PSBT	phase-shaped binary transmission
PSD	power spectral density
PSF	point spread function
PSK	phase-shift keying
PSP	principal states of polarization
PTMP	point to multipoint
PTP	precision time protocol
PtP	point-to-point
PTS	partial transmission sequence
PUCCH	physical uplink control channel
PUSCH	physical uplink shared channel
PVC	permanent virtual circuit
PW	pseudowire
PWM	pulse-width modulation

Q

QAM	quadrature amplitude modulation
QKD	quantum key distribution
QoS	quality of service
QoT	quality of transmission
QPAR	quasi-passive reconfigurable node
QPSK	quadrature phase-shift keying
QSFP	quad small form-factor pluggable

R

RAID	redundant array of independent discs
RAM	random-access memory
RAN	radio access network
RAU	radio access unit
RCC	radio cloud center

RDC	regional datacenter
RDS	relative dispersion slope
RE	radio equipment
REC	radio equipment controller
REN	research and education network
REST	representational state transfer
RF	radio frequency
RFC	request for comments
RFL	reflection
RH	remote hub
RH-LIONS	reconfigurable hierarchical low-latency interconnect optical network switch
RIN	relative intensity noise
RLC	radio link control
RLS	recursive least squares
RMS	root mean square
RMSA	routing, modulation, and spectrum assignment
RMT	reconfigurable match table
RN	remote node
RNG	regional network gateway
ROADM	reconfigurable optical add-drop multiplexer
RoF	radio over fiber
ROLEX	rapid optical layer end-to-end X-connection
RPC	remote procedure call
RRC	radio resource control
RRH	remote radio head
RRU	remote radio unit
RS	Reed–Solomon
RSA	routing and spectrum assignment
RSCA	routing, spectrum and core assignment
RSOA	reflective semiconductor optical amplifier
RSP	restricted shortest path
RSSI	received signal strength indication
RSTP	rapid spanning tree protocol
RSVP	resource reservation protocol
RTS	request to send
RTT	round-trip time
RV	random variable
RWA	routing and wavelength assignment
RX	receiver

S

SAG	separated absorption and gain
SAN	storage area network
SAR	segmentation and reassembly
SBI	southbound interface
SBS	stimulated Brillouin scattering
SBVT	sliceable bandwidth variable transponder
SCM	subcarrier multiplexing
SD-WAN	software-defined wide-area network
SDA	software-defined access
SDH	synchronous digital hierarchy
SDM	space-division multiplexing
SDN	software-defined network
SDN	software-defined networking
SE	spectral efficiency
SER	symbol error rate
SERDES	serializer/deserializer
SFC	service function chain
SFF	small form-factor
SFL	substrate fiber link
SFP	small form-factor pluggable
SG-DBR	sampled-grating distributed Bragg reflector
SILEX	semiconductor laser inter-satellite link experiment
SINR	signal interference-to-noise ratio
SISO	soft-input soft-output
SLA	service level agreement
SMF	single-mode fiber
SMSR	side-mode suppression ratio
SNA	switched network access
SNIA	Storage Networking Industry Association
SNMP	simple network management protocol
SNR	signal-to-noise ratio
SOA	semiconductor optical amplifier
SONET	synchronous optical network
SPM	self-phase modulation
SRAM	static random access memory
SRLG	shared-risk link group
sRN	smart remote node
SRS	stimulated Raman scattering
SSG-DBR	super-structure grating DBR
SSH	secure shell
SSM	synchronization status message
SSMF	standard single-mode fiber
STAG	service tag
SWDM	short-wave wavelength-division multiplexing

T

T-API	transport application program interface
TAOGS	transportable adaptive optics ground station
TAPI	transport application programming interface
TCAM	ternary content-addressable memory
TCM	trellis-coded modulation
TCO	total cost of ownership
TCP	transmission control protocol
TDHF	time-domain hybrid format
TDM	time-division multiplexing
TDMA	time-division multiple access
TE	traffic engineering
TE	bridging-traffic engineering
TEC	thermoelectric cooler
TED	traffic engineering database
TFF	thin-film filter
TIA	transimpedance amplifier
TIR	total internal reflection
ToR	top-of-rack

TP	transport profile
TPC	turbo product code
TRCV	transceiver
TRX	transceiver
TSN	time-sensitive networking
TTT	timing transparent transcoding
TWC	tunable wavelength converter
TWDM	time and wavelength division multiplexing
TX	transmitter

U

U-MCF	uncoupled multicore fiber
UART	universal asynchronous receiver-transmitter
UCB	uncorrectable block
UCLP	user-controlled lightpath
UDWDM	ultra-dense wavelength-division multiplexing
UE	user equipment
UL	uplink
UL-JR	uplink joint reception
ULH	ultra-long-haul
UMTS	Universal Mobile Telecommunications Service
UNI	user–network interface
UPSR	unidirectional path-switched ring
URLLC	ultra-reliable low-latency communication
US	upstream
UVLWC	underwater visible-light wireless communication

V

VCAT	virtual concatenation
VCO	voltage-controlled oscillator
VCSEL	vertical-cavity surface-emitting laser
VDSL	very high speed digital subscriber line
VHTS	very-high-throughput satellites
VID	VLAN ID
VLAN	virtual local area network
VLC	visible light communication
VLM	virtual link mapping
VLSI	very large-scale integration
VLW-LAN	visible-light wireless LAN
VM	virtual machine
VNF	virtualized network function
VNM	virtual node mapping
VNTM	virtual network topology manager
VOA	variable optical attenuator
VOL	virtual optical link
VON	virtual optical network
VSOTA	very small optical transponder

W

W-LAN	wireless local area network
WA	wavelength assignment
WAN	wide area network
WBA	wavelength and bandwidth allocation
WDM	wavelength-division multiplexing
WR	wavelength routed
WRON	wavelength-routed optical network
WSS	wavelength-selective switch
WXC	wavelength cross-connect

X

XENPAK	10 Gb Ethernet transceiver package
XFP	10 Gb small form-factor pluggable
XG-PON	10-gigabit-capable PON
XGEM	10G encapsulation method
XGS-PON	10-gigabit-capable symmetric PON
XGTC	XG-PON transmission convergence
XML	extensible markup language
XOR	exclusive OR
XPM	cross-phase modulation

Y

YANG	yet another network generator

1. The Evolution of Optical Transport Networks

Rod C. Alferness

This introductory chapter describes the role of optics in networks, their capabilities, and their scaling limitations (different multiplexing techniques, i.e., TDM (time-division multiplexing), CDM (code-division multiplexing), FDM (frequency-division multiplexing), SDM (space-division multiplexing)). Types of optical networks installed around the globe are summarized, as well as their impact on society, market structure, and future perspectives.

Optical fiber transmission links were first deployed in the mid 1970s to provide 45 Mb/s capacity in metropolitan networks at a time when most traffic was wireline telephone communication. Few would have imagined then, when bandwidth demand on the telephone network was growing only as rapidly as population growth, that this new technology would radically alter business and everyday life by enabling the worldwide Internet. As the Internet grew in popularity and extended to all parts of the world, new devices and transmission technologies were invented and developed to cost-effectively achieve the required higher capacity transmission, and fiber was laid across continents and under the oceans to cover the globe. Thus began a virtuous cycle of higher capacity optical transmission systems and, ultimately, reconfigurable optical networks enabled by new technologies to meet increased demand, which resulted in new applications and services, such as video, which drove ever greater capacity and flexibility demand, which was, again, achieved via new technology innovation at significantly lower costs/capacity.

First-era systems grew the capacity of optical links by increasing the bit rate of information carried on the fiber. The second era, which was achieved cost-effectively by the optical fiber amplifier, increased capacity by multiplexing many wavelengths, each carrying independent information, onto a single fiber. The result was an increase in single fiber transmission equal to the number of wavelength channels and long spans over which all the wavelength channel signals could be periodically boosted with a single optically-powered optical fiber amplifier. The next era took the giant step of moving optics from transmission links only to fully reconfigurable, wavelength-channel-based networks to achieve higher efficiency (and, therefore, capacity), flexibility, and restorability by allotting and managing network capacity at the optical layer level.

This step required cost-effective optical switching components to provide the reconfigurable wavelength add/drop and cross connect functions. Today's optical networks provide superhighways of information bandwidth of the order 100 wavelength-defined lanes each providing hundreds of Gb/s information capacity. These networks provide network optimization to address changing capacity needs under software control via configurable wavelength on-and-off ramps and route switching centers. Without these optical networks, the global internet, cloud computing, and high bandwidth mobile services, including video, would not be possible. This chapter provides a view of the evolution of these optical networks—the market drivers, network architectures, transmission system innovations and enabling devices, and module technologies—which was a result of the efforts of a global community of researchers, developers, and, ultimately, manufacturers.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_1

1.1 Fundamental Developments

It is hard to imagine a world without optical communication networks. While commercial fiber optic transmission systems were introduced less than 50 years ago, today, because they underpin the global Internet, optical communications networks enable and enhance nearly every part of our daily lives. They are virtually invisible to the public because they are often in buried ducts or embedded in telco central offices. Yet, optical networks, including the fiber cables and the optoelectronics and optical switches that light and guide them, provide an infrastructure of long haul, metro, access, and, increasingly, local data-center networks, that have enabled broadband local and global information and communication and data interconnecting highways that drive the economy and have enhanced our daily lives. Now ubiquitously deployed under the oceans, across continents, and deep into metropolitan areas, fiber optics networks provide the high-bandwidth pipes over which all communication and data and Internet services are carried. To cost-effectively meet the bandwidth requirements of the incredible service demand growth, especially driven by video and more recently cloud services, the research and development community has been able to increase the capacity carried on a single fiber roughly a hundredfold every 10 years. The result of that invention, innovation, and commercialization is that new generations of optical networks to meet consumer demand for capacity growth have been deployed at costs that have grown at a substantially lower rate than the capacity demand. The result is that costs to the ultimate consumer for every expanded broadband service have been kept under control.

This sustained increase in the information bandwidth achieved on a single fiber is a result of continuous device invention and system innovation and introductions of new technologies. It is worth noting that to meet this demand, rather than increasing the capacity on a single fiber, one could employ multiple parallel fiber transmission systems. In this case, the network cost would grow roughly at the same rate as capacity. As we will see, the advantage of maximizing the capacity on a single fiber is, of course, avoidance of the extra cost and space of additional fibers to meet capacity demands. Yet what is even more important is to do so economically by cost-effectively scaling the other components and networking functions on the transmission links, rather than duplicating them on a second, parallel fiber.

An excellent example is the invention and development of the fiber optic amplifier that, together with wavelength multiplexing technologies, enabled the cost-effective commercialization of wavelength-division multiplexed (WDM) transmission. Innovation of wavelength add/drop and optical cross connects systems together with automated control software subsequently enabled fully reconfigurable optical networks that enable flexible bandwidth management and restoration capability in today's global networks. The currently-deployed optical networks have transmission links that are able to operate at Tb/s capacity on a single fiber by combining many wavelengths, each operating today at rates as high as 100 Gb/s with 400 Gb/s on the way. These networks are reconfigurable, under network control, at a wavelength granularity whose capacity is based upon the information bit rate carried by the wavelength channel [1.1, 2].

Without cost-effective high-capacity optical networks that span continents and connect them via undersea routes, the worldwide Internet would not be possible. Optical access systems are also essential to bring broadband access to that global Internet all the way to the home and to businesses. Increasingly important is that ubiquitous broadband optical networks provide the high-bandwidth backhaul essential for wireless access networks that have enabled today's smartphone users. These networks also provide the always-available, broadband access that will make cost-effective and energy-efficient cloud services available to all in the future.

Fundamental to the development and application of optical communications systems is the inherent ability to encode high-capacity information on an optical carrier. Additional fundamental enablers include the ability to achieve low transmission loss and low chromatic dispersion in silica fiber—to minimize pulse spreading and resulting interpulse (bit) interference during transmission over large distances—in a silica fiber fabricated as an optical guided-wave media. Finally, with

the addition of potentially low cost, low power, and small footprint semiconductor lasers, high-speed optical modulators and sensitive photo detectors, we have the key technologies to achieve bandwidth and distance performance that greatly exceeds that possible with electrical or microwave-based communication links.

These advantageous characteristics of optical communication can be achieved across the full spectrum of communication networks, undersea, terrestrial long haul, metro, access and for local networks, including data centers, but the great potential for advantage and first applications have been for high capacity over long distances—initially for long-haul undersea and terrestrial applications. With increasing bandwidth demands these advantages became cost-effective for the more modest distances of metropolitan networks as well. In this introductory chapter, we focus on optical transport networks for metro, national, and undersea applications, where the impact has been absolutely transformational.

1.2 Transport Networks

To appreciate the value of optical networks and, especially, to understand the importance of optical-switching (reconfigurability) elements in these networks, a short overview of transport networks is helpful. In the telecommunication network architecture, transport networks provide the very high capacity highways that enable logical pairwise connectivity (for simplicity assumed to be duplex) between all nodes. Large cities of the national network, as well as larger hubs in metropolitan networks, are the primary nodes in these networks. In Fig. 1.1 the key hardware elements of a transport network are shown [1.3]. Ultimately, transport networks provide high-capacity connections between all node pairs on the network. Rather than actually deploying a dedicated, separate physical connection (wire, cable, or fiber) between each node pair, which would be very expensive and operationally rigid, transport networks provide the connectivity between node-pair endpoints with virtual channels, each of which takes only a fraction of the capacity of the transmission medium. To be cost effective and to efficiently utilize the bandwidth of the transmission medium these lower-capacity channels (often referred to as *circuits* or *lines*) are multiplexed (aggregated) together on a single physical medium via a multiplexer (Fig. 1.1a) to provide connectivity via node-to-node links that operate at that bundled or aggregated bandwidth.

This aggregated information/data stream, encoded onto a carrier (radio-frequency, microwave, or optical) is sent over a transmission link on which the signal is re-

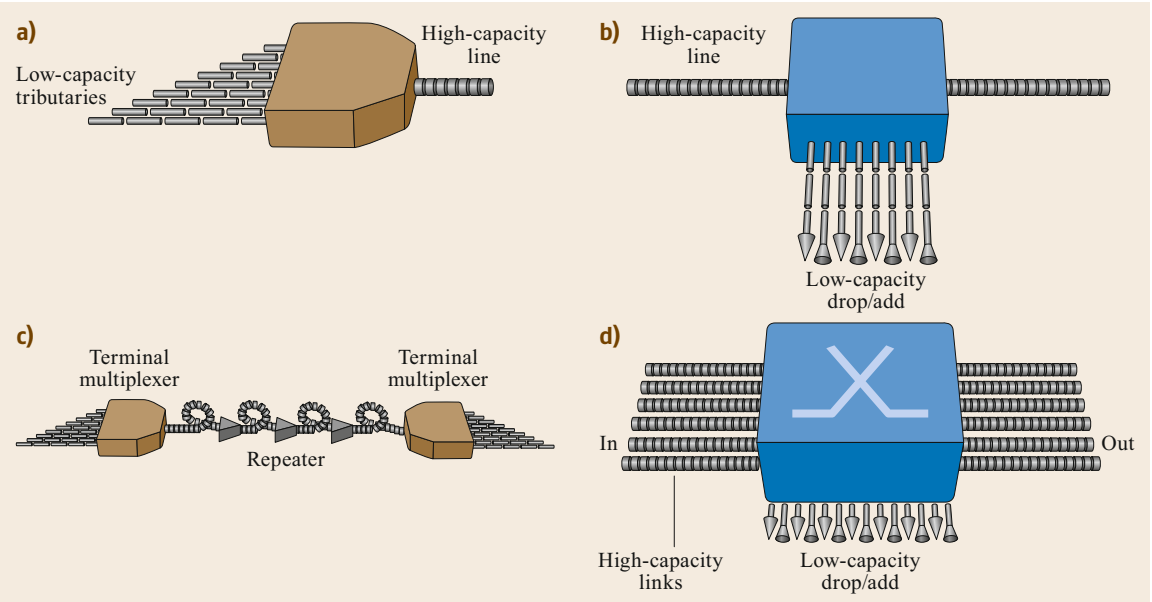


Fig. 1.1a–d Functional elements of telecommunication transport networks. **(a)** Terminal multiplexer, **(b)** add/drop multiplexer, **(c)** high-capacity transmission link, **(d)** cross-connect

generated periodically to compensate for transmission loss and impairments (Fig. 1.1c). At any node on the network, there are logical connections that represent the pairwise connectivity between that node and all other nodes on the network. Those logical connections need to be made at the destination node. This is achieved by unbundling the aggregated channels to access the channels that represent the pairwise connections to that node from the other network nodes. Those channels are then terminated, i.e., *dropped* at that node. Other pairwise connections may also be *added* at that node.

An important feature of transport networks is the different topologies of the interconnection of the transmission links between the nodes. Linearly connected architectures like a bus or a ring (typical for metropolitan networks) are characterized by a single link in and out of the node. In that case, terminated channels must be *dropped*, and all others are simply *passed through*. In addition to providing connectivity between that node and other nodes on the *output* side of the node, channels originating at the node must be added to the output link. The network element that performs this function in transport networks is an add/drop multiplexer (Fig. 1.1b).

In mesh-configured networks (used in national networks), each node may be connected by links to more than just the two nearest neighbors used for linear and ring networks. In mesh networks, multiple links (transmission routes) enter and exit the node. The transport switch element to route-aggregated traffic is more sophisticated in this case because in addition to providing the add/drop function for tributary channels for each transmission link entering the node, there is also the need to switch/route some disaggregated tributary (lower bit-rate *circuits* in the TDM case) traffic between output links to provide the desired pairwise connectivity across the network. The network switch element that provides this functionality is the cross connect (Fig. 1.1d). Because of the need to switch, at the tributary rate, between any of multiple input routes to any output route and do add/drop for all routes, cross connects are significantly more complex than add/drops, generally requiring more extensive switching fabrics. Both reconfigurable add/drops and cross connects are essential to provide the ability to change the network connectivity or the bandwidth of that connection to address capacity demand changes (increase) between node pairs or to facilitate the addition of nodes or to configure capacity around a network failure (restoration).

In early transport networks, prior to the use of optical transmission, the transmission was purely electrical, and the method of aggregation was electrical time-division multiplexing (TDM) where multiple lower bit-rate information-bearing channels were combined to form a single higher bit-rate channel (represented schematically in Fig. 1.1a). To do so requires synchronization of the lower bit-rate channels that depended upon distribution of centralized clock and electronics to do pulse narrowing and bit stream interleaving to form the higher rate-aggregated stream.

1.3 Enter Optical Fiber Transmission

The use of optical carriers to increase transmission capacity in transport networks was very natural because its higher carrier frequency inherently offered higher signal encoding bandwidth. Indeed, microwave frequency line-of-sight systems had already been investigated, as had exploration of a microwave that offers both ultralow loss and low chromatic dispersion. As was noted earlier, the latter results in minimal pulse spreading and resulting interpulse (bit) interference during transmission over large distances. At its most basic implementation, an optical transmission system requires an optical source whose generated dc optical signal can be modulated with information at the information bandwidth of interest, a low-loss fiber and an optical detector and receiver. First field trials were demonstrated in the mid 1970s. For example, AT&T demonstrated 45 Mb/s over one of its loop routes in Chicago in 1977 [1.4].

Today's networks provide tens of Tb/s capacity on a single fiber with nearly 200 wavelengths, each carrying an information capacity of 100 Gb/s. New products are pushing to 250 Gb/s, and shorter-reach systems operate at 400 Gb/s per wavelength [1.1]. In many cases, those optical signals—with the application of optical fiber amplifiers—travel over a thousand kilometers without the need for electrical regeneration. We have been able to reach this capability, driven by society's demand, by a virtuous cycle over the years of increased market demand driving innovations to cost-effectively increase capacity, which, in turn, drove the ability for the network to provide new applications, which then drove another round of market demands, etc.

The ability of fiber optics systems, and ultimately, reconfigurable wavelength-routed networks to respond to the increasing demand for ever higher capacity with increased reach and without depending upon power hungry and costly electronic regeneration, is a result of global efforts to continuously invent and innovate at the material, device, transmission system, network, and manufacturing level.

1.4 Single-Fiber Capacity as the Key Metric for Cost-Effective Optical Transmission Systems

To examine the evolution of optical transmission systems and, ultimately, reconfigurable optical networks, over the decades since the initial deployments, it has been useful to track a parameter that provides some measure of the cost-effective value that optical systems offer. That figure of merit is the information capacity that can be sent over a single fiber.

As we have noted, the communication transport architecture depends heavily on aggregation and the network demand to transmit information over relatively long distances. Furthermore, there is significant cost advantage when the transmitted information can be reconditioned—either regenerated or simply amplified—at the aggregated rate rather than requiring demultiplexing and reconditioning multiple signals at the tributary rates. This is the case regardless of the multiplexing parameter, time, wavelength or, potentially, as we note later, space. As a result, from the very beginning of innovation on optical fiber transmission systems, researchers have focused on the device and system technologies and transmission techniques that have enabled ever-higher aggregated transmission capacity that can be carried over a single fiber. Essential for this metric to provide the important economic value is that the technologies to cost-effectively recondition at the aggregated rate are commercially available.

In Fig. 1.2, we show the *hero* research transmission experiment summary of the maximum information capacity carried on a single fiber versus the year those results were achieved [1.1, 2]. These extraordinary *hero* transmission systems experiment results became the highlight of the postdeadline sessions at international fiber optics conferences, including the Optical Fiber Conference (OFC), the European Conference on Optical Communications (ECOC), and the Asia Communications and Photonics Conference (ACP). Commercial results typically followed a number of years later. For the purposes of this chapter, it is convenient and informative to describe the progress in information distribution capacity over the years in generations or eras of fiber optic transmission and, ultimately network, capacity increases. Those generations include time-division multiplexed systems, wavelength-division multiplexing, which includes also the introduction of reconfigurable optical networks, and, most recently, very high speed WDM networks. Each generation has built on top of the capability of earlier generations. We use these roughly-defined eras to provide an overview of the advances and achievements in optical communications over the years. Importantly, we also note the set of critically important optical component or system technology innovations that enabled each era.

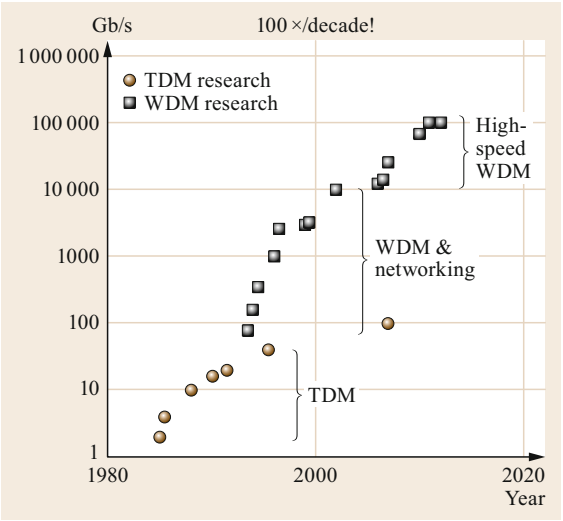


Fig. 1.2 Summary of trends of high-capacity *hero* transmission experimental results on a single fiber over recent decades

1.5 High-Speed Time-Division Multiplexing (TDM) Transmission Systems

Following the first field trials and initial deployment at rates of $\approx 50\,\mathrm{Mb/s}$, transmission capacity was increased by increasing the bit rate. The application was primarily long haul, where traffic from all users in a large city was aggregated together and sent over a fiber destined for other metropolitan areas. The basic service at that time was voice, which required only several kb/s, and demand was over local and regional distances. Aggregated rates on the long-haul fiber represented many aggregated voice sessions. Increase in capacity demands was still driven mostly by population growth, as well as some increase in new services such as fax. As capacity demand increased, given the broad bandwidth capability of fiber, new systems or upgrades of deployed systems could be achieved by increasing the speed of the transmitter and receiver and the regenerators. In the early days, the capacity increase per product generation was typically fourfold. That was

a number that had generally been large enough to handle demand for some time (typically ≈ 7 y) while not being overly demanding of the required advances in electronics. Of course, the economics demanded that cost of new systems scaled sublinearly with capacity, so that as systems moved to higher capacity to meet demand, the cost per bit was decreased.

Assuming higher bit-rate electronics (say $\times 4$) were available, or could be cost-effectively developed, as were the optical transmitter and photodetector, upgrading to a $\times 4$ transmission systems was more cost effective than building four-parallel transmission systems at the lower rate. The typical rule of thumb was that, once one paid for the cost of developing the next generation of electronics, a fourfold increase in transmission capacity could be achieved with only a roughly twofold increase in cost, resulting in a cost per bit that was reduced by a factor of 2.

The emergence and growth of multinational companies around the world put increased pressure on the transoceanic undersea-transmission facilities that had moved to optical fiber in 1988 on the TAT-8 undersea system. In addition, as computers began to become more common in business, the vision of data networks—computers talking to computers—began to suggest the need for an acceleration in network capacity. These new technology advances, together with growing global business opportunities, would dramatically change the global communication needs and, in turn, drive the demand for ever higher connection bandwidth.

After several generations of increasing the capacity by increasing the transmission bit rate, by the end of the 1980s commercial systems were operating at roughly 1 Gb/s. The first systems deployed were designed to be as simple as possible and to minimize the components. Information encoding was achieved by directly modulating the output power of the source laser by modulating the drive current. That function was accompanied with an undesirable wavelength *chirp* that resulted in optical pulse spreading (see below) if the transmissions fiber was dispersive at the transmission wavelength. To avoid that issue, especially at initial bit rates of less than ≈ 1 Gb/s, these early systems were dispersion limited and, thus, operated at fiber zero dispersion, 1.3 μm, wavelength.

To reach higher TDM rates required first and foremost the next generation high-speed electronics. In addition, for the same distance between electrical regeneration, both the signal strength relative to noise and quality of the detected signal with respect to interference are important. For the same fiber loss over a fiber optic link, the received electrical power scales as the inverse square of the signal bit rate. To help overcome the reduced receiver power at higher bit rates, researchers focused on systems operating at the lowest loss wavelength window around 1.55 μm. However, for the standard single-mode fiber deployed in the early 1990s, there was significant chromatic dispersion in the 1.55 μm window. This dispersion was problematic when using direct laser modulation where creating digital zeros and ones was by modulating the drive current of the semiconductor laser source to go from *low* output to *high* output. Unfortunately, to achieve a high *on/off* ratio, it was necessary to reduce the *off* drive current to the point where the laser was nearly turned off. As a result, in addition to the desired transition between a *zero* and *one*, there was also a small change in output wavelength. This wavelength *chirp* together with the chromatic dispersion at a wavelength of 1.55 μm resulted in pulse broadening, which limited transmission performance.

Three technology advances were instrumental to strongly mitigate these limitations and enable long-haul bit rates to go above roughly 1 Gb/s. (Fig. 1.3) First, to avoid chromatic dispersion, it was critical that the semiconductor laser operating at 1.5 μm be truly single frequency. Thus, laser structures, such as the distributed-feedback (DFB) laser that provided single longitudinal mode operation were critical.

Secondly, external optical modulators that provided optical information encoding without the *chirping* effects proved to be essential for data rates above several Gb/s. In particular, electro-optical modulators based on waveguide Mach–Zehnder (MZ) interferometers fabricated using titanium-diffused lithium niobate waveguides were shown to offer zero chirp and high on/off modulation with low insertion loss. With a traveling-wave electrode structure, such devices were shown early on to be capable of operating at ≈ 10 Gb/s with relatively low (3−5 V) drive voltage [1.5]. Hero systems experiments demonstrated that external modulators for information encoding became advantageous compared to direct laser modulation and, indeed, required for bit rates above about 4 Gb/s for long-haul applications [1.6]. Because of the ability to integrate multiple modulators, both amplitude and phase, this technology would also serve well the later generation systems' needs for advanced modulation formats, as we will see later.

Finally, optical detectors capable of operating at high bit rates and with reasonable optical to electrical conversion efficiency were also needed for high-speed TDM systems. High-gain, high-bandwidth avalanche photodiodes with separated absorption and gain regions SAG-APDs provided both [1.7]. The combination (Fig. 1.3) of single-frequency lasers operating at 1.5 μm externally modulated with waveguide modulators and detected with SAG APDs resulted in record transmis-

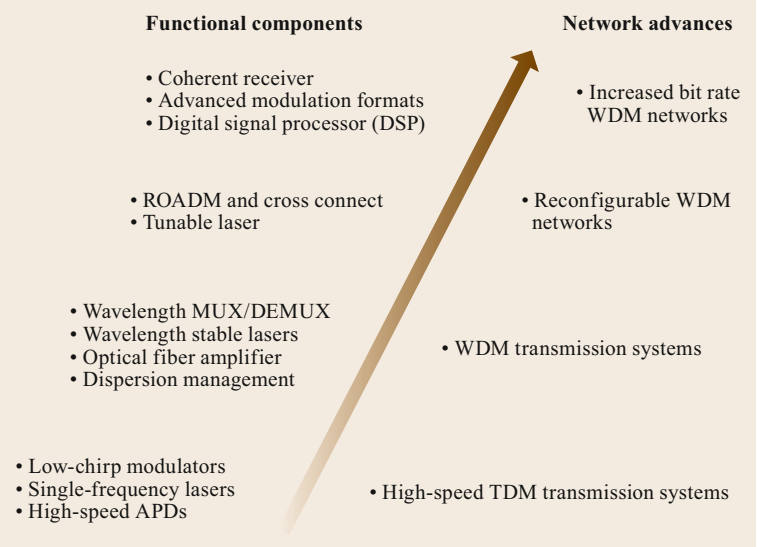


Fig. 1.3 Evolution of key underlying technologies that have enabled enhanced transmission systems and optical network functionality

sion experiments (2−16 Gb/s) in the late 1980s to early 1990s which led to commercially-deployed 10 Gb/s systems in the mid to late 1990s.

In TDM systems, leveraging improvements in the sources, receivers, and regenerators, the system's cost scaled roughly as the square root of the carried information capacity as long as the increased cost of these new higher-speed components (at volume) was not too high, and transmission impairments at the higher speed did not preclude the achievement of the fixed span distance between regenerator huts. So, starting with the first systems at 45 Mb/s to 2.5 Gb/s and 10 Gb/s, each generation at a higher bit rate substantially reduced the cost per bit of transmitted information. Thus, this, roughly, factor of 200 increase in capacity was achieved very cost-effectively.

In these TDM systems, electronics was still used to aggregate data to high data rates to take advantage of the bandwidth of fiber and TDM add/drop multiplexers, and cross connects continued to be done electronically. As we will see, that changed as the optical fiber amplifier and wavelength-division multiplexing were introduced, providing an additional layer of multiplexing and bandwidth management.

1.6 Wavelength-Division Multiplexed Transmission Systems

Increasing the TDM bit rate beyond a few Gb/s with the standard single-mode fiber that was, for the most part, the only fiber type commercially deployed in the mid 1990s, proved to be very challenging. To achieve the necessary loss budget with the laser launch power available or necessary to avoid fiber nonlinear effects, operation at 1.5 μm was essential. However, the high chromatic dispersion in standard single-mode fiber was challenging even with nonchirp external modulators. In addition, it became apparent that much of the deployed single-mode fiber displayed polarization mode dispersion that further impaired the transmitted signal. A potential solution to the first impairment was to use so-called dispersion-shifted fiber in which the zero-chromatic dispersion point is moved to the 1.5 μm window by choosing the appropriate material refractive indices and the fiber core dimensions. This approach was heavily researched and ultimately deployed for 10 Gb/s-TDM systems, especially in Japan.

The concept of wavelength-division multiplexed (WDM) transmission systems where multiple wavelengths, each separately encoded with information, are passively multiplexed together onto a single single-mode fiber, transmitted over some distance, then wavelength demultiplexed into separate channels whose information is detected and received, was well known from microwave transmission systems. While similar systems had been proposed for optical transmission systems in the early years, they had not gained popularity, partially because the necessary optical components were not available. More importantly, compared to increasing capacity via TDM, the approach did not scale capacity as cost-effectively as TDM, because signal regeneration required an array of electronics regenerators,

one for each wavelength rather than a single regenerator operating at the higher aggregated capacity. In addition, in such systems, a transmitter, a receiver, and a regenerator are needed for each wavelength on the fiber at each regenerator site. Thus, for electrically regenerated transmission systems, WDM did not offer a significant cost-effective capacity upgrade option.

That fundamental practical limitation to the use of WDM transmission systems changed dramatically with the demonstration and development of the erbium-doped fiber amplifier. While not a pulse regenerator, the optical amplifier provides a relatively low-noise 20–30 dB amplification, see, for example, [1.8]. From a system's value-add point of view, the optical amplifier offers the simultaneous amplification of multiple wavelengths, each carrying a high-capacity TDM signal. Importantly, there is no mixing of the information that is carried by the individual wavelengths. In addition, the fiber amplifier provides amplification for signals at essentially arbitrarily high bit rates for each wavelength. This is an enormously important and valuable feature of optical amplifiers that enables the upgrade of appropriately designed optical transmission systems to higher per wavelength bit rates without the need to replace amplifiers.

The earliest and most successful optical fiber amplifier is the erbium-doped amplifier in which the preform is doped with erbium prior to the fiber draw. When pumped with an intense laser source in the 0.98 or 1.48 μm region, the erbium-doped fiber amplifier provides optical gain over a, roughly, 40 nm-bandwidth centered around 1.55 μm, fortuitously well matched to the loss minima of single-mode transmission fiber. A gain of 20–30 db is achievable with reasonable pump power and is sufficient to compensate for loss of an ≈ 100 km span. Typical fiber amplifier length is 10–20 m. Amplifiers based upon Raman gain, which offer flexibility to control the center wavelength of the gain spectrum by tuning the pump laser, have also been deployed.

These gain characteristics combined with the other features above, make the optical amplifier an ideal amplifier for high-capacity WDM transmission systems. Long-haul transmission systems—including undersea ones that can greatly benefit from the ability to upgrade capacity by simply increasing the bit rate at the land-based transmitters and receivers—were the first natural application areas for this new multiplexing technology. Amplified WDM transmission systems, while still having competition from single-channel 10 Gb/s over the rather limited installed fiber plant that would support it, took off in the marketplace. It is important to note that from a market value point of view, it was not just increasing the network capacity that was important but the drastic reduction in the cost per bit of transmission capacity that this new WDM, fiber amplified technology made possible. It was not WDM alone, but rather the elimination of a number of single-channel electronic regenerators and transceivers with a single-multi-wavelength fiber amplifier enabled by WDM that provided a significant increase in cost-effective transmission capacity that made this new technology transformational for optical transport networks.

The other key enabling technologies for WDM transmission systems are the wavelength multiplexing and demultiplexing devices and single-frequency lasers that can be precisely set to a wavelength grid that allows matching source wavelengths to mux/demux devices wavelength response (Fig. 1.3). It is essential for these devices and for the entire value proposition of WDM to be viable that there was the agreement and standardization of the wavelength grid to ensure that devices from different vendors could be used by systems' vendors. A variety of technologies have been employed for wavelength multiplexers (mux) and demultiplexers (demux). Waveguide grating routers based on silica waveguide technology were typically employed for large wavelength applications. The 80-wavelength output from an early silica-based arrayed waveguide gratings router is shown in Fig. 1.4. Also important are the high-power (≈ 100 MW output power) semiconductor lasers to pump the optical fiber amplifier. Because fiber-amplified transmission systems are essentially analog systems, amplifier noise from each repeater builds up, as does dispersive and nonlinear pulse spreading. Therefore, low-noise amplifier operation and careful dispersion management are also important. Such considerations also make high-performance optical modulators that enable long-distance transmission especially important for signal encoding as one seeks to leverage the cost-effectiveness of the amplifier over longer distances without the need for costly electrically regenerators.

While the fiber-amplifier enabled the cost-effective deployment of WDM, there was another factor that drove lightwave systems to WDM. That was the performance of *standard* single-mode fiber that was in the field in the early 1990s. The deployed fiber had its minimum loss at 1.55 μm with the zero-dispersion wavelength of 1.32 μm. Prior to the WDM deployment, as work was focused on increasing fiber capacity by increasing the bit rate from 2.5 to 10 Gb/s, workers were hampered by deleterious pulse interference effects. Further study found that the limitations resulted from polarization mode dispersion—an effective mode refractive index that was polarization dependent. These results suggested that for such fibers—which represented a significant portion of the buried fiber—increasing TDM capacity beyond, roughly, 2.5 Gb/s

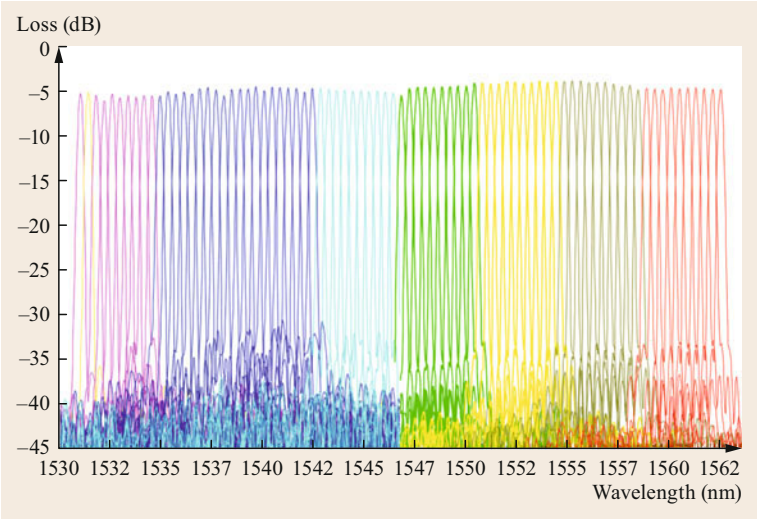


Fig. 1.4 Measured wavelength spectrum of a silica waveguide-based grating router 80-wavelength multiplexer/demultiplexer with 50 GHz channel spacing. Courtesy of Chris Doerr

over distances of typical regenerator spans would not be possible. Polarization dispersion limitations were not necessarily an issue for all embedded fiber at that time, and some single-channel 10 Gb/s-systems were deployed.

Another important enabler for WDM transmission systems was the nonzero dispersion shifted fiber [1.9]. For high-bit-rate systems, low chromatic dispersion is important to avoid the pulse spreading that results in interbit overlap and loss of information. However, in wavelength-division multiplexed systems a small amount of chromatic dispersion is useful. The nonlinearity of single-mode fiber is small but over distances of hundreds of kilometers typical of long-haul systems that four-wave mixing between wavelengths generate waves that spectrally overlap existing wavelength channels causing signal interference and degradation. A carefully designed small amount of chromatic dispersion can provide enough phase mismatch to effectively eliminate such deleterious effects.

Nonzero dispersion-shifted fiber can be designed as the transmission fiber, in which case the fiber is designed with sufficiently high dispersion in the 1.55 μm-wavelength region to minimize deleterious effects of four-wave mixing while ensuring that the dispersion is low enough to minimize pulse spreading for the planned per wavelength bit rate. Nonzero dispersion-shifted fibers can also be used as dispersion compensating fiber placed at the amplifier site to undo dispersion in the transmission fiber. In this case, the transmission fiber has sufficient dispersion over the transmission distance to avoid four-wave mixing, but the resulting pulse spreading is undone by the compensating fiber.

As alluded to earlier, undersea lightwave systems were an important driver and early adopter of fiber-amplified WDM transmission technology. An important issue for deployment of undersea TDM optical transmission systems was the length of time to develop and do reliability and accelerated aging for the new electronics and optoelectronics devices to be deployed in the submerged regenerator pods because of the high cost to replace undersea electronics should it fail prematurely. This requirement meant that every new generation of higher capacity required the higher-speed electronics needed to pass stringent reliability standards. With fiber-amplified WDM systems, there are no high-speed electronics under water, so issues associated with reliability testing required less lead time. The first such system—a transatlantic system—included 16 wavelengths at 2.5 Gb/s, each with repeater spacing of 100 km.

As multiplexing devices and amplifier performance were improved and techniques to mitigate dispersive and nonlinear transmission impairments developed, single/fiber transmission capacity results in research labs around the world improved—sometimes quite dramatically—every year. These extraordinary *hero* transmission-systems' experimental results became the highlight of postdeadline sessions at major optical conferences around the world (Fig. 1.2). Increased capacity in transmission systems' experimental results over the years (Fig. 1.2) were achieved by increasing the per-wavelength bit rate from 2.5 to 10 to 100 Gb/s and by increasing the number of wavelength channels. The latter was also achieved by adding additional gain bands. With respect to increasing the capacity on each wavelength, the issues that needed to be addressed included demonstrating the high-speed electronics, modulator, and receivers at the higher rates, mitigating fiber nonlinearities and managing dispersive effects. Increasing

the total single-fiber capacity by increasing the number of wavelengths was achieved either by enhancing the spectral efficiency, i.e., reducing the wavelength spacing, without reducing the information rate/wavelength or by increasing the spectral bandwidth of the amplifier.

The first WDM commercial systems, deployed in terrestrial long-haul applications in the mid-1990s, employed eight wavelengths at 2.5 Gb/s, a tenfold improvement over the single-channel systems previously available. The typical amplifier span distance was 50−100 km, and the maximum distance between regenerators was typically roughly 600 km—a distance that was limited more by network considerations than the maximum achievable distance before full regeneration was necessary. (That network consideration was that for distances longer than 600 km, typically some portion of the traffic needed to be dropped and traffic added for the route to be cost-effective. Prior to adoption of optical wavelength add/drops, as discussed below, that could only be done electronically, in which case it made sense to do full electrical regeneration.) As multiplexing devices and amplifier performance were improved, the number of wavelengths was soon doubled and then quadrupled. In research labs, work focused on WDM transmission at higher TDM rates—10 Gb/s and 40 Gb/s. In addition, wavelength mux/demux technology was improved, so that by the early 2000s commercial amplified WDM systems with 160 wavelengths at rates of 10 Gb/s per wavelength for 1.6 Tb/s total capacity were being deployed. These systems also operated over a distance of about 600 km, again limited mostly by network connectivity constraints, as noted. Tunable lasers have also become essential for cost-effective WDM systems [1.10, 11]. Having a single laser that can be tuned to any of the wavelengths carried by a given system drastically reduces the cost of excess inventory of lasers needed both for the system's vendors during deployment and also for network carriers for replacement or for adding wavelengths. The latter became even more important as reconfigurable networks evolved, as we will see below.

1.7 Bandwidth Management at the Optical Layer Via Optical Add/Drops and Cross Connects

The adoption of wavelength-division multiplexed transmission enabled by multiwavelength optical amplifiers has, in line with our earlier discussion of the general features of transport networks, rather naturally led to wavelength-based reconfigurable optical networks that provide high-level, cost-effective network bandwidth management, including restoration after fiber cuts or equipment failures. The roadmap of that evolution is shown schematically in Fig. 1.5 [1.2].

These optical layer networks, as noted earlier, are the optical wavelength-based equivalent to the time slot-based electrical TDM networks. However, this is a higher layer transport network whose networking granularity is a wavelength channel with its rate dictated by the TDM channel it is carrying. For a given wavelength, TDM networks continued to be used to provide bandwidth management at a lower total capacity and at finer bandwidth granularity via TDM add/drops and cross-connects.

Initially, WDM was employed over linear transmissions links where all wavelengths were aggregated onto the fiber at one node and carried with periodic amplification to an end node, where all wavelength channels were terminated (Fig. 1.5a). At those nodes, typically signals on a wavelength would be received and converted to electrical signals demultiplexed into lower TDM tributaries, some of which were dropped locally; others might be cross connected to other routes via TDM cross connects. In the mid to late 1990s, on average the longest distance over which the long-haul transmission went before it passed a major node at which traffic needed to be dropped or added was about 600 km. Amplified WDM systems that leveraged the value of the amplifier could carry signals to longer distances without regeneration but without the ability to get traffic on and off the line; it was not cost-effective to do so. Optical wavelength-add/drop multiplexers (Fig. 1.5b) provide those high-capacity on/off ramps with a full wavelength of capacity and allow all other wavelengths to pass through the node, benefiting from the amplification of the optical amplifier but not requiring it for each by-passed wavelength. This ability to optically add and drop selected wavelengths and at the same time allow the *pass-through* wavelengths to leverage the amplifier and not require a regenerator for each wavelength provides the ability to reduce opto-electronic regenerators while meeting the connectivity requirements of the network nodes. The result is the elimination of regenerators and extension of the basic value proposition of the amplifier to a multinode linear network rather than to a single node-to-node link.

Initially, to minimize capital costs, these add/drops were fixed in the sense that the number and actual wavelengths that could be added or dropped was set at

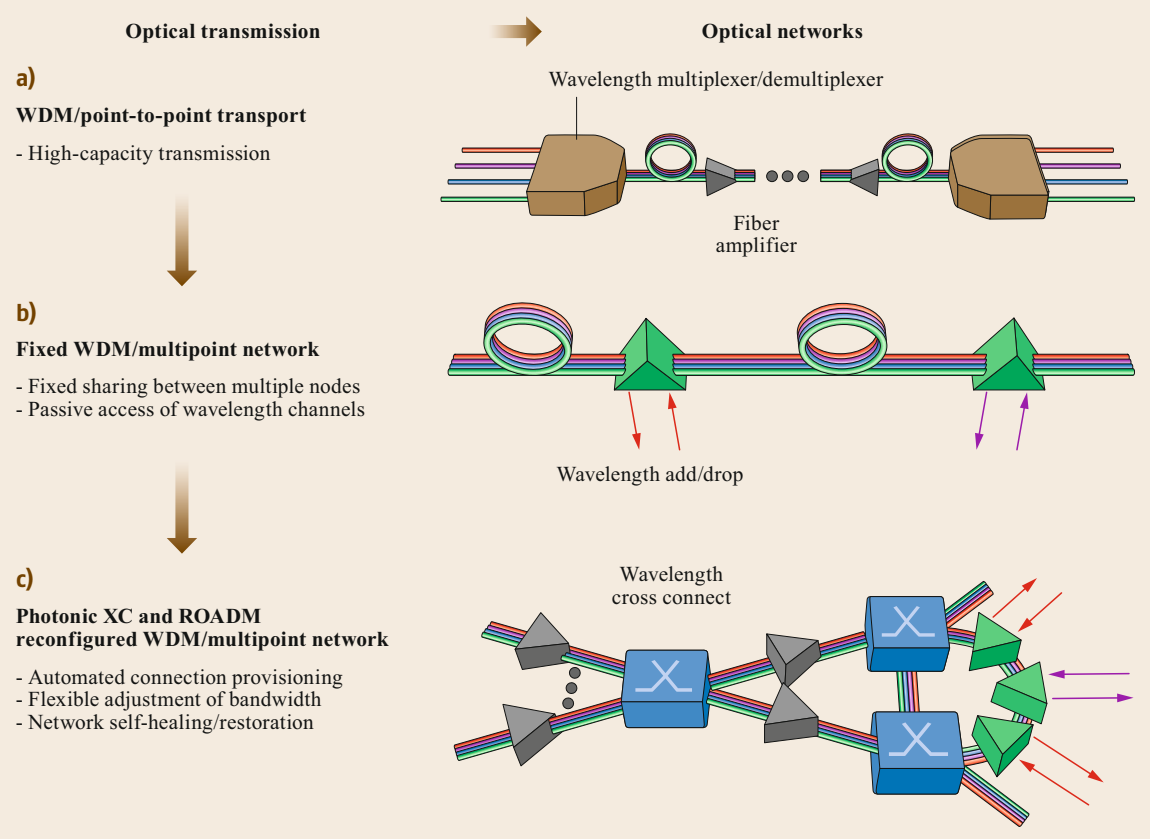


Fig. 1.5a–c Evolution path from WDM transmission systems to fully reconfigurable wavelength-routed optical networks that employ ROADMs and optical cross connects

deployment. Typically, the number of possible added or dropped channels is only a fraction of the total number of wavelengths on the fiber, enough to serve the traffic demands of that node. However, capacity requirements change, and service providers want the ability to change the configuration. Therefore, even though most of the add/drop multiplexers were initially fixed to contain cost, the direction toward reconfigurable optical add/drop multiplexers, referred to as ROADMs (reconfigurable wavelength add/drop multiplexers) was clear (Fig. 1.5c). As technology has improved and costs been reduced, ROADMs with their flexibility for remote reconfigurability have become central to commercially deployed networks. In mesh networks at nodes where two or more routes enter and exit (Fig. 1.5c) reconfigurable cross connects provide the ability to add/drop wavelengths of capacity for each route in each direction and also provide the ability to directly cross connect wavelengths between routes.

The vision of reconfigurable optical layer networks did not come out of a vacuum. Prior to the amplifier and commercial introduction of WDM transmission systems there was considerable research activity to explore reconfigurable optical networks, but mostly for LAN and local-access applications. Optical 2×2 space-division switches and switch arrays using integrated optics techniques, especially using titanium-diffused lithium niobate waveguide technology, were demonstrated with both time-multiplexed switching and transport network cross-connect applications as potential drivers [1.12]. Tunable wavelength-selective 2×2 waveguide switches that demonstrated the ROADM function were also demonstrated [1.13, 14]. So, innovating and applying optical switching technologies to achieve reconfigurable optical networks were explored. However, without deployed or planned networks whose multiplexing to aggregate bandwidth was based upon an optical parameter, all that work was a solution looking for a problem.

The emergence of the optical amplifier and the commercial deployment of WDM transmission systems dramatically changed that picture. The multiplexing parameter by definition becomes the parameter around which network bandwidth management is achieved via

add/drop and cross connection. In addition, the availability of highly efficient, high-gain optical amplification provides loss compensation for both the transmission losses, as well as the inevitable loss in the optical switching elements.

WDM transmission systems and the emerging optical networking vision came at a time (mid to late 1990s) when globalization and computer networks were driving predictions of high network capacity demand and reach. It was also a time of intense venture investment. The WDM optical networking vision gained the support of funding agencies and companies in Europe, the United States, and Japan. Several generations of optical cross connects and WDM network projects were funded by the European Framework Programs that fostered cooperation between universities and corporations. Japan initially chose the strategy to higher bit rates and deployed dispersion-shifted fiber to mitigate signal distortion issues at bit rates beyond 2.5 Gb/s. Nevertheless, there was considerable interest in WDM-network research in Japan as well. In the US, DARPA (Defense Advanced Research Project Agency) funded several generations of WDM and optical switching programs, including both local-area network and switched-network programs even before it was clear that WDM transmission would break into commercial deployment, see, for example [1.15]. While a couple of simple metrics—total bandwidth per fiber and distance without requiring regeneration—define transmission systems' performance, networks are more complicated. To help understand the initial challenges and hurdles that needed to be overcome to pave the way for commercial networks, below we describe the DARPA-funded MONET (multiple wavelength optical network) project that was an early research, prototyping, and field trial example of a reconfigurable WDM transport network.

DARPA, through its ARPANET (Defense Advanced Research Project Agency Network) programs had been a strong champion of packet-based data networks that gave them a vision of the importance and requirements for future data networks. Aware of the growing future needs for defense, they understood the communication bandwidth that would be needed to support communications, as national security networks increasingly moved from person-to-person to machine-to-machine connections. They were especially interested in a network model that would meet demanding government needs and grow cost-effectively. There was experience at the time with so-called *dual-use-networks* using TDM technology. The idea was for commercial service providers to build and operate a network that could have secure, separate subnetworks for government and normal commercial traffic. Included in this model was the view that under unusual circumstances of heavy demand, the partitioning of the network could be reconfigured to give priority capacity where needed by government offices. In the case of a WDM network that suggested full wavelengths, literally *virtual fibers*, that could be used for the subnetworks.

A consortium that included United States service providers—AT&T, Bell Atlantic (now Verizon), Bell South (now AT&T) and Southwest Bell (now AT&T), and Bellcore, the regional service providers research arm—was one of the awardees of the DARPA call to drive the vision outlined above. The consortium's program was called the multiple wavelength optical network (MONET). AT&T was also the parent of the AT&T Network Solutions business, which built optical infrastructure (including, shortly after the call, one of the first WDM transmission systems) and Bell Labs, the research arm. On the government side, in addition to DARPA, there was strong partnering and close field trial collaborations with NSA, NRL, DISA, and others. The overall goal, through test beds and field trials, was to demonstrate the technical feasibility of national and metro reconfigurable WDM networks with prototype network elements—ROADM and optical cross connects—that could be prototyped with the research-grade components that were available at the time [1.16].

MONET was vertically integrated. The program included research, network architecture design, and the necessary prototyping of ROADMs and reconfigurable optical cross connects and network elements. Both metropolitan and long-haul networks were built and ultimately connected via the cross connect test bed in a metro to long haul to metro (actually the same metro) end-to-end network. The network management system controlled the network elements to provide the desired end-to-end configuration. Bellcore's metro network and the long-haul and optical cross connect test beds of AT&T, including Bell Labs, were connected by fiber to demonstrate end-to-end networking [1.17]. Later, over commercially laid (Bell Atlantic) fiber connections, several US government agencies in the Washington DC area, a successful field trial of a reconfigurable optical network that included both ROADMs and a reconfigurable [1.18] optical cross connects was achieved.

The work included careful consideration of network operations and management that resulted in device requirements previously not recognized as important. As an example, in a discussion between device researchers and carrier network operators, the question was raised about how the amplifier gain characteristics might change in time as several input wavelength channels were suddenly dropped out of the input stream. Research identified the time-dependent gain changes to the pass-through wavelengths when the number of wavelengths input to the amplifiers changed rapidly.

Research further showed that these gain transients were unacceptable, and mitigation techniques to circumvent them had to be developed [1.18].

As noted above, the MONET network element prototyping was done with available lab prototype optical component technologies. Liquid crystal technology was used as the active optical switching element in the metro ROADM. For the wavelength cross connect and larger ROADMs, a wavelength demux/mux combined with a layer of space-division switches for each wavelength was selected. Lithium niobate-integrated optic switch matrices were employed as the space switches [1.12]. Some research was also performed on new component technologies that offered greater promise for the future. As an example, MONET researchers demonstrated one of the first multiwavelength receivers on a chip [1.18].

The program, which gave strong support for the potential value and practical feasibility of reconfigurable, wavelength-routed networks was completed in the end of the 1990s, just as globalization and the Internet were driving increased focus on ever-higher capacity demands for national and international networks,

It is important to emphasize that several conditions do need to be met for the use of add/drop multiplexers to be preferred to full demultiplexers at each network node. To be cost-effective, a significant fraction of network node pairs should have capacity demands that require or will soon require a full channel (wavelength) at the maximum supported TDM. Adequate transmission reach to ensure that the information-bearing wavelengths launched at one end of the network can be transmitted across the network without unacceptable noise or distortion is also necessary. Even though they carry digitally-encoded information, the transmission links of optical networks with optical amplifiers for gain and transmission-induced distortions are inherently analog-like systems. Consequently, optical signal-encoding technologies that are least susceptible to transmission impairments are essential to achieve increased reach and the economic benefit that results because of eliminating costly regenerators. In the early days of optical transmission, that really meant no chirp in the modulated signal—something that could be achieved with external electro-optic modulators but not with directly-modulated lasers. The evolution to optical networks with their ability to get traffic on and off the network optically meant that reach well beyond 600 km could also bring value by reducing electronic regenerators while meeting the network-access and connectivity requirements. That drove the demand for even longer optical layer reach. Today's optical networks span the nation, providing optical wavelength connectivity between major cities without requiring electronic regeneration in the transmission path. This ability of externally modulated signals to travel over long distances with limited transmission-induced distortion is valuable to cost-effective optical networks. It has assured an important place in the marketplace for optical modulators. That trend has held up even today as at modulation rates of 100 and 400 Gb/s advance modulations schemes that use both amplitude and phase modulation in nested configuration are required. These critically-important modulation techniques can only be achieved with electro-optic modulators.

Relatively low-cost ROADMs are also important to make the *optical bypass* value proposition complete. Roughly, the cost of ROADM plus the high-gain amplifier should to be less than that of a simple multiplexer/demultiplexer pair plus the cost of the N electrical regenerators and N transceivers, where N is the total number of wavelengths on the line. It is important not to forget that at the system level as one tries to extend the optical bypass value proposition by extending the reach of the network there is also potential additional transmission cost to enable optical signals to travel over the extended length without unacceptable noise accumulation and signal distortion. This becomes particularly important at elevated TDM rates. Some optical line-conditioning elements, such as dispersion-compensating fiber, e.g., maybe required. Such cost, normalized by the number of nodes (add/drop sites), should really be added to the add/drop multiplexer plus amplifier cost when doing the comparison with full regeneration cost.

Given the above caveats, it is reasonable that if the amplified WDM-transmission system wins over a regenerated system for intracity systems separated by several hundred miles, then, if achievable from a transmission reach point of view, a linear WDM network extending over several thousand miles, with wavelength add/drop to allow on and off traffic at cities along the route, will also provide a good business case, perhaps an even better one. This is the amplified WDM value proposition extended to the linear multinode network enabled by a reconfigurable wavelength add/drop multiplexer (ROADM) switch.

With the understanding and appreciation of the value of ROADMs in a linear network, including a ring, one can see that a similar value proposition applies to mesh-based architectures (Fig. 1.5c) where any node can be connected directly to several other nodes. It is important to keep in mind that optical cross connects provide both wavelength-based route switching, as well as the local wavelength add/drop. As an example, consider a national network with a node (medium to large city) that serves as a hub with connections to other cities via a north/south route and an east/west route going through it. At this hub, there may be traffic terminating at the node from both the north/south traffic and from the east/

west traffic. There is also *pass-through* traffic that maybe passing through on the same route or passing through from one route to another (switching routes). Given the multitude of possible pathways, this is a much more complex switching function than a ROADM. Nevertheless, the optical cross connect that includes local wavelength add/drop that provide this function achieves value in the same way as the ROADM—eliminating costly regenerators at branching points in the network. It also offers the direct value of switching a wavelength between routes. If there were no optical switching layer, and in the absence of line rate electrical cross connects (the case in the early days of 2.5 and 10 Gb/s and today at 100 Gb/s), the signals on wavelengths that needed to be switched to a different route would have to first be detected and demultiplexed down to an electrical TDM rate for which digital cross connect fabrics are available.

The enabling technologies for ROADMs and cross connects (Fig. 1.3) are electrically-controlled optical switches, either broadband space switches together with a wavelength multiplexer/demultiplexer component or a single-module wavelength-selective switch (WSS). Integrated wavelength-selective switch modules that include an internal grating-based free-space wavelength to position mapping onto an array of electrically-controlled variable angle-deflecting elements provide a modularized $1 \times N$ WSS that is a compact building block for both ROADMS and for full optical cross connects, see, for example, [1.19, 20]. The N output fibers in this case can include multiple wavelengths on bypass fibers that represent the different network routes. In a simple ROADM on a ring topology, there would be only one of these, whereas for cross connects there may be several, reflecting the different output routes. The other output fibers from this module would typically be single-wavelength client-side drop (add) ports. A variety of technologies have been used for optical switch fabrics, including micromechanical adjustable angle mirrors (MEMs) on silicon, liquid crystal on silicon (LCOS), and electro-optical or thermo-optical waveguide space switches, see, for example, [1.19, 20].

Commercial wavelength-reconfigurable optical networks have been widely deployed. Initially, these networks were deployed in national networks, but with the dramatic growth in transport bandwidth needs driven by video and broadband wireless, reconfigurable optical networks are now massively deployed in metropolitan applications as well.

There have been significant advances in both the architecture and the component technologies of both ROADMs and cross-connect systems since the initial deployments. With the high-performance $1 \times N$ wavelength-selective switch (WSS) components described above, the add/drop architecture becomes one in which only the added or dropped wavelengths are simultaneously mux/demuxed and switched to the desired output port. All the other wavelengths remain on the output fiber. This provides a significant component count and cost reduction, especially for the case where the percentage of the channels added or dropped at any given node is small. It also provides a powerful component for the creative design of optical cross connects for applications in network nodes where a limited number of routes intersect.

Overall, the design considerations and tradeoffs for ROADMs and cross connects are very similar to those of their electronic predecessors. At the highest level, there is a tradeoff between operation flexibility versus complexity, size, and cost. In general, the complexity, and thus the cost, of wavelength add/drop modules increases with the number of wavelengths on the transmission fiber, the fraction of wavelengths that can be added/dropped, and the fraction of wavelengths from which one can chose to add/drop. While systems are designed with node-to-node connection capacity requirements in mind, the ability to accommodate unpredicted capacity needs is inevitable. Network operators need to balance that potential need against the initial increased cost to provide it.

As with electronic add/drops and cross connects, operators want operational ease and flexibility with ROADMs and optical cross connects. That has driven the requirement that it should not be necessary, as it was in first-generation optical add/drop multiplexers, to require the assignment of any particular spatial slot on add/drop frame to be used only for a specific transmitted wavelength. That is, there should be complete freedom to insert a transmitter line card of any wavelength into any available line-card slot. ROADMs that provide this capability are referred to as colorless ROADMs. To provide this flexibility and ease of use generally requires additional optical-switch fabric with its additional cost. Once the line card has been inserted, internal software controls this optical-switch fabric to achieve the desired connectivity.

A simple way to think about this is the example of a simple $M \times 1$ passive wavelength multiplexer, for example a waveguide-grating router, that multiplexes M different colors onto a fiber. To operate effectively the right color has to be launched into a particular input port on that multiplexer. Alternatively, if in front of the M input lines to that mux you have an $M \times M$ switch (or, perhaps, M $1 \times M$ switches), you could insert any of the M colors into the switch at any port and under switch control get it to the right port on the mux to work.

Similar operational advantages exist for placing no restrictions on line-card placement based upon the ultimate route direction or any internal blocking or

contention. Therefore, there has been considerable research and development over recent years to design and manufacture cost-effective colorless, directionless and contentionless optical ROADMs. These so-called CDC ROADMs offer flexibility and simplify operations in ways that carriers have come to expect based upon their experience with TDM (SONET) add/drop multiplexers (ADM). While they provide substantial operational flexibility, these CDC ROADMS do require substantial increase and complexity of the optical switch fabric. To achieve this capability cost-effectively will require continued innovation in ROADM architecture and WSS technology.

Another feature available in electrical digital add/drop multiplexers is the ability to do bandwidth defragmentation as well as bandwidth concatenation. Several approaches to implement this function in ROADM have been proposed, including the use of time-slotted WDM and add/drop elements that have adjustable bandwidth per wavelength. The latter enable wavelength channels to be dynamically allocated only the bandwidth needed for the bit rates to be carried. This would provide more efficient utilization of the bandwidth in networks for which the bandwidth demands between node pairs differs across the network, which, as noted above, is typically the case [1.2].

Overall, the value that reconfigurable WDM networks bring in the marketplace can be summarized [1.2]:

- They provide high-level (wavelength) bandwidth management at relatively low cost, power, and space requirements.
- They provide traffic on/off ramps to satisfy network connection demands while extending the value of the multiwavelength amplification via optical bypass of the pass-through wavelengths.
- They reduce network costs and power by eliminating O-E-O conversion and electrical regeneration.
- They offer the opportunity to simplify and reduce the costs of capacity upgrades.

The second point suggests that the cumulative benefit increases with network reach. That is moderated at some point, as the extra cost to increase the reach is greater than the O-E-O reduction savings. An example of the fourth point above is that a reconfigurable WDM network designed for 10 and 40 Gb/s and with reserve gain in the amplifier could be upgraded to 100 Gb/s by simply adding long-reach 100 Gb/s transmitters and receivers, and providing additional power to the optical amplifier suggests significant potential value for network operators [1.21]. This is especially the case for undersea optical networks. This capability is a natural and fortuitous consequence of the ability of the fiber amplifier to amplify arbitrarily high bit channels carried by the wavelengths. In addition, the performance of the optical switching elements used for add/drop and cross connects are, of course, independent of bit rate.

As was noted earlier, in general for optical networks, it is necessary that a significant number of the network node pairs have a capacity demand between them equal to or exceeding the information bandwidth to be carried by a wavelength. It is also important to map TDM channels into wavelength expeditiously. If either of these is not the case, insufficient wavelengths will not be fully packed when they depart a node and will have to be dropped and groomed electronically with additional traffic destined to the same endpoint to effectively utilize the bandwidth of the transmission links. In this situation, few wavelengths will effectively benefit from optical bypass, and the fundamental value proposition of optical networks is lost.

At the time of this writing, the commercial reconfigurable optical networks available and deployed for national and metro applications have capacities of ≈ 10 Tb/s (100–160 wavelengths at 100 Gb/s) with fully reconfigurable wavelength add/drop capability. Transoceanic commercial systems, which include undersea ROADMs, are operating at capacities of ≈ 4 Tb/s. Increasingly, the ROADMs in these networks provide some CDC capability.

1.8 Driving Ultrahigh-Speed WDM and Beyond

The ubiquitous deployment of broadband wireless systems together with massive sharing of consumer-produced video and the growing demand to access *cloud*-based computational services continues to drive bandwidth demand. The next generation of wireless, 5G, networks will drive the capacity demands on optical networks even further. There is every indication that demand will continue to grow $\approx 100\times$ over the next 10 years. Given the state of current commercial systems this suggests the need for 1 Pb/s per-fiber systems in the not-too-distant future.

As the practical implementation of wavelength networking was being pursued during this period, the quest to increase the capacity per wavelength also continued. The next goal beyond 40 Gb/s eventually became 100 Gb/s. The issues that needed to be ad-

dressed the implied 2.5 ×-increase in the bit rate were similar to those for earlier generations of TDM—higher-speed modulators, photodetectors, and, most importantly, electronics to operate at 100 Gb/s. As has happened repeatedly over the decades of evolution of optical systems, researchers addressed those issue directly but also looked to alternative approaches that would not require the 2.5 × increased electronics speed.

To achieve higher effective per-wavelength channel capacity beyond the 10 and 40 Gb/s level, while limiting the bit-rate requirement, research focused on advanced modulation techniques that depend upon encoding information on both amplitude and phase information as offered by coherent detection techniques. In addition, polarization multiplexing to double the per-wavelength capacity without increasing the bit rate has always been viewed as a potential option, but its implementation using direct-detection techniques had been problematic. Coherent detection techniques were explored extensively in the late 1980s and 1990s, especially as a means to achieve higher receiver sensitivity than what was achievable with direct detection. That advantage became much less important with the advent of the fiber amplifier. Coherent detection also requires mixing with a local oscillator that must be frequency and phase-locked to the received signal wavelength. The linewidth of early lasers made that impractical. Today, using lasers with reduced linewidth (≈ 1 GHz), advanced modulation formats, polarization multiplexing, and high-performance digital signal to processors (DSPs) to facilitate frequency-locking transmission systems with an effective information capacity of 100 Gb/s are achievable with electronic speeds of 25 Gb/s or less. The higher-order modulation formats require nested circuits of waveguide Mach–Zehnder modulators (MZM) and phase modulators to provide both amplitude and differential phase modulation. An added benefit of this approach is that the DSP can also be used to mitigate deleterious chromatic and polarization dispersion effects (Fig. 1.3). This solution to achieve 100 Gb/s per wavelength has been massively deployed in both national and metro WDM networks [1.1].

Research efforts continue to focus on both increasing the number of wavelengths and the bit rate per wavelength. Commercially, the next targeted bit rate is likely to be 400 Gb/s, potentially followed by 1 Tb/s. To achieve these higher speeds requires continued advancement in high-speed electronics, photodetectors, and modulators. It also requires the ability to launch higher optical power into the fiber without causing nonlinear effects. This is an area where new fibers maybe beneficial. The number of wavelength channels is limited by the required bandwidth per channel and the total transmission bandwidth, which is limited by the amplifier bandwidth. The goal is to optimize spectral efficiency. To leverage the value of optical amplifiers and to avoid electrical regeneration, the ability to achieve long-distance transmission is also important.

These modulation formats, including quadrature phase-shifted keying (QPSK) and quadrature amplitude modulation (QAM) require the use of high-speed digital signal processors to convert the input signal information to the coded amplitude and phase modulation signals to drive complex nested amplitude and phase optical modulators to encode the optical signal. As an example, with polarization multiplexing and 64-QAM (64 symbols per bit) one can transmit at an effective rate of 320 Gb/s with electronics, modulator and receiver operating at only 80 Gbaud/s [1.1]. Unfortunately, the benefits come with transmission tradeoffs, as well as the complexity, power requirements, and cost of high-speed digital signal processors.

1.9 Space-Division Multiplexing: The Next Era of Optical Networking

As has been a common theme throughout the evolution of optical networks, there is growing concern that to keep up with bandwidth demand another major technological leap—an additional dimension of multiplexing—will be needed. Arguably, the only remaining dimension is space. There has been keen interest and growing activity in using space division as implemented either via multiple cores in a fiber or multimodes of a single-core fiber (Fig. 1.6), see, for example, [1.22]. As we learned with WDM, for space-division multiplexed (SDM) transmission systems to be cost effective compared to simply building parallel fiber optic systems of the current type, it likely will be essential to also demonstrate optical amplifiers that provide uniform gain simultaneously for all spatial modes of a single-core fiber or multicore transmission medium. There has been extensive activity over the past nearly 10 years to demonstrate suitable multicore or multimode fibers, means to selectively couple into the spatial modes, as well as suitable optical amplifiers. As an example of recent research progress in this area, a 10 Pb/s SDM/WDM transmission demonstration over limited distance (11 km) was achieved using a 6-mode 19-core fiber and a C_L band optical amplifier [1.23].

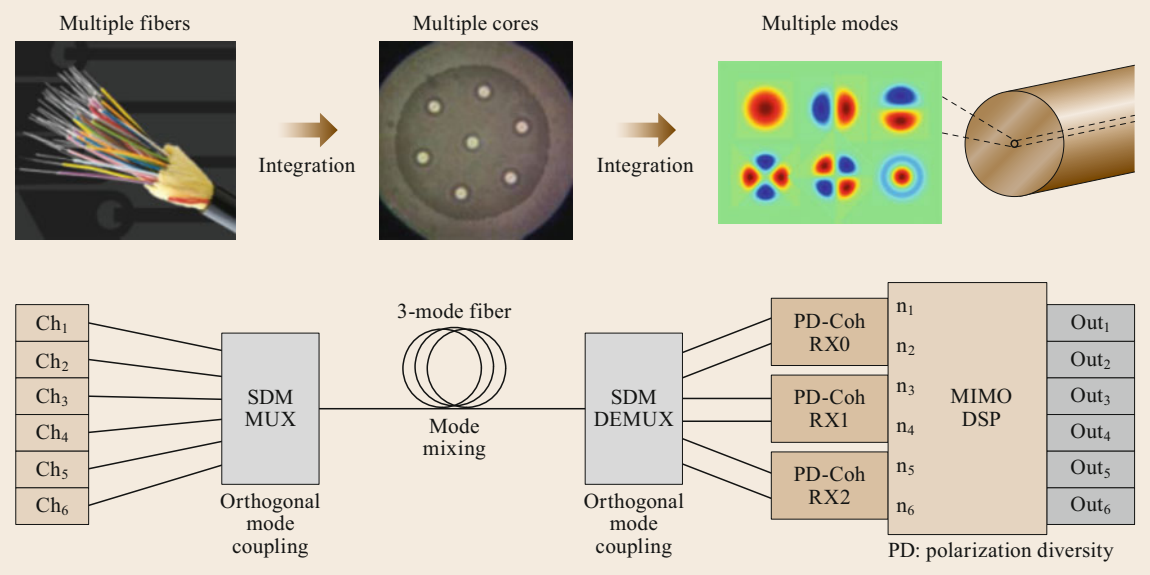


Fig. 1.6 Optical space-division multiplexed transmission. Adapted from [1.22], © 2012 IEEE

As we saw in the case of WDM, in addition to a cost-effective multiwavelength, multispatial mode amplifier, to reap the full benefit of a reconfigurable network to manage bandwidth at the optical wavelength and spatial modes layer, components and modules that would allow reconfigurable add/drops and cross connection with respect to spatial modes, as well as wavelength, would also be required. Ultimately, one would like these to work in a hierarchical manner with channels defined by both wavelength and spatial mode to achieve the best granularity of bandwidth management. It is also useful to note that the spatial modes of a multimode fiber may have some level of coupling over typical transmission distances. While these coupling effects can be mitigated with MIMO-like electronic processing, see for example [1.24], any need to do so would require demultiplexing other spatial modes and would result in a reduced ability to reap the full benefit of optical by-pass. While work remains to explore and demonstrate these additional add/drop and cross-connect functions and to demonstrate the practical, cost-effective value of SDM for flexible optical networks, given the continuing growth in bandwidth demand, the potential for a factor of 10 or more in network bandwidth offers strong potential for opening the next era of optical communications [1.4].

1.10 Optical Networking Applications Beyond Global Transport Networks

We have focused here on the role of optics in long-haul and metro transport networks because to date they have had the largest deployment and have clearly demonstrated the value of not only optical transmission but also reconfigurable networks for cost-effective, very high capacity networks. As discussed elsewhere in this book, there is growing application of optics in residential and small business access, wireless backhaul, and last mile to residential consumers. In addition, much of the cable distribution network up to the last link to the home is already fiber-based, and fiber-to-the-home continues to grow. Researchers are tapping into the massive deployment of LEDs for lighting in homes and businesses as a platform for optical-based LANS, so-called *LI-Fi*, which leverage the ability of the light output of LEDs to be modulated at relatively low speeds.

However, perhaps the biggest opportunity for optics to show its value for high-capacity interconnectivity is in the rapidly growing number of data centers around the US and the world, which increasingly provide the computing power that underpins the e-services that consumers, companies, and governments depend upon in their daily lives. Of course, these data centers are only viable because of the *cloud* provided by the global transport networks described above and the data network that runs over it. These data centers are home

to literally hundreds of thousands of servers that need to be interconnected where the total internal information capacity flow is substantially larger than that coming in or going out of the data center. Low-cost optical transmitters and receivers will be essential to provide the required massive low-power connectivity. Given the limited distances, the role of optical amplifiers and WDM needs to be more fully demonstrated. Once that has been established, the value of reconfiguring optical interconnectivity, even if initially only relatively infrequently, as is the case for the optical cross connect in transport networks. Time-slotted interconnection leveraging high-speed optical switching is also being explored to rapidly set up and take down high-capacity interconnections pipes, see, for example, [1.2].

References

1.1 P.J. Winzer, D. Nielsen: From scaling disparities to integrated parallelism, J. Lightwave Technol. **35**(5), 1099–1115 (2017)

1.2 R.C. Alferness: The evolution of configurable wavelength multiplexed optical networks: a historical perspective, Proc. IEEE **100**(5), 1023–1034 (2012)

1.3 I. Jacobs: Lightwave systems development: looking back and ahead, Opt. Photonics News **6**(2), 19–23 (1995)

1.4 P.J. Winzer, D.T. Nielsen, A.R. Chraplyvy: Fiber-optic transmission and networking: the previous 20 years and the next 20 years, Opt. Express **26**(18), 24190–24239 (2018)

1.5 S.K. Korotky, G. Eisenstein, R.C. Alferness, J. Veselka, L.L. Buhl, G.T. Harvey, P.H. Read: Fully connectorized high-speed Ti:$LiNbO_3$ switch/modulator for time-division multiplexing and data encoding, J. Lightwave Technol. (1985), https://doi.org/10.1109/JLT.1985.1074169

1.6 S.K. Korotky, G. Eisenstein, A. Gnauck, B. Kasper, J. Veselka, R. Alferness, L. Buhl, C. Burrus, T. Huo, L. Stulz, K. Nelson, L. Cohen, R. Dawson, J. Campbell: 4-Gb/s transmission over 117 km of optical fiber using a Ti:LiN60z external modulator, J. Lightwave Technol. **3**(5), 1027–1031 (1985)

1.7 J.C. Campbell, A.G. Dentai, W.S. Holden, B.L. Kasper: High-performance avalanche photodiode with separate absorption, "grading", and multiplication regions, Electron. Lett. **19**(20), 818–820 (1983)

1.8 E. Desurvire, J.R. Simpson, P.C. Becker: High-gain erbium-doped traveling wave, Opt. Lett. **12**(11), 888–890 (1987)

1.9 A.R. Chraplyvy, R.W. Trach, K.L. Walker: Optical fiber for wavelength division multiplexing, U.S. Patent 5327516 (1994)

1.10 L.A. Coldren: Multi-section tunable laser with differing multi-element mirrors, U.S. Patent 4896325 (1990), filed 1988

1.11 V. Jayaraman, Z.M. Chuang, L.A. Coldren: Theory, design and performance of extended tuning range semiconductor lasers with sampled gratings, IEEE J. Quantum Electron. **29**(6), 1824–1834 (1993)

1.12 E.J. Murphy, T.O. Murphy, A.F. Andrews, R.W. Irvin, B.H. Lee, P. Peng, G.W. Richards, A. Yorkinks: 16×16 strictly non-blocking guided wave optical switching systems, J. Lightwave Syst. Tech. **14**, 352–358 (1996)

1.13 R.C. Alferness: Guided wave devices for optical communications, IEEE J. Quantum Electron. **17**(6), 946–959 (1981)

1.14 R.C. Alferness, R.V. Schmidt: Tunable optical waveguide directional coupler filter, Appl. Phys. Lett. **33**(2), 161–163 (1978)

1.15 S.B. Alexander, R.S. Bondurant, D. Byrne, V.W.S. Chan, S.G. Finn, R. Gallager, B.S. Glance, H.A. Haus, P. Humblet, R. Jain, I.P. Kaminow, M. Karol, R.S. Kennedy, A. Kirby, H.Q. Le, A.A.M. Saleh, B.A. Schofield, J.H. Shapiro, N.K. Shankaranarayanan, R.E. Thomas, R.C. Williamson, R.W. Wilson: A precompetitive consortium on wide-band all optical networks, J. Lightwave Technol. **11**(5), 714–735 (1993)

1.16 R.E. Wagner, R.C. Alferness, A.A.M. Saleh, M.S. Goodman: MONET: Multiwavelength optical networking, J. Lightwave Technol. **14**(6), 1349–1355 (1996)

1.17 R.C. Alferness, J.E. Bethold, D. Pomey, R.W. Tkach: MONET: New Jersey demonstration network results. In: *Opt. Fiber Conf., February 16, Dallas* (1997), paper WI1

1.18 W.T. Anderson, J. Jackel, G.-K. Chang, H. Dai, W. Xin, M. Goodman, C. Allyn, M. Alvarez, O. Clarke, A. Gottlieb, F. Kleytman, J. Morreale, V. Nichols, A. Tzathas, R. Vora, L. Mercer, H. Dardy, E. Renaud, L. Williard, J. Perreault, R. McFarland, T. Gibbons: The MONET Project – a final report, J. Lightwave Technol. **18**(12), 1988–2009 (2000)

1.19 D.M. Marom, D.T. Neilson, D.S. Greywall, N.R. Basavanhally, P.R. Kolodner, Y.L. Low, F. Pardo, C.A. Bolle, S. Chandrasekhar, L. Buhl, C.R. Giles, S.-H. Oh, C.S. Pai, K. Werder, H.T. Soh, G.R. Bogart, E. Ferry, F.P. Klemens, K. Teffeau, J.F. Miner, S. Rogers, J.E. Bower, R.C. Keller, W. Mansfield: Wavelength selective 1x4 switch for 128 WDM channels at 50 GHz spacing. In: *Proc. Optical Fiber Comm. Conf. (OFC)* (2002) p. 857

1.20 B. Collings: New devices enabling software – defined optical networks, IEEE Commun. Mag. **51**(3), 66–71 (2013)

1.21 P.J. Winzer, G. Raybon, H. Song, A. Adamiecki, S. Corteselli, A.H. Gnauck, D.A. Fishman, C.R. Doerr, S. Chandrasekhar, L.L. Buhl, T.J. Xia, G. Wellbrock, W. Lee, B. Basch, T. Kawanishi, K. Higuma, Y. Painchaud: 100-Gb/s DQPSK transmission: from

laboratory experiments to field trials, J. Lightwave Technol. **26**, 3388–3402 (2008)
1.22 P.J. Winzer: Optical networking beyond WDM, IEEE Photonics J. **4**(2), 647–650 (2012)
1.23 D. Soma, Y. Wakayama, S. Beppu, S. Sumita, T. Tsuritani, T. Hayashi, T. Nagashima, M. Suzuki, H. Takahashi, K. Igarashi, I. Morita, M. Suzuki: 10.16 Peta-bit/s dense SDM/WDM transmission over low-DMD 6-mode 19-core fibre across C+L Band. In: *Proc. Eur. Conf. Opt. Comm. (ECOC)* (2017), Th.PDP.A1
1.24 S. Randel, R. Ryf, A. Sierra, P.J. Winzer, A.H. Gnauck, C.A. Bolle, R.-J. Essiambre, D.W. Peckham, A. McCurdy, R. Lingle Jr.: 6×56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6×6 MIMO equalization, Opt. Express **19**(17), 16697–16707 (2011)

Rod C. Alferness

University of California Santa Barbara
Santa Barbara, CA, USA
alferness@ucsb.edu

Rod C. Alferness is the Richard A. Auhll Professor and Dean of the College of Engineering at the University of California, Santa Barbara. He received his BS in Physics and Math (1968) from Hamline University and his MS and PhD in Physics from the University of Michigan. He has authored over 100 papers and 5 book chapters and holds over 30 issued patents. He has given numerous plenary and invited lectures.

Optical Su

Part A

Part A Optical Subsystems for Transmission and Switching

Ed. by Peter J. Winzer

The Internet, with its myriad of applications and services that fuel today's communication and information society, is used on an everyday basis by more than half of the world's population, and this number continues to grow. Despite its deep penetration, few people are aware that *fiber-optic networks* are at the heart of the Internet. In fact, without its utilization of highly cost-efficient and energy-efficient fiber-optic transmission and switching technologies, the Internet would not have been able to grow the way it has to become a ubiquitous and universally accessible resource serving the global population. For instance, global optical fiber deployments have exceeded 4 billion km, meaning that the resulting fibers could be wrapped around the globe approximately 100 000 times. Using infrared light, more than three exabits (one exabit $= 10^{18}$, bits) of traffic are transported per second through these fibers between users, datacenters, machines, and (increasingly) things. Essentially every bit of information that is transmitted beyond the confines of a home or enterprise, and virtually every phone call (whether placed from a fixed or wireless telephone), spends part of its life as a photon in an optical fiber. How would global societies have evolved without this amazing fiber-optic infrastructure?

Chapter 2 of this Handbook examines the evolution of optical fibers, from Charles Kao's prediction of low-loss optical fiber in 1966—for which he was awarded the 2009 Nobel Prize in Physics—to the state-of-the-art commercial telecommunications fibers that are widely deployed today. The chapter presents key performance characteristics of optical fibers in a communication systems context and discusses the outlook for fiber technologies that are currently being explored within the fiber-optic research community, such as photonic bandgap fibers, multicore fibers, and few-mode fibers. The chapter also covers optical fiber cabling technologies and deployment-related aspects, thus providing a comprehensive account of this key medium for information transmission.

Despite the low loss afforded by modern optical fibers, the maximum achievable transmission distance between two fiber-optic transponders remains loss limited to about 100 km. Since an optical fiber supports the simultaneous transmission of up to a few hundred signals at slightly different wavelengths (wavelength-division multiplexing, WDM), up to a few hundred optical transponders would be needed every 100 km were it not for optical amplification techniques. These permit the simultaneous low-noise amplification of all wavelength signal channels by a single, highly efficient amplification device inserted into the link about every 100 km, enabling transcontinental transmission distances of several thousand km and even transoceanic links exceeding 10 000 km. Chapter 3 focuses on the two most commercially important optical amplification technologies: the erbium-doped fiber amplifier (EDFA) and Raman amplification. A brief overview of other amplification techniques, such as semiconductor optical amplifiers and parametric amplification, is also presented.

In order to actually transmit optical signals, transponders are needed to encode light waves with digital information through modulation. In its simplest form, modulation is accomplished by switching the light on and off, but modern optical transponders also use the optical phase and polarization of the light wave to convey information more efficiently, allowing WDM systems to reach about a hundred terabits per second (one terabit $= 10^{12}$ bits) over a single conventional optical fiber. Chapter 4 discusses various transponder technologies, ranging from short-reach client interfaces that connect servers to nearby Internet routers (for example) to long-haul line interfaces that are able to send signals across the Pacific.

The key components of an optical transponder are reviewed in Chapter 5, including lasers, optical modulators of various kinds, and photodetectors to convert received optical signals back into electrical signals. The chapter also shows how those components are assembled into transponder subsystems to enable the modulation performance addressed in Chapter 4. Typical component and subsystem characteristics are discussed for state-of-the-art commercial products as well as recently demonstrated research systems.

Enormous advances in integrated electronic chips have enabled optical transponders to perform the complex digital signal processing (DSP) functions that are needed to implement the modulation formats discussed in Chapter 4. Chapter 6 focuses on those DSP functions, showing the internal workings of modern coherent optical transponders, how they mitigate optical signal impairments, and how they adaptively equalize the optical fiber channel to achieve transmission performance close to the fundamental limits of the fiber channel.

In addition to DSP, forward error correction (FEC) is a key enabler of modern optical transponders, and FEC is discussed in Chapter 7. Modern error-correcting codes allow full bit stream correction even when several percent of the received signal is erroneous bits. In combination with modulation and DSP, coding allows transponders to achieve their optimal transmission performance.

The hundred (or even hundreds of) WDM signals supported by optical fibers and amplification technologies carry more traffic than needed for today's point-to-point demands. Individual wavelengths or groups of

wavelengths are therefore assigned to each point-to-point connection within a network. In order to avoid the need for costly transponders to reroute wavelengths within the network, optical switching nodes are employed to route individual wavelength signals to the desired fiber in a network node as well as to add and drop wavelengths at local transponders at their final destination. Chapter 8 discusses the architecture and implementation of optical switching nodes.

The optical field that propagates within an optical fiber has a diameter that is on the order of 10 μm. This means that the optical intensity within the fiber core approaches the MW/cm^2 level, even for individual WDM signal powers of only 1 mW. At such high intensities, the refractive index of glass varies with optical power, leading to power-dependent optical phase shifts along the fiber, which accumulate over the thousands of kilometers propagated and produce severe signal distortions, especially when coupled with chromatic dispersion. Chapter 9 focuses on the effect of fiber Kerr nonlinearity, which has limited fiber-optic communication systems since the invention of optical amplifiers. The chapter presents analytical and numerical models of fiber nonlinearities as well as the means to combat their performance-degrading effects in the context of modern optical transmission systems, and it describes the limits of nonlinear transmission performance.

Conventional optical fibers support a fundamentally limited information capacity, which practical systems are already approaching. Therefore, as network traffic continues to increase, fiber-optic networks are nearing a 'capacity crunch.' The only way to scale networks beyond the limits of conventional fibers is to use spatial parallelism (space-division multiplexing, SDM) techniques. Such approaches range from employing bundles of conventional fibers for parallel transmission all the way to using individual transverse modes of few-mode fibers as parallel channels. The latter leads to crosstalk between those parallel paths, meaning that DSP techniques such as multiple-input multiple-output (MIMO) are needed to combat crosstalk. Chapter 10 discusses SDM techniques that are at the forefront of fiber-optic transmission and network research.

2. Optical Fiber and Cables

Takashi Sasaki, Takemi Hasegawa, Hiroki Ishikawa

This chapter gives an overview and introduces application scenarios for optical fibers and cables in optical communications. The use of single-mode optical fibers for both short-reach and long-haul applications is growing due to continually increasing demand for higher bandwidth optical communication systems. To better understand fiber-based optical communications the chapter first focuses on the design of the single-mode fiber while the latter half focuses on the design of optical fiber cable. A wide variety of optical fiber cables have been designed and installed to meet the needs of various applications and this chapter reviews the many types of cables for fiber to the home as well as for datacenter connectivity.

We will start with a section showing the history of optical loss improvement. Then, the categories of optical fibers and their cross-sectional structure are explained. Next, the main features of single-mode fibers such as standard single-mode fiber (SSMF), bend-insensitive fiber (BIF), cutoff-shifted fiber (CSF), dispersion-shifted fiber (DSF), and nonzero dispersion-shifted fiber (NZDSF) are summarized with their major optical characteristics. Then, features of each fiber type are explained. In particular, the cutting-edge application of low-loss fiber in ultralong-haul systems such as subsea systems are discussed and the importance of their system impact is also described. Thereafter, the key characteristics of multimode fiber are explained. In the final part of the optical fiber section, emerging fiber types are introduced such as fiber for space-division multiplexing (SDM) systems, for example multicore fiber, few-mode fiber, or coupled multicore fiber.

Thereafter, in the optical fiber cable section, we start with the classification of use cases such as indoor or outdoor cables and their features. Next, we introduce the optical fiber unit, a basic element used to bundle the fiber into cable, such as an optical fiber ribbon or loose tube. Following this we present many examples of optical fiber cables and their features, such as the slotted-rod cable, loose-tube cable, central-tube cable, layered fiber core cable, and direct-jacketed cable. Next, we present key considerations in optical cable design, such as fiber density, environment conditions, temperature change, water durability, biological attacks, and mechanical durability (bend, impact, torsion, crush). Finally, specific fiber cable use cases such as air-blown fiber cables and ultrahigh-density fiber cables for datacenter applications are introduced.

For more information about the basics of optical fibers and cables, we refer the reader to text books such as [2.1, 2]

1. D. Marcuse: Theory of Dielectric Optical Waveguides, Academic, New York, (1974)
2. Barry Elliott, Mike Gilmore: Fiber Optic Cabling, Elsevier Science, Netherlands, (2002)

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_2

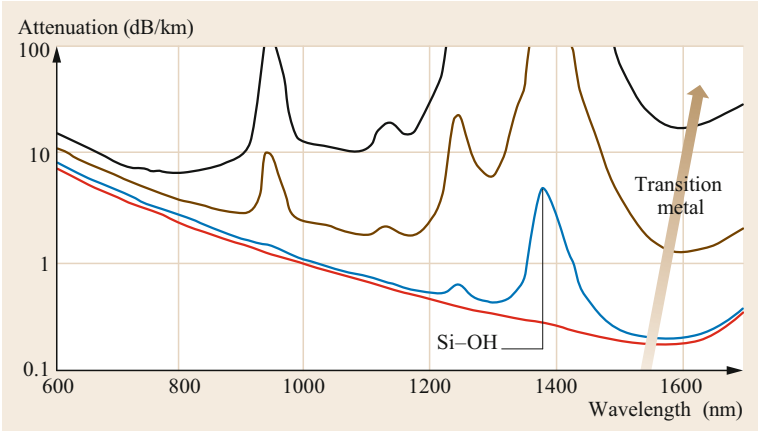


Fig. 2.1 History of loss improvement in optical fiber

Optical fibers comprise an indispensable part of the network infrastructure required to support current data services like the internet and wireless communication systems. Silica-based optical fiber is a unique transparent medium that is readily manufacturable and reliable, and both of these characteristics are required to enable global terrestrial and subsea networks to be interconnected.

In 1966, Dr. Kao at Standard Telecommunication Laboratories in the UK predicted that silica-based optical fiber would function as a low-loss transmission media equal to or less than 20 dB/km [2.3]. In 1970, Corning Glass Works demonstrated a low-loss (less than 20 dB/km at 630 nm) fiber that proved optical fiber could replace existing electric wired transmission networks [2.4]. Since then, as Dr. Kao forecasted, many enterprises worldwide have conducted comprehensive research to reduce the impurities (e.g., transition-metal and Si–OH absorption) and achieve significant improvements in regard to loss, as seen in Fig. 2.1.

In addition to the advancements in optical fibers, the steady performance improvement for semiconductor-based lasers, photo diodes, and modulators has also enabled optical communications networks to increase fiber bandwidth.

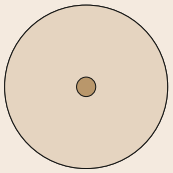
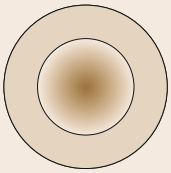
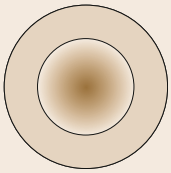
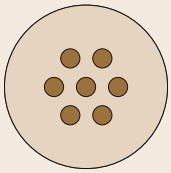
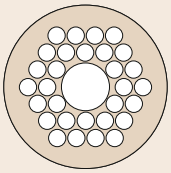
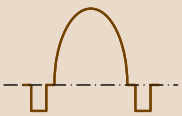
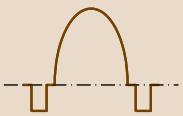
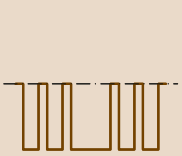
To replace traditional electric-wired transmission networks, optical fiber not only proved to be much lower loss, but it also had the advantages of wider bandwidth, lighter weight, smaller diameter, and no electrical inductance while experiencing no optical crosstalk between fibers. These features combined with silica's abundance and stability as material enabled optical fiber cables to significantly contribute to today's worldwide network connectivity and overall improvement to peoples' quality of life.

2.1 Transmission Fibers

2.1.1 Overview

As the geographical coverage and density of optical networks has increased, optical fiber links in data networks are growing in both quantity and diversity. The annual global supply of optical fibers was estimated to be approximately 511 million km in 2019 and growing by 10% annually [2.5]. Among the global supply, approximately 99% is single-mode fiber that transmits a single waveguide mode without impairment by modal dispersion (detailed in Sect. 2.2). The remaining 1% is multimode fiber, which transmits typically more than 100 waveguide modes, thereby facilitating low-cost optical coupling with devices, (detailed in Sect. 2.3). However, since the demand for capacity is expected to keep growing, a technical breakthrough in fiber will be necessary to meet the demand. Potential breakthroughs such as spatial-division multiplexing fiber and photonic band-gap fiber are described in Sect. 2.3. See Table 2.1 for the categories of transmission fibers and the sections containing corresponding descriptions.

Table 2.1 Categories of transmission fibers

	Conventional fibers		Emerging fibers		
	SMF	MMF	FMF	MCF	PBGF
Structure					
Refractive index profile (typical)					
Usage	General	Short reach	Spatial division multiplexing		Low latency
Sect.	2.2	2.3	2.4.1		2.4.2

2.2 Single-Mode Fibers (SMFs)

In this section, the general properties of SMFs, their classification (standardization) and application are described.

2.2.1 Common Properties

Single-mode fiber (SMF) is a glass fiber whose cross-section is normal to the length is composed of a circular core and an annular cladding surrounding the core. The refractive index of the core is higher than that of the cladding, so that the light wave is confined in the core and guided along the fiber length. Whereas the diameter of the core varies among the subtypes of SMF, it is typically in the range 5−15 μm corresponding to the 1.3−1.6 μm wavelength of light to be guided. As a result, only a single fundamental mode per wavelength of an electromagnetic wave can be guided in the core. Note that the two polarization modes are counted as a *single* guided mode because the two polarization modes form a pair of degenerate modes due to the rotational symmetry of the optical fiber structure.

Since the number of modes that can be guided by a finite core area increases almost in proportion to the inverse square of the wavelength, single-mode guidance is enabled within the wavelengths longer than a threshold, called the cut-off wavelength. In the industry, two definitions of cut-off wavelength are commonly used [2.6, 7]. One is the fiber cut-off wavelength often denoted as λ_c, and the other is the cable cut-off wavelength often denoted as λ_{cc}. In either case, the cut-off wavelength is defined as the minimum wavelength at which any mode other than the fundamental mode is attenuated by more than 19.3 dB by the designated transmission medium. In the case of the fiber cut-off wavelength λ_c, the transmission medium is defined as a 2 m fiber wound for one turn around a 280 mm diameter spool. In the case of the cable cut-off wavelength λ_{cc}, it is defined as a 22 m cable whose central 20 m length is wound around a 280 mm diameter spool and both 1 m length ends wound for one turn around an 80 mm diameter spool.

The core and cladding of an SMF are composed of silicate glass (SiO_2), which enables low optical attenuation. They often contain small amounts of dopants such as germania (GeO_2) and fluorine (F) which change the refractive index. The industry-standard diameter of the cladding is typically 125 μm. Since such a bare thin glass fiber is too fragile to handle, it is usually covered by a polymer coating of approximately 250 μm diameter, which is also treated as a common industrial standard. Since even coated fibers are not mechanically robust enough for deployment, they are incorporated in a protective cable whose structure depends on the environment in which it is to be deployed (details are provided in Sect. 2.5).

Because the fiber is composed of silicate glass, optical fibers have a wavelength window of low attenuation, as shown in Fig. 2.2. Typically, the attenuation is lower than 0.5 dB/km over the 1.2−1.7 μm range, that is, equal to 73 THz in the frequency domain. Within this broad bandwidth, the conventional C-band (1530−1565 nm) is the most frequently used band especially for long-distance transmission because of the coincidence of low attenuation and the gain window of the erbium-doped fiber amplifier (EDFA). The long wavelength L-band (1565−1625 nm) is also becoming common for long-distance transmission because the EDFA can be optimized to cover this wave-

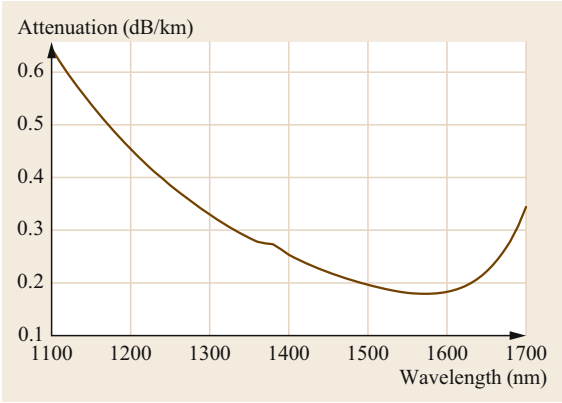


Fig. 2.2 Typical transmission loss spectra of current optical fiber

length. The original O-band (1260–1360 nm) is most frequently used for short-distance transmission without amplification because of low chromatic dispersion and low waveform distortion. The short wavelength S-band (1460–1530 nm) and extended E-band (1360–1460 nm) are also used for short-distance transmission. The above nomenclature of wavelength bands is standardized by the ITU-T (International Telecommunication Union Telecommunication Standardization Sector) [2.8].

2.2.2 Classification

In this section, classification of SMFs is described with industry standardization associated with their typical use case.

There are several types of SMF optimized for specific use cases such as fiber to the home (FTTH) and long-haul transmission. Choosing the best fiber for a specific application is important to achieve competitive cost-benefit and reliable networks. Conversely, some fiber types in use for terrestrial or subsea systems are becoming less popular for new deployment because they are optimized for legacy transmission technologies. For example, dispersion-shifted fiber, DSF (ITU-T G.653) suffers from nonlinear effects, and it has become suboptimal after the launch of wavelength-division multiplexing-transmission technologies for higher bandwidths. However, legacy fibers have been deployed in substantial volume, so that understanding the characteristics of those fibers is still important to make the most of the deployed fiber asset.

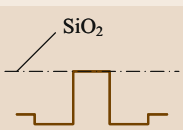
Table 2.2 summarizes the subtypes of SMF, their typical structure and characteristics. It should be noted that CSF has advantages of low attenuation, a large effective core area (A_{eff}), a low nonlinear refractive index (n_2) and large absolute chromatic dispersion, all of which improve the capacity of digital coherent transmission, as described in Sect. 2.2.6. A more detailed description of each subtype will be given in the following sections.

2.2.3 Standard Single-Mode Fiber

Standard single-mode fiber (SSMF) is the most often used fiber and often called just *single-mode fiber* (SMF). SSMF is typically composed of a core made of GeO_2-doped SiO_2 and a cladding made of pure SiO_2. Sometimes the inner part within the cladding is doped with fluorine and has a slightly lower index than the outer part, thereby improving bending loss performance. The optical and mechanical characteristics of SSMF are standardized by ITU-T G.652.D recommendations [2.9]. Compared to the other SMFs, SSMFs have the lowest cost because they are produced and employed in the largest volumes worldwide. They can be used generally for transmission across a wide wavelength range from 1.3 to 1.6 μm because the cable cut-off wavelength is shorter than 1.26 μm, and the extrinsic bending loss is low up to a wavelength of around 1.625 μm.

Mode-field diameter (MFD), a parameter representing how widely the optical energy is distributed laterally

Table 2.2 Characteristics of SMF by subtypes

	SSMF	BIF	CSF	DSF	NZDSF
ITU-T	G.652	G.657	G.654	G.653	G.655–656
Refractive index profile (typical)					
Typical attenuation (dB/km)	0.18–0.20	0.18–0.20	0.15–0.16	0.22	0.19–0.22
A_{eff} (μm^2)	80	80	80–153	46	55–72
$n_2 10^{-20}$ (m^2/W)	2.3	2.3	2.2	2.5	2.5
CD (ps/(nm km))	17	17	18.5–20.9	0	4–8
Usage	General	FTTX	Submarine long haul	–	Long haul

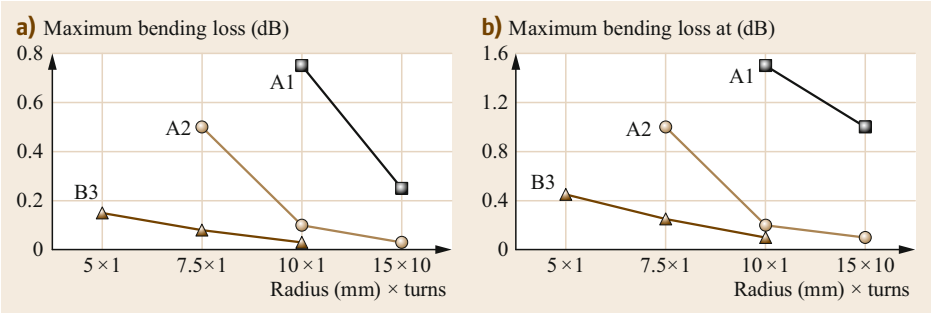


Fig. 2.3a,b Maximum bending losses of bend-insensitive fibers specified by ITU-T G.657: **(a)** at 1550 nm; **(b)** at 1625 nm

across the core [2.7], is standardized by G.652.D to be between 8.6 and 9.2 μm with a nominal tolerance of less than 0.4 μm. In that range, most of the commercial SSMFs are specified to have a 9.2 μm nominal MFD [2.8–12]. A large MFD has the advantage of lower coupling losses when coupled to another fiber with the same MFD because coupling losses due to lateral misalignment are reduced in proportion to the inverse square of MFD [2.13]. However, it should be noted that a larger MFD can cause a high coupling loss in waveguide devices or lasers, which typically have a much smaller MFD than SSMF [2.13].

2.2.4 Bend-Insensitive Fiber (BIF)

Bend-insensitive fiber (BIF), sometimes also referred to as bend-resistant fiber (BRF), is the second most frequently used fiber next to SSMF. Compared to SSMF, BIF has a lower bending loss that meets one of the ITU-T G.657 recommendations [2.14]. G.657 specifies the subclasses A1, A2 and B3, having lower bending losses in that order (B3 is the lowest), as summarized in Fig. 2.3. In terms of use, G.657.A1 is the most frequently used, and G.657.A2 is the next. These fiber types are mainly used for applications in fiber to the home (FTTH) or to the premise (FTTP), sometimes also called FTTX. In any of these applications, fibers are deployed near to the user and can be exposed to tight bends and localized loss because they are integrated into thin cables that are deployed in narrow pathways, and users often handle the cables harshly.

G.657.A1 fits for usage in a bending radius larger than 10 mm. Some G.657.A1 products have 9.2 μm MFDs compatible with typical G.652.D products [2.15, 16], whereas others have slightly smaller 8.9–9.0 μm MFDs [2.10, 17].

G.657.A2 fits for usage in a bending radius larger than 7.5 mm. Most G.657.A2 products have 8.6–8.8 μm MFDs, still in the range of G.652.D recommendations [2.10, 18–20].

G.657.B3 has the most advanced bend insensitivity and fits for usage in a bending radius larger than 5 mm. Like G.657.A2, most G.657.B3 products have 8.6–8.8 μm MFDs in the range of G.652.D recommendations [2.21–23].

As indicated by the MFDs of G.657 subclasses, reducing MFD and confining the light in the core is one of the effective ways to reduce bending loss of the fundamental mode, while enhancing the bending losses of higher order modes so that single-mode operation is maintained. The trade-off among bending loss, higher order mode cut-off, and mode-field diameter is described by the MAC factor

$$\mathrm{MAC} = \frac{\mathrm{MFD}}{\lambda_c}\,, \tag{2.1}$$

where it is empirically known that bending loss decreases as the MAC factor decreases [2.24].

Another way to reduce bending loss of the guided fundamental mode is to reduce the refractive index of a selected annular region in the cladding, which is often referred to as a trench [2.25, 26]. Such a trench can be made by synthesizing SiO_2 glass, including a large amount of fluorine, for example, using the plasma-enhanced chemical-vapor deposition (PCVD) technique [2.26].

Incorporating voids or air holes in SiO_2 fiber glass is another effective way to substantially reduce refractive index because the relative refractive index between SiO_2 glass ($n = 1.45$) and air ($n = 1$) is as large as 26%, in contrast to less than the 1% that is available by practical fluorine doping. Hole-assisted fiber (HAF), having several air holes of micrometer-order scale, is known to have very low bending losses [2.25, 27]. Further, optimized bending loss performance can be realized by using an effective trench made by a plurality of voids having tens or hundreds nanometer-order sizes [2.28].

2.2.5 Reduced Outer-Diameter Fiber

Whereas the outer diameter of coated optical fibers has typically been 250 μm, reduced-diameter fibers having an approximately 200 μm diameter are becoming commercially available. Their glass optical properties

are almost the same as those of one of the typical 250 μm products and comply with G.657.A1 [2.29, 30] or A2 [2.31, 32] recommendations. Also, their core glass diameter is kept unchanged at 125 μm, so only the outer coating thickness has to be reduced. Even with reduced thickness, the coating has to protect the inside glass fiber from breakage or microbending loss even under external forces in actual deployment conditions.

Microbending loss is an extrinsic loss caused by longitudinally distributed microbends of the glass waveguide that couples the fundamental mode to higher order modes [2.1, 33]. Those microbends are the outcome of the combination between external disturbances on the outer surface of the coated fiber and the mechanical characteristics of the coating that diminish the disturbances felt by the internal glass fiber. Therefore, the thicknesses and elastic moduli of the layers composing the coating should be designed optimally to keep low microbending loss [2.34, 35]. In addition, optimized waveguide designs, such as a trench-assisted structure [2.26], are also effective in suppressing microbending loss [2.36]. Either by improved coating or glass, 200 μm reduced-diameter fibers have been successfully demonstrated to exhibit comparable optical and mechanical performances compared to the conventional-diameter fibers [2.37].

2.2.6 Cutoff-Shifted Fiber (CSF)

Cutoff-shifted fiber (CSF) is a single-mode fiber that has an upper limit of cutoff wavelength equal to 1530 nm, shifted from the 1260 nm of SSMF. Although the longer cutoff wavelength reduces the wavelength range of operation, it enables an increase in the effective core area (A_{eff}) [2.38], which reduces the density of guided optical power and hence nonlinear optical effects. Such an increase in A_{eff} is enabled because larger core diameter accompanies longer cutoff wavelength. It can be understood if one assumes a simple matched-cladding step index profile, the theoretical cutoff wavelength λ_{ct} is given by

$$\lambda_{\text{ct}} = \frac{2\pi a}{2.405} n_1 \sqrt{2\Delta}\,, \tag{2.2}$$

where a is the core radius, n_1 is the core refractive index, and Δ is the refractive index difference between core to cladding.

Such low nonlinearity allows higher signal power to be launched into the fiber and thus benefits longer distance transmission by reducing the number of amplifiers and/or enhancing the signal power at the receiver. As a result, CSF is optimal for submarine and long-haul terrestrial transmission using the C-band and/or L-band, where erbium-doped fiber amplifiers can be used as repeaters. Moreover, since digital coherent transmission is widely adopted and its performance strongly depends on the optical signal-to-noise ratio (OSNR), CSFs are becoming even more important because of their low nonlinearity and low attenuation.

CSFs are usually composed of a pure silica (SiO_2) core and a fluorine-doped cladding. The pure silica core, in contrast to a relatively lossy GeO_2-doped silica core, enables low attenuation and low nonlinear refractive index. Typical refractive index structures of CSFs have a W-shaped depressed cladding or a trenched cladding [2.39–41], to achieve a suitable balance between large effective area, low bending loss of the fundamental mode, and a short cut-off wavelength. An intuitive explanation is that higher order modes have a longer tailed electric field encroaching into the outermost elevated cladding and are influenced by W-shaped or trenched cladding more than by the guided fundamental mode. As a result, a W-shaped or trenched cladding can reduce the bending loss of the fundamental mode relative to the higher order modes.

In addition to the cladding structure, a ring-shaped core is also used for some types of CSF [2.42, 43]. A ring core causes the intensity distribution of the fundamental mode to deviate slightly from that of a Gaussian distribution. As a result, if one keeps a constant MFD, a ring core has a larger effective area than a simple-step index core, as shown in Fig. 2.4. Since an increase in MFD for a CSF causes a higher loss when it is spliced to the SSMF typically used in repeater interfaces, the ring core provides an effective means to enlarge the effective area without a large penalty in terms of increased splice losses.

As the effective area gets larger (and its electromagnetic field gets spread more broadly), the fibers become more sensitive to microbending loss. Hence it is important when enlarging the effective area to improve the fiber coatings that isolate the glass waveguide from the

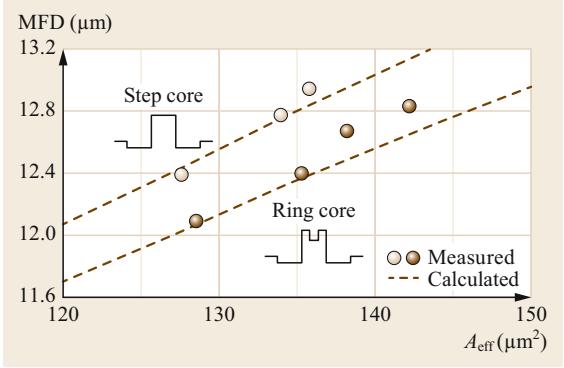


Fig. 2.4 The effect of a ring-shaped core to expand A_{eff} with the same MFD (at 1550 nm)

Table 2.3 ITU-T G.654 by subclass

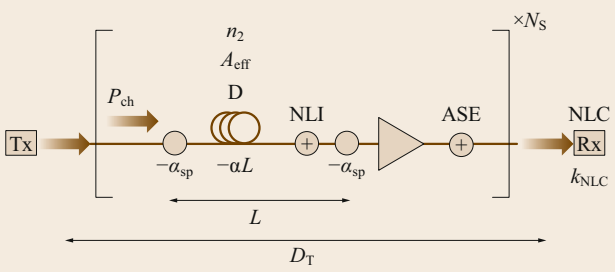
ITU-T	G.654.B	G.654.C	G.654.D	G.654.E
MFD at 1550 nm (nominal) (μm)	9.5–13	9.5–10.5	11.5–15.0	11.5–12.5
Attenuation at 1550 nm (cable) (dB/km)	≤ 0.22	≤ 0.22	≤ 0.20	≤ 0.23
CD at 1550 nm (ps/(nm km))	≤ 22	≤ 20	≤ 23	17–23
Bending loss $r = 30\,\text{mm} \times 100$ turns (dB)	≤ 0.50	≤ 0.50	≤ 0.50	≤ 0.1
Usage	Regional and middle-distance	Submarine	Long-distance submarine	Terrestrial long haul

external disturbances that cause microbending and as a result, an increase of loss. A fiber coating is typically composed of an innermost soft layer, a middle hard layer, and an outermost colored thin hard layer, typically called the primary, secondary, and ink layers, respectively. The state-of-the-art CSFs have a primary coating with a reduced elastic modulus that enables an increase in effective area to $130\,\mu\text{m}^2$ [2.43] or $150\,\mu\text{m}^2$ [2.44] without compromising microbending loss performance.

The characteristics of CSFs are standardized by ITU-T G.654 recommendations [2.45], as summarized in Table 2.3. Among them, G.654.D is the most popular standard for long-distance submarine transmission to an extent of, or greater than, several thousand kilometers because its large MFD correlates to a large A_{eff}. The latest G.654.D products have effective areas of $130–153\,\mu\text{m}^2$, the largest among SMFs and attenuations of 0.152–0.157 dB/km, also the lowest among SMFs [2.46–48]. Because of such advanced characteristics for attenuation and A_{eff}, CSFs have been contributing to the *hero* transmission experiments in laboratories [2.49–51] and also in submarine applications [2.52].

Whereas submarine cable use has been the main application of G.654 fibers, a subclass G.654.E intended for long-haul terrestrial transmission was newly defined in 2016 to have a tighter MFD range and lower bending loss. Field trials are also underway to verify the performance of G.654.E fibers in real terrestrial transmission links in China [2.53, 54]. In this trial, the fibers from seven vendors were cabled, deployed, and used for transmitting 400 Gb/s signals. The results have shown stable performance even under varying environmental conditions.

The value of low attenuation and large effective-area transmission fiber can be expressed in terms of improvement in generalized OSNR, which includes nonlinear impairment as additive noise to the conventional amplified-spontaneous-emission (ASE) noise [2.55]. As the generalized OSNR increases, the theoretical upper limit of transmission capacity increases almost in proportion to its logarithm [2.56], and one can increase the actual capacity using advanced modulation format signals such as higher order quadrature-amplitude modulation (QAM) and/or probabilistic shaping [2.57]. In addition, digital coherent WDM transmission provides another benefit because nonlinear impairment noise can be approximated by an additive white Gaussian noise described by an analytical formula, the so-called Gaussian noise model (GN-model) [2.55]. Because of this property, the generalized OSNR can be approximated by a simple formula called the fiber figure-of-merit (FOM) [2.58–60]. Assuming a definition of symbols as



$$L_{\text{eff}} = \{1 - \exp[-(0.1\ \ln 10)\alpha L]\}/[(0.1\ \ln 10)\alpha]$$
$$\gamma = 2\pi\, n_2/(\lambda\, A_{\text{eff}})$$
$$F = h\nu B\,\text{NF}$$
$$D_{\text{T}} = N_{\text{S}} L$$

A_{eff}	Fiber effective core area (μm^2)
B	Reference band width of OSNR (Hz)
D	Fiber chromatic dispersion (ps/(nm km))
D_{T}	Transmission distance (km)
F	ASE coefficient (W)
h	Planck constant (J s)
L	Span length (km)
k_{NLC}	NLI mitigation coefficient
L_{eff}	Effective span length (km)
n_2	Fiber nonlinear refractive index (μm^2/GW)
NF	Amplifier noise figure (linear)
N_{S}	Number of spans
P_{ch}	Launched signal power per channel (W)
α	Fiber loss coefficient (dB/km)
α_{sp}	Dissimilar splice loss (dB)
γ	Fiber nonlinear coefficient ($\text{W}^{-1}\text{km}^{-1}$)
λ	Signal wavelength in vacuum (μm)
ν	Optical frequency of signal (Hz)

Fig. 2.5 Link model in fiber FOM theory

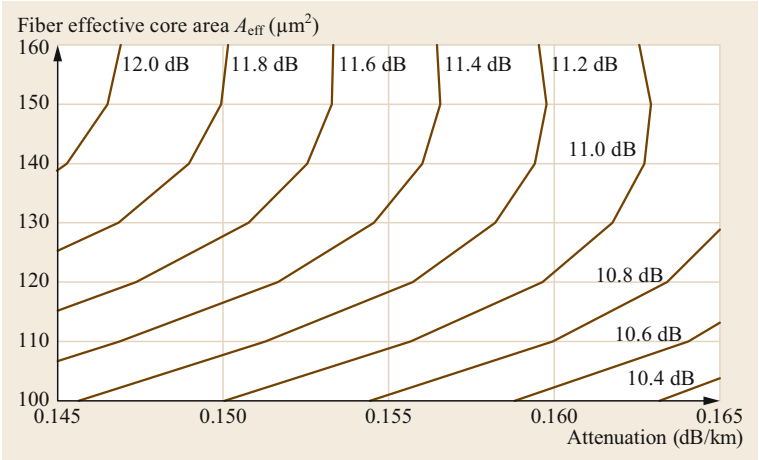


Fig. 2.6 Example FOM map on 80 km span system

shown in Fig. 2.5, the FOM is calculated as

$$\mathrm{FOM\ (dB)} = -\frac{10}{3}\log\left(\gamma^2 L_{\mathrm{eff}}|D|^{-1}\right) - \frac{2}{3}\alpha L - \frac{2}{3}\alpha_{\mathrm{sp}} + 10\log L\,, \qquad (2.3)$$

where the symbols are defined as dimensionless numbers defined by dividing the designated physical quantity by the designated unit. This formula gives an intuitive idea of the better fiber for digital coherent transmission systems. For example, if one assumes a repeated transmission system with a span length of 80 km and maximum signal power of −2 dBm/ch, typical for submarine transmission, the FOM can be calculated as a function of fiber attenuation and effective area and expressed as a contour map as shown in Fig. 2.6. On the basis of this kind of FOM map, one can intuitively understand the performance merits provided by using advanced fibers. Note that the exact shape of contours can vary depending on system assumptions such as span length, maximum signal power, and nonlinear compensation [2.60].

Last but not least, CSF is one of the most active fields of development in transmission fibers because CSFs with lower attenuation and nonlinearity are the most practical and effective way to expand the transmission capacity for the long-distance submarine cables which interconnect the world. In the most state-of-the-art CSF (Fig. 2.7), the attenuation at 1550 nm is 0.1424 dB/km and the lowest attenuation is 0.1419 dB/km at 1560 nm [2.44]. Such low losses are realized by reducing the Rayleigh scattering caused by microscopic inhomogeneity in the density of SiO_2 glass, and this can be ascribed to the manufacturing technology that has been continuously refined due to increasing demand since digital coherent transmission became the mainstream in long-distance submarine transmission. As a result, improvement in lowest attenuation has accelerated in recent years.

2.2.7 Dispersion-Shifted Fiber (DSF)

Dispersion-shifted fiber (DSF) is an SMF whose zero dispersion wavelength is shifted to near 1550 nm wavelength, from near 1310 nm typical of SSMF [2.61], as

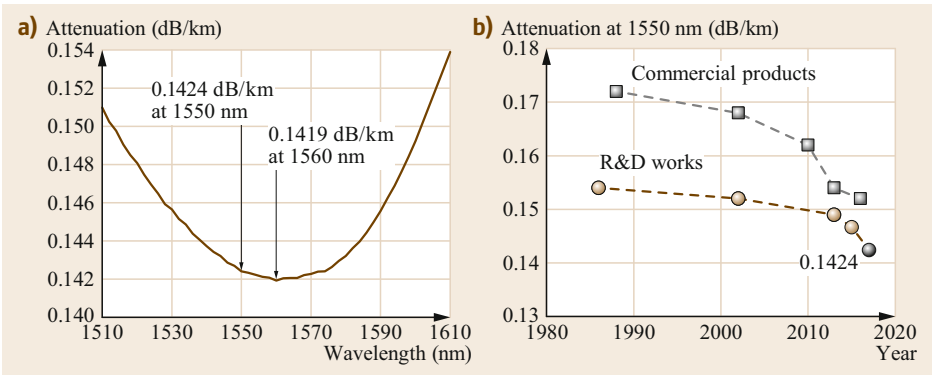


Fig. 2.7a,b The lowest attenuation fiber **(a)** and the history of reduction in attenuation **(b)**

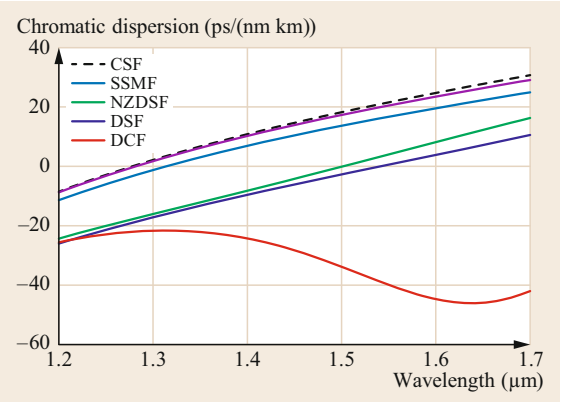


Fig. 2.8 Examples of chromatic dispersion characteristics for different materials

shown in Fig. 2.8. Such a shift in chromatic dispersion can be realized by increasing the refractive-index difference between the core and the cladding and decreasing the diameter of the core [2.62]. Since 1550 nm is the wavelength where the attenuation of optical fiber is the lowest, having the smallest chromatic dispersion at this wavelength could benefit transmission performance. However, in the case of wavelength-division multiplexed (WDM) transmission, the smaller chromatic dispersion generates crosstalk noise due to four-wave mixing, meaning that DSF is not usually suited for WDM transmission near the 1550 nm wavelength. For this reason, DSF is no longer actively deployed, but legacy fibers remain in the field.

2.2.8 Nonzero Dispersion-Shifted Fiber (NZDSF)

Nonzero dispersion-shifted fiber (NZDSF) is an SMF whose absolute chromatic dispersion over the C-band is smaller than that of SSMF but is not actually zero, so that it can support WDM transmission in the C-band. Whereas either positive or negative dispersion can support WDM, most products have positive chromatic dispersion in the C-band because getting positive dispersion requires less of a shift in dispersion than getting negative dispersion. Those dispersion characteristics can be tailored by using complex refractive-index profiles [2.63], but they also induce a smaller effective core area, or equivalently a higher nonlinearity. In addition, although the small dispersion reduces the need for dispersion compensation, it also increases the nonlinear impairment in WDM transmission [2.55]. Therefore, NZDSF has an advantage over other SMFs in cases where one must rely on costly optical-dispersion compensation solutions. Conversely, in cases where one can compensate dispersion electronically instead of optically, the merit of NZDSF decreases and the penalty of higher nonlinearity increases. The former was the case before the mid-2000s when long-distance transmission relied on intensity-modulated direct-detection (IM-DD) techniques, but after the mid-2000s the latter has become the case because digital coherent techniques are used in long-distance transmission.

Although NZDSFs are becoming a legacy for the above reasons, they have already been deployed widely and are still supplied by several vendors [2.64–67]. Among the various products, some have relatively large effective areas but large dispersion slope [2.64, 65], whereas others have small dispersion slopes but small effective areas [2.66], or intermediary characteristics between these extremes [2.67]. In contrast to the above-mentioned NZDSFs having positive dispersion, some NZDSF has negative chromatic dispersion [2.47] that suffers less from modulation instability even when high-power signals are launched for submarine transmission.

2.2.9 Dispersion-Compensating Fiber (DCF)

Dispersion-compensating fiber (DCF) is an SMF that has negative chromatic dispersion and dispersion slope. Since most transmission fibers, such as SSMF, BIF, CSF, and NZDSF have positive dispersion and dispersion slope, the overall amount of accumulated dispersion can be reduced over a wavelength range by joining a DCF with the transmission fiber [2.68]. In order to reduce the accumulated dispersion over a wide wavelength range, it is important to have a suitable relative dispersion slope (RDS), which is defined by normalizing the dispersion slope by the dispersion to be matched between the transmission fiber and DCF. Since RDS varies among transmission fibers such as SSMF and NZDSF, DCF is designed to compensate specific transmission fibers. There are two ways to use DCFs. One is a discrete DCF module where a length of DCF is coiled and packaged [2.69]. Another is a cabled DCF where DCF composes a span of a transmission link in combination with a matched positive-dispersion transmission fiber [2.47], such a transmission link is also called a dispersion-managed link.

Since the chromatic dispersion and dispersion slope of SiO_2 composing DCFs are both positive, a complex waveguide structure, such as a triple clad design, is used for DCF to realize the large negative dispersion and controlled dispersion slope [2.68]. However, such complex index profiles and high-density dopants also cause a slightly increased fiber attenuation and decreased effective area for a cable-type DCF [2.70]. The result is that a dispersion-managed link is now less often adopted for new deployment than positive-dispersion CSFs which perform better with digital coherent techniques.

2.3 Multimode Fibers (MMFs)

This section describes MMFs applied to short-reach transmission applications. Multimode fiber (MMF) is, in the context of transmission fiber, a glass fiber having a circular core and an annular cladding surrounding the core, where the refractive index of the core is higher than that of the cladding and has a graded profile expressed by αth power of radius. Typically, the diameter of the core is 50 μm and the core-cladding relative index difference is 1%. The αth power profile can suppress modal dispersion that is the difference in group delay among guided modes. The optimum α that minimizes modal dispersion is approximately 2 [2.71], but the precise value depends on the operating wavelength and the dopants in the SiO_2 core, which affects the material chromatic dispersion [2.72]. The MMF core is typically composed of GeO_2-doped SiO_2, whereas additional codoping of F can widen the wavelength range with low modal dispersion [2.72–74]. Regarding the cladding, an F-doped trenched cladding is widely used to provide a lower bending loss [2.75, 76].

MMFs are typically used for short-reach transmission at 850 nm wavelengths in combination with a vertical-cavity surface-emitting laser (VCSEL) transceiver, because of the high-energy efficiency [2.77] and cost efficiency [2.78, 79]. The maximum possible transmission distance, typically less than 2 km, is inversely proportional to the transmission bandwidth so that the transmission performance of MMFs are generally specified by a bandwidth–distance product. Whereas MMFs with precisely fabricated index profiles generally have larger bandwidth–distance products, the exact values of bandwidth–distance products can vary from VCSEL to VCSEL, which excite different populations of modes in the MMF. To cope with this problem, it has been agreed in the industry to specify the bandwidth–distance product of an MMF based on the measurement of the differential mode delay (DMD), a series of time-domain impulse responses by radially offset SMF launches [2.80] and assumption of several patterns of predetermined launching conditions. One widely used metric is its calculated minimum effective modal bandwidth (minEMBc), which is obtained by calculating virtual impulse responses to various assumed launching conditions and then taking the worst-case value [2.81, 82]. The second and traditionally used metric is the overfilled launch (OFL) modal bandwidth, which is defined as the bandwidth when all the guided modes are excited [2.81]. Conversely, OFL is fairly realistic when the signal is launched into MMF from a light-emitting diode (LED) [2.83]. It generally underestimates the performance when the MMF is excited by a VCSEL. OFL can be substituted by a calculated overfilled modal bandwidth (OMBc), which can be calculated from DMD measurements [2.81].

Among several equivalent standards of MMF, ISO/IEC 11801 and TIA-568.3-D are the most frequently [2.84, 85] used. MMFs are classified into four subclasses, OM1 to OM4. OM1 has a large 62.5 μm diameter core, which is optimal for LED rather than VCSEL signals, and has an OFL bandwidth of 500 MHz km at 1300 nm. OM2, OM3 and OM4 have a 50 μm core and have OFL bandwidths at 850 nm of 500, 1500 and 3500 MHz km, respectively. OM4 is backward compatible to the other two subclasses and is supplied by several vendors [2.86–88].

Whereas OM2 to OM4 specifies modal bandwidths only at 850 and 1300 nm, the growing need for short-wave wavelength-division multiplexing (SWDM) means a further advanced subclass of MMF that supports SWDM in the wavelength range between 850 and 950 nm is required. Such an advanced MMF, called a wideband MMF (WB-MMF) or OM5, is also becoming commercially available [2.89–91] and a state-of-the-art transmission experiment using four wavelength and four-level pulse amplitude modulation (PAM4) verified 206 Gb/s transmission over a distance of 150 m [2.92].

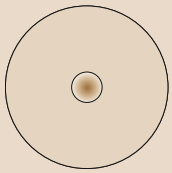
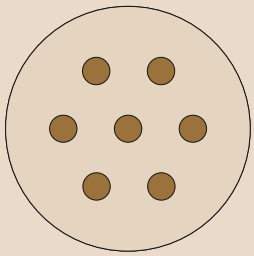
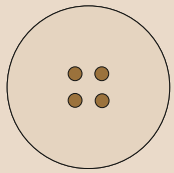
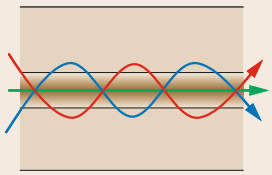
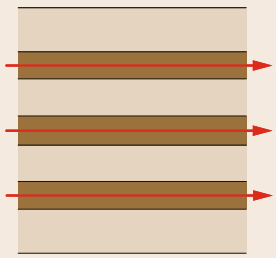
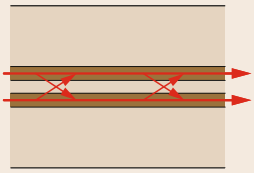
2.4 Emerging Fibers

This section describes new state-of-art fibers for future optical communication systems. There have been significant advances in transmission fibers with low attenuation and low nonlinearity that increase the theoretical capacity limit. However, the demand for capacity has expanded at a much faster rate, so that it is anticipated that we will soon face the challenge of an increasing gap between demand and supply [2.93].

2.4.1 Space-Division Multiplexing (SDM) Fiber

In the past, advances in transmission techniques, such as WDM and digital coherent transmission, contributed to significant expansions in capacity. However, the C and L bands supported by the EDFA are already fully used and leave almost no room for an increase in capacity by WDM, and digital coherent transmission using

Table 2.4 Comparison of SDM schemes

	FMF	U-MCF	C-MCF
Reference	[2.94]	[2.95]	[2.96]
Structure			
Propagation			
Cladding diameter (μm)	125	188	125
Attenuation (dB/km)	0.23	0.168	0.158
A_{eff} (μm^2)	108	125	112
Need MIMO	Yes	No	Yes
Differential mode delay (ps/km)	20	–	6.1

advanced modulation format signals is also approaching the theoretical nonlinear Shannon limit. Therefore, expanding the number of spatial channels represents the last available dimension to increase capacity.

For this purpose, various types of space-division multiplexing (SDM) fibers have been proposed and demonstrated [2.97]. Among the wide variation, realistic candidates for long-distance transmission fiber include few-mode fiber FMF [2.94], uncoupled multicore fiber (U-MCF) [2.95], and coupled multicore fiber (C-MCF) [2.96], as summarized in Table 2.4.

FMF supports several guided modes and therefore is similar to MMF. Actually, a standard MMF can also be used as FMF [2.98]. However, whereas every guided mode of legacy MMF carries the same signal, each guided mode of FMF is loaded with a different signal. Although those signals are mixed during transmission, because of mode-coupling in FMF, multi-input multi-output (MIMO) digital signal processing can unbundle the mode-coupling and recover the original signal. Since MIMO processing becomes more complex with larger differential mode delay (DMD), it is important to reduce the DMD of FMF. To reduce DMD, it is effective to concatenate two FMFs having DMDs of opposite signs. With this technique, a low DMD of 20 ps/km has been achieved [2.94].

U-MCF contains multiple independent cores in its cross-section. Since it can be treated simply as parallel fibers, it has the advantage that conventional transceivers can be used without change. However, the challenge is that isolation between the cores requires a large separation and/or a deep depression in the cladding index. A large separation can result in a cladding diameter larger than the conventional 125 μm, which might affect reliability due to enhanced bend-induced stress. A deep depression in the cladding might degrade productivity and cost-efficiency. Therefore, significant work will be necessary before practical use as transmission fiber.

C-MCF contains multiple cores that are positioned closely enough to experience mode-coupling during transmission. It has the advantage that the conventional 125 μm diameter cladding can contain as many as 12 cores [2.99]. Whereas it requires MIMO processing as FMF does, the cost of the MIMO calculation is expected to be lower than that in the case of FMF because the DMD of C-MCF is smaller than that of FMF. Since the spatial modes of C-MCF propagate having random coupling with each other, the differential delay between the modes accumulates with random changes in the sign and magnitude of the coupling, so that the DMD grows in proportion to the square root of the distance, not directly in proportion to the distance as in the deterministic case [2.96]. In addition to these advantages, it was also demonstrated that a C-MCF can be fabricated with an attenuation as low as 0.158 dB/km at 1550 nm, this

is comparable to those of the most advanced CSF products for submarine applications [2.96]. Further, a lab transmission experiment achieved 5500 km transmission over a C-MCF and superior transmission quality that is presumed to be due to the reduced nonlinearity caused by mode coupling [2.100]. In spite of those promising results, many challenges still remain, such as integration of a suitable multicore amplifier, low-loss and low-cost splice techniques, and control of the mode-coupling characteristics of cabled fiber before this new fiber may be used in the field.

2.4.2 Photonic Band-Gap Fiber (PBGF)

In this section the features of PBGF are described. Photonic band-gap fiber (PBGF) is an optical fiber that guides light by the photonic band gap (PBG) effect, in contrast to the conventional optical fiber that guides light by total internal reflection (TIR) [2.101]. Whereas TIR guidance requires the core to have a higher refractive index than the cladding, PBG guidance requires the cladding to have a micron-scale periodic structure that prohibits propagation of light in a range of frequency that is the photonic band gap. The unique guiding mechanism generates an important advantage of PBGF in that it can guide light in a low-index hollow core [2.102]. Guidance in a hollow core could have a great impact over a conventional glass core due to low attenuation, low nonlinearity, low latency, and low sensitivity to environmental disturbances such as temperature and radiation. Although the lowest achieved attenuation of a hollow-core PBGF (HC-PBGF) is still as high as 1.2 dB/km at 1620 nm, and presumed to be limited by the surface roughness of the hollow core [2.103], nevertheless the advantage of low latency have been demonstrated by transmitting a 10 Gb/s signal at 1550 nm over 11 km of HC-PBGF, resulting in more than a 15 μs reduction in latency [2.104].

2.5 Optical Fiber Cables

Optical fiber cables are produced to protect the optical fibers in the field, for example outside the plant, intrabuilding, or central office. Hence, by considering the use case, several factors need to be considered and optimized:

- Number of fibers per cable
- Environmental conditions
- Mechanical durability
- Fire retardancy.

Fiber count can be defined by use case. A *drop cable* is used to connect the subscriber's house and is typically a single fiber-cable, while a *feeder cable* is used to distribute fibers to a residential area and contains several hundred to one thousand fibers.

The environmental conditions can affect the design of the cable structure. The key considerations are temperature and moisture (or surrounding water) or tensile force in the case of an aerial cable. Optical cables are designed to maintain their performance under the anticipated environmental conditions.

Mechanical durability has two aspects. One is a durability during deployment, and the other is residual external forces after being deployed.

Fire retardancy is required in order to avoid fire propagation along the cables and smoke and/or toxic gas generation.

In this section the structure of optical cables is described, and we present a variety of actual optical fiber cables in use, highlighting in particular FTTH (fiber-to-the-home) network and intrabuilding network applications.

2.5.1 Cable Structure Used in Access Network and Intrabuilding Network

The *access network* is built to bring fiber(s) from the central office to the customer's residence, multi-dwelling unit, or business office. Figure 2.9 schematically describes the FTTH access network and the intrabuilding network. Here the fibers are connected via feeder cable, distribution cable, and drop cable [2.105–107].

The intrabuilding network consists of a backbone cable and horizontal cable. The backbone cable carries fibers to higher floors. Horizontal cables are connected from the branching point of the backbone cable to each subscriber.

Types of Optical Cables

Various kinds of optical cables are used to build an optical network, Table 2.5 summarizes the relationships about usages, cable types and design factors.

2.5.2 Cable Types

In this section, several types of optical fiber units and cable structures are described.

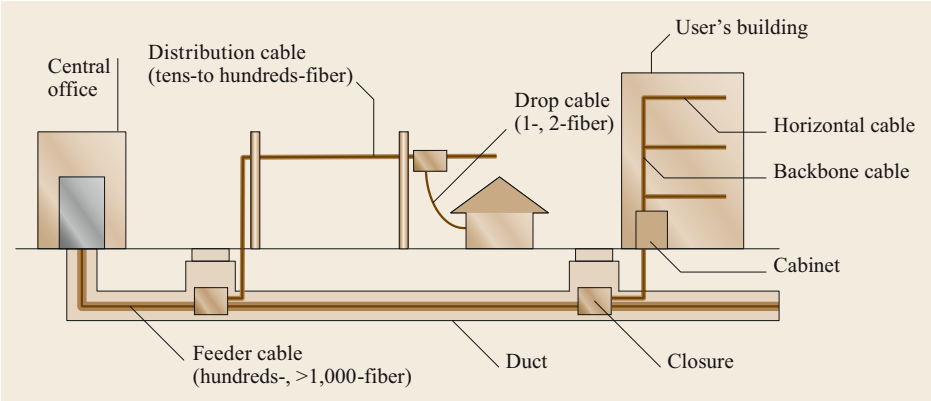


Fig. 2.9 FTTH access network and intrabuilding network

Table 2.5 Usages, cable types, and design factors in the FTTH access network

Cable/usagel		Fiber count	Cable types	Additional design factors
Outdoor network	Feeder	Hundreds, > 1000	• Ribbon slot • Loose tube • Central tube	• Tensile strength for installation • Water blocking • Armored
	Distribution	8 to hundreds	• Ribbon slot • Loose tube • Central tube • Direct jacketed	• Self-supporting • Durability to sunshine • Fire retardancy • Biological attack
	Drop	1–2	• Direct jacketed	
Intrabuilding network	Backbone	8, 12, 100–300	• Layered • Ribbon slot • Loose tube • Central tube	• Fire retardancy • Smoke/gas generation
	Horizontal	1–12	• Direct jacketed • Loose buffered	

Optical Fiber Unit

An optical fiber *unit* is a preassembled cable consisting of a plurality of fibers that are bonded, bundled, or tubed. To avoid misconnections during deployment, identification of each fiber in the cable is critical.

The unit helps operators organize and identify each fiber before and after installation. Identification of each fiber in the unit is usually conducted by visually checking the color of the fiber coating.

There are two types of units deployed into the field, consisting of either optical fiber ribbon or loose tube.

Optical Fiber Ribbon. Optical fiber ribbon is a unit in which a number of optical fibers are arrayed and held together using resin. The number of optical fibers in one ribbon typically available in the market are 4, 8, 12, and 24. If one cable needs many fibers, a higher count fiber ribbon can be selected.

It should be noted that the ribbon structure makes fiber identification easy, as the colored fiber can show the fiber number and ribbon number. Additionally, they can be connected by using mass-fusion splicers [2.108–111], and this serves as an efficient deployment method. Examples of fiber ribbons are shown in Fig. 2.10.

Loose Tube. A loose-tube unit comprises a plastic tube filled with a buffer compound and a number of fibers. The compound is typically a jelly-like material and acts

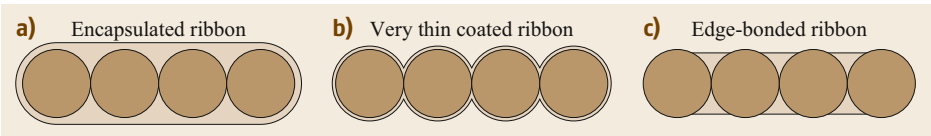


Fig. 2.10a–c Cross-sections of examples of optical fiber ribbons

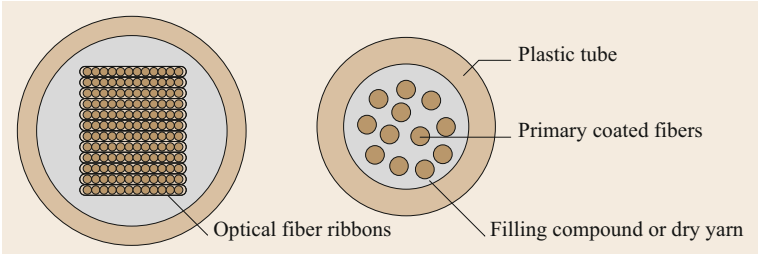


Fig. 2.11 Cross-sections of examples of loose tubes

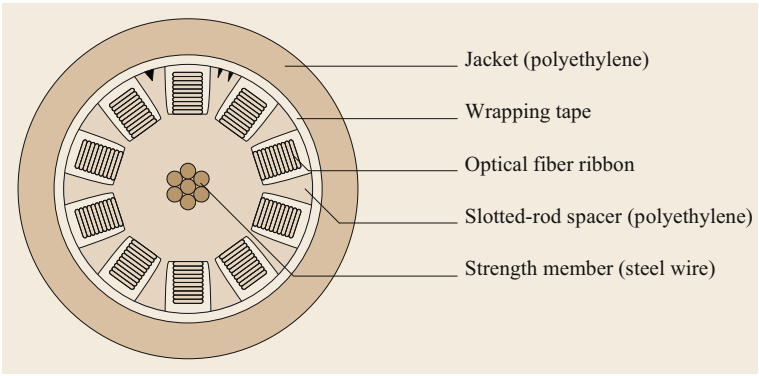


Fig. 2.12 Example of a cross-section of a cable with a slotted-rod spacer

as a lubricant that ensures fibers are free to move inside and prevents an attenuation increase by local fiber bending. In addition, the compound helps prevent longitudinal water penetration and improves mechanical stability.

For higher fiber counts, a stacked and twisted number of fiber ribbons is used to facilitate fiber identification and mass-fusion splicing.

Figure 2.11 shows an example of a ribbon-type and nonribbon-type loose-tube unit.

Cable Structure

Cable structure is classified by the differences in the cable core located at the center of the fiber cable. The cable core consists of one or more *units* and elements, such as a strength member, spacer, wrapping tape, or filling material. Usually, a cable core forms a cylindrical shape and is surrounded by a protective sheath.

The cable-core structure is selected by considering the fiber count, installation, jointing, and required environmental conditions.

Slotted-Rod Cable. A slotted-rod structure consists of a spacer with a central strength member and a polyethylene rod with slots trenched helically. The unit(s) are located in the stiff slot to avoid external forces being applied to the cable; it also makes identifying the fiber easy because the slots can bundle the fiber units independently. Figure 2.12 shows an example of a ribbon slotted-rod cable [2.112, 113].

Loose-Tube Cable. A loose-tube-cable core is formed by stranding a number of loose-tube units around a central strength member. For this cable, the tube itself needs to be stiff enough to protect the fibers from external forces.

In some cases, a jelly-like water blocking material or water-swellable material is applied within the space between the tubes and the strength member.

It is relatively easy to avoid loss increases caused by cable temperature changes or bending in a loose-tube cable because the jelly-like compound enables the fibers to move freely, therefore mitigating fiber-bending and the effect of environmental condition change. Conversely, the jelly-like compounds require extra cleaning when jointing. Figure 2.13 shows examples of loose-tube cables [2.114–122].

Central-Tube Cable. In a central-tube cable, typically 12-, 24-, or 36-fiber ribbons are stacked and twisted within a plastic tube. There are two typical structures: in one the tube space is filled with a jelly-like compound and in the other yarn with water-swellable material is applied to prevent longitudinal water penetration.

Figure 2.14a shows a 576-fiber cable with aramid-yarn strength members surrounding the central tube. Figure 2.14b shows another example of an 864-fiber ca-

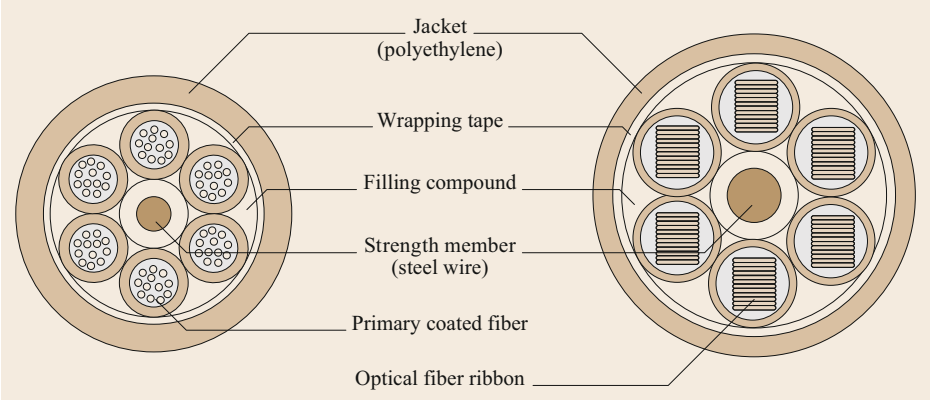


Fig. 2.13 Examples of cross-sections of loose-tube cables

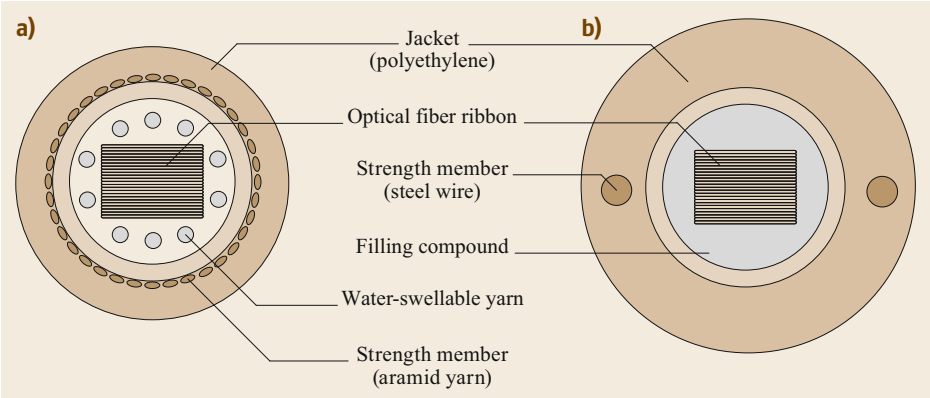


Fig. 2.14a,b Examples of cross-sections of central-tube cables: **(a)** 576-fiber cable; **(b)** 864-fiber cable

ble with two steel strength members contained in the sheath [2.123–125].

Layered Fiber Core Cable. Layered fiber core cable is formed by stranding the fibers around a central strength member; it is used for smaller fiber-count cables such as 6- to 12-fiber cables. Usually, 900 μm-diameter primary coated fiber is used for this type of cable. Figure 2.15 shows an example of a layered fiber cable.

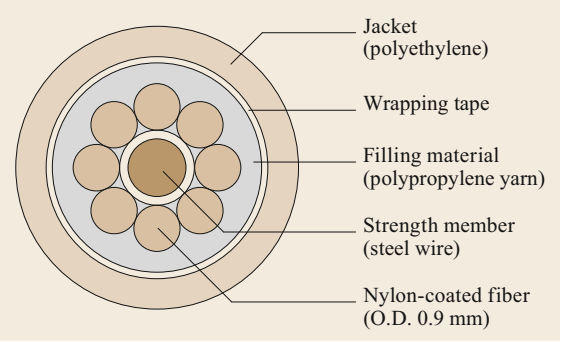


Fig. 2.15 Layered fiber cable

Direct Jacketed Cable

For fewer fiber count cabling, a direct jacketed cable structure is also used. This jacket is applied directly to the primary coated fiber(s) or fiber ribbon(s).

Because this structure is of lighter weight, it is suitable for aerial drop and indoor cable applications. Thinner and lighter cable also enables the suspension wire of an aerial drop cable to be thinner. For indoor cable that is installed on the room wall, thinner cable is more aesthetically pleasing and involves a lower risk of becoming an *eyesore*. Figure 2.16 shows examples of cross-sections of direct jacketed cables [2.126–133].

2.5.3 Cable Design

Each optical fiber cable should be designed to achieve the required performance and reliability under field conditions. This section describes the factors that should be considered in cable design.

Optical Fiber Density

If high optical fiber density in a cable is required, the fibers can experience random lateral forces from neighboring fibers or the walls of the slotted-rod spacers. This

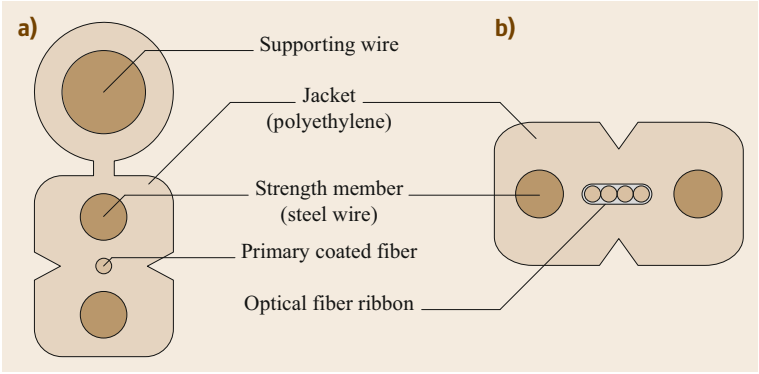


Fig. 2.16a,b Examples of cross-sections of direct jacketed cables: **(a)** drop cable, **(b)** indoor cable

force results in microdeformation of the fiber, called microbending, that causes attenuation to increase. To avoid microbending loss, the optical fiber density of a cable is selected within the appropriate range [2.6, 134, 135].

Environment Conditions

This section describes how cables need to be optimized for various environmental conditions in order to ensure stable loss performance.

Temperature Change. Temperature change results in stresses on the cable, such as elongation or shrinkage of the components within a cable caused by the differences in coefficients of thermal expansion between the fiber and the cable elements (normally made from polymer plastic). This stress may cause microbending to fibers and loss increases. Especially at low temperature, cable elements shrink and therefore create an excess length of fiber that then needs to be accommodated within the cable. This causes microbending in the fibers and increased loss.

Therefore, optical fiber cables should be designed so that the fibers have enough space not to experience the stresses or microbending within the expected temperature range [2.136].

Surrounding Water (Water Blocking). When the cable is installed in an underground duct, the cable can unintentionally get soaked with water when the ducts are flooded by groundwater. Once a part of the cable jacket is damaged, water penetrates into the cable and may reach the fibers. The existence of water or moisture around optical fibers may cause the degradation of fiber strength because water molecules react with silica (SiO_2) and accelerate the growth of glass-surface microflaws. To prevent this, underground cables require a water-blocking function.

To protect against water penetration, ribbon slotted-core cable wraps use water-swellable tapes, while loose-tube cables use jelly-like filling compound as a water-blocking material.

Biological Attacks. For outdoor cables, especially the aerial cable type, these may suffer biological attacks from mammals, birds or insects, so cables are designed to have effective protection [2.137–145].

One example of a mammal pest are squirrels that bite aerial cables and cut the fiber. Woodpeckers have been known to peck the cables and make holes in the jackets. Some types of cicada have been known to lay eggs within thin drop cables and so break the fibers.

For protection against these pests, wrapping the cable with metallic tape or net will work. To avoid cicadas laying eggs in drop cables, harder jacket material has been shown to be effective. Figure 2.17 depicts the example of a cicada attack on a drop cable.

Mechanical Durability

Optical fiber cables incur various external forces during both installation and operation, hence the cables need to be designed to maintain the required properties even when external forces are being applied [2.146, 147].

Tensile Forces in Installation. To install a cable in a building duct, the cable is connected to a preinstalled rope and is pulled through. During this pulling operation, the cable experiences tensile forces due to friction inside the duct. If the cable piece is elongated by the tensile force, the optical fibers in the cable may be broken. Therefore, the tension member(s) in the cable should be designed to limit cable elongation during this operation. For example, a backbone intrabuilding cable is specified to withstand the weight of itself because it is installed by being pulled vertically from the top of the building, as shown in Fig. 2.18 [2.148].

For the strength member, steel wire is commonly used. If electric induction needs to be avoided, then aramid, FRP (fiber reinforced plastic), or yarn is used.

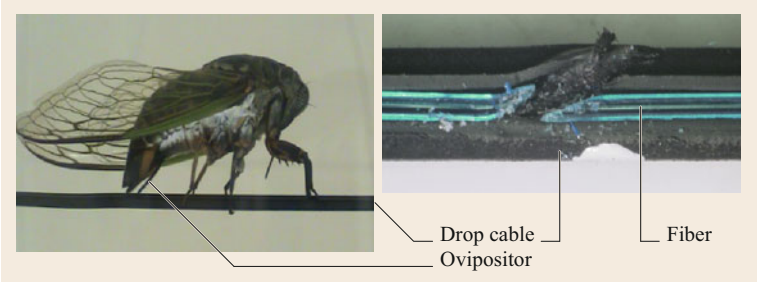


Fig. 2.17 Cicada attack on a drop cable

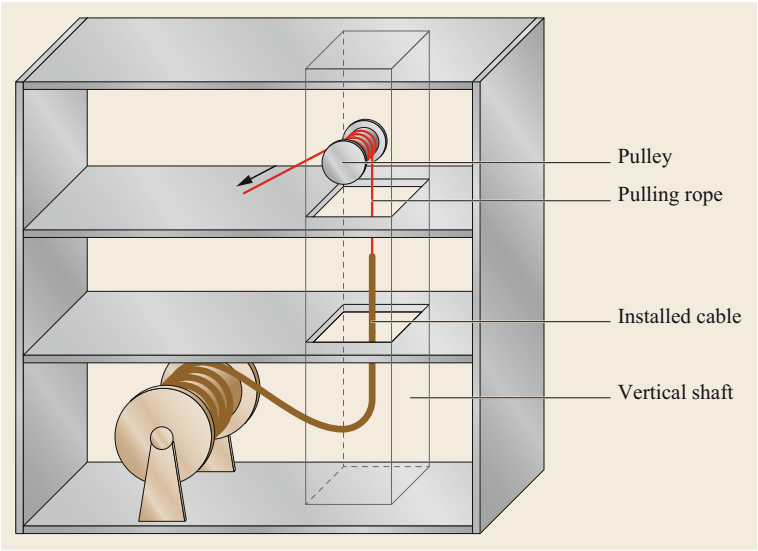


Fig. 2.18 Schematic layout of vertical installation in building

External Forces During Installation. In addition to tensile stress during the installation, operation cables are subjected to various external forces, such as crushing, twisting, bending, and impact. Cables should provide an effective protective means against the expected external forces.

As for the slotted-rod cable, the spacer wall should be thick and hard enough to withstand crushing and impact, and the size of slot should be large enough to withstand stress on the fibers under twisting and bending. Conversely, loose-tube and central-tube cable protects the fiber due to the hardened tube material.

The mechanical strength requirements are standardized by the IEC (International Electrotechnical Commission). Table 2.6 shows examples of the IEC standards for each external force. There are various conditions or criteria described in IEC 60794-3-10, as well as other standards.

Fire Retardancy

Because the outermost cable material is usually made of plastic, fire retardancy is required to avoid fire propagation along the cables and toxic gas and/or smoke generation, especially for office buildings with high-rise multidwelling units, or tunnels. The regulations on fire

Table 2.6 Example of IEC standards for external forces

External force	Test condition	Criteria
Bend	Radius: 20 times cable diameter Bend angle: 180°	No increase in attenuation at 1550 nm during and after the test.
Impact	Impact energy: 3 J (300 g hammer drop from 1 m height)	Under visual examination, there shall be no damage to the sheath or to the cable elements.
Torsion	180° twist over 2 m length	
Crush	Load: 1.5 kN/100 mm plate Duration: 1 min	

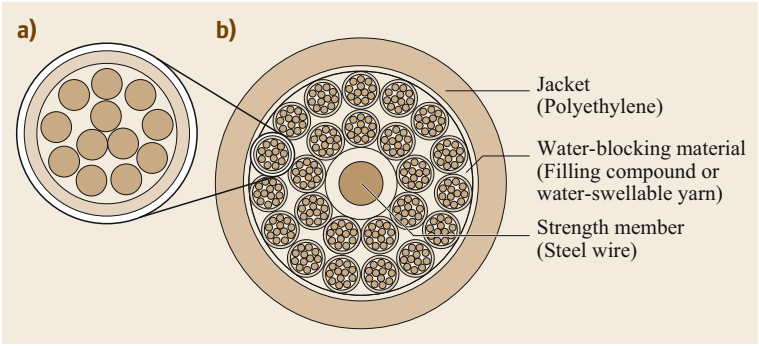


Fig. 2.19a,b An example of microduct cable **(a)** and micromodule cable **(b)**

retardancy may differ by country or usage [2.125, 149, 150].

If fire retardancy is required, fire-retardant materials are selected for the outer sheath. Usually these are made from polyethylene with the addition of some self-extinguishing material.

2.5.4 Cables for Specific Needs

Cables for Air-Blown Fiber Installation

Air-blown fiber cable is applied because of the simplicity of fiber deployment in buildings. The cable is literally blown into the duct or tube by applying high-pressure air; this operation greatly reduces the tension during installation. Moreover, the length of cable installation is known to be longer and the operation time shorter than for conventional pulling operations. This is beneficial for simplicity of the installation operation [2.151–162].

In the case of outside plant installations, microduct cables are used. An example of the structure of a microduct cable is shown in Fig. 2.19. This cable can be installed up to lengths of several kilometers.

As seen in Fig. 2.20, a fiber bundle unit is also used. Here, the bundle unit means a number of fibers that are sheathed with thin polyethylene formers in order to improve installation. This technique is very suitable for intrabuilding installations because tube cables are installed beforehand. Fiber bundles are installed when fiber needs arise and, if needed, bundles can be removed from the tube cable by applying an air flow in the opposite direction [2.106, 163–167].

Ultrahigh-Count Density Cable

Ultrahigh density and high-count fiber cable are among the newest products. This cable adopts the ribbon-slotted-rod type or central-tube-type structure, however the optical fiber ribbon is different from conventional cables. The applied ribbon unit has a feature of ultrahigh density known as *pliable ribbon*.

The ribbon arrays of fibers are incorporated in a mesh-like structure, as shown in Fig. 2.21. Pliable ribbons can be packed tighter than ordinary one-

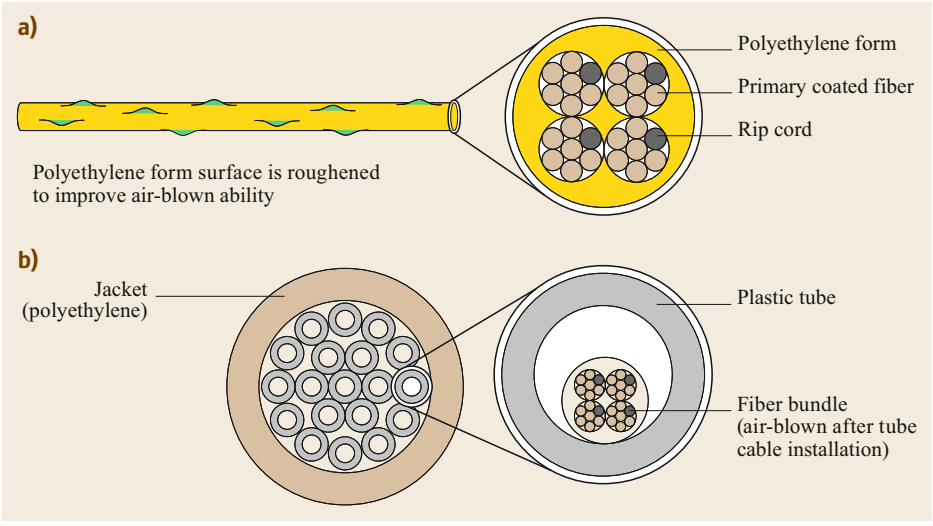


Fig. 2.20a,b Fiber bundle **(a)** and tube cable **(b)** for air-blown installation

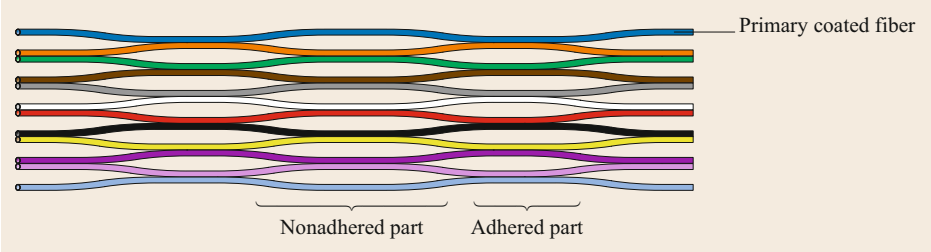


Fig. 2.21 Pliable ribbon unit

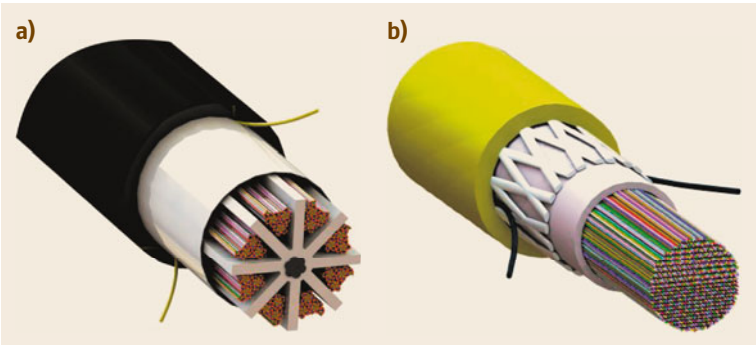


Fig. 2.22a,b Example of ultrahigh-count high-density cables: **(a)** slotted rod spacer type 3456-fiber cable; **(b)** central tube type 1728-fiber cable

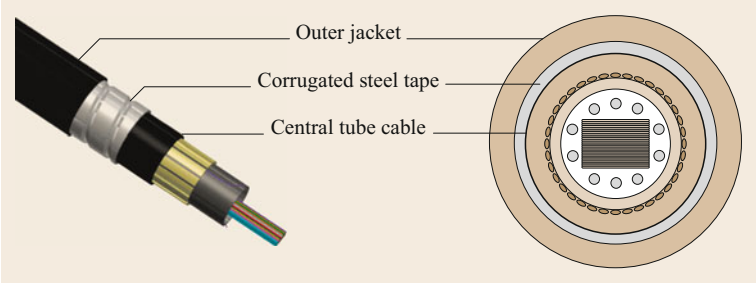


Fig. 2.23 Example of armored-sheath cable

dimensional ribbon units because they can pile the fiber from a one-direction array into a second dimension, thus filling the empty space. One of the highest fiber-count cables is a 3456-fiber cable with a 34 mm diameter that can be installed into a two-inch duct [2.107, 108, 168–174] as seen in Fig. 2.22.

Armored Sheath

In rural areas, to reduce cost and increase efficiency, the cable will occasionally be directly buried and installed into the ground by trenching during fiber-cable deployment, in contrast to urban-area deployments where fiber is placed within high-cost, but protective, conduits. Additional protection of the buried ground cable is required for it to withstand ground pressure and other elements. The most commonly used protection is a corrugated metallic-tape armored sheath over the ordinary jacket.

This kind of cable can also be used for duct, aerial or indoor installation, in the case of harsh conditions or where rodent attacks are expected [2.175, 176]. An example is shown in Fig. 2.23.

Induction-Free Cable

Optical fiber cable may be installed close to a high-voltage power cable. To avoid the electrical induction between a steel strength member and the power cable, an induction-free cable is required. This type is free from metal elements and instead uses aramid, FRP, or yarn applied as the strength member.

2.5.5 ITU-T Technical Reports

ITU (International Telecommunication Union) is the United Nations' specialized agency for information and communication technologies—ICTs, and ITU-T is the telecommunication standardization sector in the ITU. The ITU-T provides international standards for telecommunication systems and facilities *recommendations*. Besides the recommendations, technical re-

ports and technical papers are also published for guidance or reference. Among the technical reports, *ITU-T Technical Paper on Optical Fibres, Cables and Systems* is a very useful overview of optical-fiber cable structure, installation, connection, and systems [2.177–179].

References

2.1 D. Marcuse: *Theory of Dielectric Optical Waveguides* (Academic Press, New York 1974)
2.2 B. Elliott, M. Gilmore: *Fiber Optic Cabling*, 2nd edn. (Newnes, Oxford 2002)
2.3 K.C. Kao, G.A. Hockham: Dielectric fiber surface waveguides for optical frequencies, Proc. IEEE Inst. Electr. Electron Eng. **113**(7), 1151–1158 (1966)
2.4 F.P. Kapron, D.B. Keck, R.D. Maurer: Radiation losses in glass optical waveguides, Appl. Phys. Lett. **17**(7), 423–425 (1970)
2.5 M. Finch: China's 'transition phase': How deep, how long, and what are the implications on international markets? In: *CRU Wire Cable Conf.* (2019)
2.6 IEC 60793-1-44:2011: Measurement methods and test procedures—Cut-off wavelength (2011)
2.7 ITU-T G.650.1: Definitions and test methods for linear, deterministic attributes of single-mode fibre and cable (2010)
2.8 ITU-T G-Series Recommendations—Supplement 39: Optical system design and engineering considerations (2016)
2.9 Recommendation ITU-T G.652: Characteristics of a single-mode optical fibre and cable (2016)
2.10 Sumitomo Electric Industries: Sumitomo Electric fiber optics products for FTTx solutions, http://global-sei.com/fttx/images_n/EMEA.pdf (2016)
2.11 Corning: SMF-28e+ optical fiber, http://www.corning.com/media/worldwide/coc/documents/Fiber/PI1463_07-14_English.pdf (2014)
2.12 OFS: AllWave fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/AllWave-117-web.pdf (2016)
2.13 D. Marcuse: Loss analysis of single-mode fiber splices, Bell Sys. Tech. J. **56**(5), 703–718 (1977)
2.14 Recommendation ITU-T G.657: Characteristics of a bending-loss insensitive single-mode optical fibre and cable (2016)
2.15 Corning: SMF28 Ultra optical fiber, http://www.corning.com/content/dam/corning/media/worldwide/coc/documents/Fiber/PI1450.pdf (2014)
2.16 OFS: AllWave One fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/AllWave-One-Fiber-160-web.pdf (2016)
2.17 Prysmian: BendBright single-mode optical fiber, https://www.prysmiangroup.com/sites/default/files/business_markets/markets/downloads/datasheets/SMF---BendBright-Single-Mode-Optical-Fiber.pdf (2010)
2.18 Corning: ClearCurve LBL Ultra optical fiber, http://www.corning.com/media/worldwide/coc/documents/Fiber/PI1488_07-14_English.pdf (2014)
2.19 OFS: AllWave Flex+ fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/AllWave-FLEX-PLUS-144-web.pdf (2016)
2.20 Prysmian: BendBright-XS single-mode optical fiber, https://www.prysmiangroup.com/sites/default/files/business_markets/markets/downloads/datasheets/SMF---BendBright-XS-Single-Mode-Optical-Fiber.pdf (2010)
2.21 Corning: ClearCurve ZBL Ultra optical fiber, http://www.corning.com/media/worldwide/coc/documents/Fiber/PI1464_07-14_English_2.pdf (2014)
2.22 OFS: AllWave Flex Max fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/AllWave-FLEX-Max-Fiber-154-web.pdf (2016)
2.23 Prysmian: BendBright-Elite single-mode optical fiber, https://www.prysmiangroup.com/sites/default/files/business_markets/markets/downloads/datasheets/DrakaElite-BendBright-Elite-Fiber-for-Patch-Cord_1.pdf (2010)
2.24 P. Matthijsse, L.-A. de Montmorrillon, G.-J. Krabshuis, F. Gooijer: Bend-optimized G.652 compatible single mode fibers. In: *Proc. 54th IWCS Conf.* (2005) pp. 327–331
2.25 K. Himeno, S. Matsuo, N. Guan, A. Wada: Low-bending-loss single-mode fibers for fiber-to-the-home, J. Lightwave Technol. **23**(11), 3494–3499 (2005)
2.26 L. de Montmorillon, P. Matthijsse, F. Gooijer, F. Achten, D. Molin, N. Montaigne, J. Maury: Bend-optimized G.652D compatible trench-assisted single mode fibers. In: *Proc. 55th IWCS Conf.* (2006) pp. 342–347
2.27 K. Nakajima, K. Hogari, J. Zhou, K. Tajima, I. Sankawa: Hole-assisted fiber design for small bending and splice losses, Photon. Tech. Lett. **15**(12), 1737–1739 (2003)
2.28 M.-J. Li, P. Tandon, D.C. Bookbinder, S.R. Bickham, M.A. McDermott, R.B. Desorcie, D.A. Nolan, J.J. Johnson, K.A. Lewis, J.J. Englebert: Ultra-low bending loss single-mode fiber for FTTH, J. Lightwave Technol. **27**(3), 376–382 (2009)
2.29 Corning: SMF-28 Ultra 200 optical fiber, http://www.corning.com/media/worldwide/coc/documents/Fiber/PI1450.pdf (2015)
2.30 OFS: AllWave Flex 200 μm fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/AllWave-FLEX-200-um-Fiber-146-web.pdf (2017)
2.31 OFS: AllWave Flex+ 200 μm fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/AllWave-FLEX+-200-um-Fiber-147-web.pdf (2017)
2.32 Prysmian: BendBright-XS 200 μm single-mode optical fiber, https://www.prysmiangroup.com/sites/default/files/business_markets/markets/downloads/datasheets/SMF---BendBright-XS-200-m-Single-Mode-Optical-Fiber.pdf (2010)
2.33 R. Olshansky: Mode coupling effects in graded-index optical fibers, Appl. Opt. **14**(4), 935–945 (1975)

2.34 D. Gloge: Optical-fiber packaging and its influence on fiber straightness and loss, Bell Sys. Tech. J. **54**(2), 245–262 (1975)

2.35 J. Baldauf, N. Okada, M. Miyamoto: Relationship of mechanical characteristics of dual coated single mode optical fibers and microbending loss, Trans. IEICE Commun. **E76-B**(4), 352–357 (1993)

2.36 P. Sillard, S. Richard, L. de Montmorillon, M. Bigot-Astruc: Micro-bend losses of trench-assisted single-mode fibers. In: *Proc. ECOC* (2010), Paper We.8.F.3

2.37 B. Overton, D. Gharbi, O. Delwal, O. Tatat, L. Provost, L. de Montmorillon: Reduced diameter optical fiber and cable. In: *Proc. 59th IWCS Conf.* (2010) pp. 117–122

2.38 Recommendation ITU-T G.650.2: Definitions and test methods for statistical and nonlinear related attributes of single-mode fibre and cable (2015)

2.39 T. Kato, M. Hirano, M. Onishi, M. Nishimura: Ultra-low nonlinearity low-loss pure silica core fibre for long-haul WDM transmission, Electron. Lett. **35**(19), 1615–1617 (1999)

2.40 S. Bickham: Ultimate limits of effective area and attenuation for high data rate fibers. In: *Proc. OFC* (2011), Paper OWA5

2.41 D. Peckham, A. Klein, P. Borel, R. Jensen, O. Levring, K. Carlson, M. Yan, P. Wisk, D. Trevor, R. Lingle Jr., A. McCurdy, B. Zhu, Y. Zou, R. Norris, B. Palsdottir, D. Vaidya: Optimization of large area, low loss fiber designs for C+L band transmission. In: *Proc. OFC* (2016), Paper Tu3G.1

2.42 M. Hirano, Y. Yamamoto, Y. Tamura, T. Haruna, T. Sasaki: Aeff-enlarged pure-silica-core fiber having ring-core profile. In: *Proc. OFC* (2012), Paper OTh4I.2

2.43 Y. Yamamoto, Y. Kawaguchi, M. Hirano: Low-loss and low-nonlinearity pure-silica-core fiber for C- and L-band broadband transmission, J. Lightwave Technol. **34**(2), 321–326 (2016)

2.44 Y. Tamura, H. Sakuma, K. Morita, M. Suzuki, Y. Yamamoto, K. Shimada, Y. Honma, K. Sohma, T. Fujii, T. Hasegawa: Lowest-ever 0.1419-dB/km loss optical fiber. In: *Proc. OFC* (2017), Paper Th5D.1

2.45 Recommendation ITU-T G.654: Characteristics of a cut-off shifted single-mode optical fibre and cable (2016)

2.46 Sumitomo Electric Industries: Sumitomo submarine optical fibers, http://global-sei.com/fttx/images_n/TR_EG010_201703_SumitomoSubmarineFibers.pdf (2017)

2.47 Corning: Vascade optical fiber, http://www.corning.com/media/worldwide/coc/documents/Fiber/PI1445_3.17.pdf (2017)

2.48 OFS: TeraWave SCUBA Ocean optical fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/TeraWave-Scuba-Ocean-Fibers-fiber-168-web.pdf (2017)

2.49 J. Cai, Y. Sun, H. Zhang, H. Batshon, M. Mazurczyk, O. Sinkin, D. Foursa, A. Pilipetskii: 49.3 Tb/s transmission over 9100 km using C+L EDFA and 54 Tb/s transmission over 9150 km using hybrid-Raman EDFA, J. Lightwave Technol. **33**(13), 2724–2734 (2015)

2.50 J. Cai, H. Batshon, M. Mazurczyk, O. Sinkin, D. Wang, M. Paskov, W. Patterson, C. Davidson, P. Corbett, G. Wolter, T. Hammon, M. Bolshtyansky, D. Foursa, A. Pilipetskii: 70.4 Tb/s capacity over 7,600 km in C+L band using coded modulation with hybrid constellation shaping and nonlinearity compensation. In: *Proc. OFC* (2017), Paper Th5B.2

2.51 H. Bissessur, C. Bastide, S. Etienne, S. Dupont: 24 Tb/s unrepeated C-Band transmission of real-time processed 200 Gb/s PDM-16-QAM over 349 km. In: *Proc. OFC* (2017), Paper Th4D.2

2.52 J. Cho, X. Chen, S. Chandrasekhar, G. Raybon, R. Dar, L. Schmalen, E. Burrows, A. Adamiecki, S. Corteselli, Y. Pan, D. Correa, B. McKay, S. Zsigmond, P. Winzer, S. Grubb: Trans-Atlantic field trial using probabilistically shaped 64-QAM at high spectral efficiencies and single-carrier real-time 250-Gb/s 16-QAM. In: *Proc. OFC* (2017), Paper Th5B.3

2.53 S. Shen, G. Wang, Y. He, S. Wang, C. Zhang: G.654 fibre and cable evaluation for terrestrial high bit-rate transport application. In: *Proc. 65th IWCS Conf.* (2016) pp. 470–475

2.54 S. Shen, G. Wang, H. Wang, Y. He, S. Wang, C. Zhang, C. Zhao, J. Li, H. Chen: G.654.E fibre deployment in terrestrial transport system. In: *Proc. OFC* (2017), Paper M3G.4

2.55 P. Poggiolini: The GN model of non-linear propagation in uncompensated coherent optical systems, J. Lightwave Technol. **30**(24), 3857–3879 (2012)

2.56 R. Essiambre, G. Kramer, P. Winzer, G. Foschini, B. Goebel: Capacity limits of optical fiber networks, J. Lightwave Technol. **28**(4), 662–701 (2010)

2.57 F. Buchali, F. Steiner, G. Bocherer, L. Schmalen, P. Schulte, W. Idler: Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration, J. Lightwave Technol. **34**(7), 1599–1609 (2016)

2.58 M. Hirano, Y. Yamamoto, V. Sleiffer, T. Sasaki: Analytical OSNR formulation validated with 100G-WDM experiments and optimal subsea fiber proposal. In: *Proc. OFC* (2013), Paper OTu2B.6

2.59 V. Curri, A. Carena, G. Bosco, P. Poggiolini, M. Hirano, Y. Yamamoto, F. Forghieri: Fiber figure of merit based on maximum reach. In: *Proc. OFC* (2013), Paper OTh3G.2

2.60 T. Hasegawa, Y. Yamamoto, M. Hirano: Optimal fiber design for large capacity long haul coherent transmission, Opt. Express **25**(2), 706–712 (2017)

2.61 Recommendation ITU-T G.653: Characteristics of a dispersion-shifted, single-mode optical fibre and cable (2010)

2.62 L. Cohen, C. Lin, W. French: Tailoring zero chromatic dispersion into the 1.5–1.6 μm low-loss spectral region of single-mode fibres, Electron. Lett. **15**(12), 334–335 (1979)

2.63 M. Li, D. Nolan: Optical transmission fiber design evolution, J. Lightwave Technol. **26**(9), 1079–1092 (2008)

2.64 OFS: TrueWave RS optical fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/TrueWaveRSLWP-120-web.pdf (2017)

2.65 OFS: TrueWave LA low water peak optical fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/TrueWave-LA-Fiber-158-web.pdf (2017)

2.66 Corning: LEAF optical fiber, http://www.corning.com/media/worldwide/coc/documents/Fiber/PI1107_07-14_English.pdf (2014)

2.67 Prysmian: Draka TeraLight optical fiber, https://www.prysmiangroup.com/sites/default/files/business_markets/markets/downloads/datasheets/SMF–TeraLight-Optical-Fiber.pdf (2010)

2.68 L. Grüner-Nielsen, M. Wandel, P. Kristensen, C. Jørgensen, L. Jørgensen, B. Edvold, B. Pálsdóttir, D. Jakobsen: Dispersion-compensating fibers, J. Lightwave Technol. **23**(11), 3566–3579 (2005)

2.69 OFS: Dispersion slope compensating modules, http://fiber-optic-catalog.ofsoptics.com/viewitems/modules/dispersion-slope-compensating-modules? (2017)

2.70 OFS: TrueWave® Ocean optical fiber HD (high dispersion), http://fiber-optic-catalog.ofsoptics.com/viewitems/modules/dispersion-slope-compensating-modules? (2017)

2.71 R. Oishansky, D. Keck: Pulse broadening in graded-index optical fibers, Appl. Opt. **15**(2), 483–491 (1976)

2.72 O. Butova, K. Golanta, A. Tomashuka, M. van Stralenb, A. Breuls: Refractive index dispersion of doped silica for fiber optics, Opt. Commun. **213**, 301–308 (2002)

2.73 N. Guan, K. Takenaga, S. Matsuo, K. Himeno: A fluorine and germanium co-doped multimode fibre for wide operation window. In: *Proc. ECOC* (2006), Paper Mo.3.3.4

2.74 M. Bigot, D. Molin, F. Achten, A. Amezcua-Correa, P. Sillard: Extra-wide-band OM4 MMF for future 1.6Tbps data communications. In: *Proc. OFC* (2015), Paper M2C.4

2.75 O. Kogan, S. Bickham, M. Li, P. Tandon, J. Abbott, S. Garner: Design and characterization of bend-insensitive multimode fiber. In: *Proc. 60th IWCS Conf.* (2011) pp. 154–159

2.76 D. Molin, M. Bigot-Astruc, K. de Jongh, G. Kuyt, P. Sillard: Trench-assisted bend-resistant OM4 multi-mode fibers. In: *Proc. 59th IWCS Conf.* (2010) pp. 439–443

2.77 P. Moser, J. Lott, P. Wolf, G. Larisch, H. Li, D. Bimberg: 85-fJ dissipated energy per bit at 30 Gb/s across 500-m multimode fiber using 850-nm VCSELs, J. Lightwave Technol. **26**(16), 1638–1641 (2013)

2.78 P. Pepeljugoski, S. Golowich, A. Ritger, P. Kolesar, A. Risteski: Modeling and simulation of next-generation multimode fiber links, J. Lightwave Technol. **21**(5), 1242–1255 (2003)

2.79 P. Pepeljugoski, M. Hackert, J. Abbott, S. Swanson, S. Golowich, A. Ritger, P. Kolesar, Y. Chen, P. Pleunis: Development of system specification for laser-optimized 50 μm multimode fiber for multigigabit short-wavelength LANs, J. Lightwave Technol. **21**(5), 1256–1275 (2003)

2.80 L. Jeunhomme, J. Pocholle: Selective mode excitation of graded index optical fibers, Appl. Opt. **17**(3), 463–468 (1978)

2.81 International standard IEC 60793-2-10 Edition 5.0: Product specifications – Sectional specification for category A1 multimode fibres (2015)

2.82 A. Sengupta: Comparison of min-EMBc and DMD template based qualification of high bandwidth multimode fibers. In: *Proc. 56th IWCS Conf.* (2007) pp. 154–160

2.83 G. Yabre: Comprehensive theory of dispersion in graded-Index optical fibers, J. Lightwave Technol. **18**(2), 166–177 (2000)

2.84 International standard ISO/IEC 11801 2nd edition: Information technology – Generic cabling for customer premises (2002)

2.85 TIA standard TIA-568.3-D: Optical fiber cabling and components standard (2016)

2.86 Corning: ClearCurve multimode optical fiber, http://www.corning.com/media/worldwide/coc/documents/Fiber/PI1468_1-16.pdf (2017)

2.87 OFS: LaserWave FLEX 300/550 (OM3/OM4) optical fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/LaserWave-FLEX-Fiber-139-web.pdf (2016)

2.88 Prysmian: Bend-insensitive 10, 40, 100 Gb/s graded-index multimode fibre, https://www.prysmiangroup.com/sites/default/files/business_markets/markets/downloads/datasheets/MaxCap-BB-OM2-3-4_2013-03.pdf (2013)

2.89 Corning: ClearCurve OM5 wide band multimode optical fiber, http://www.corning.com/media/worldwide/coc/documents/Fiber/PI1469_12.16.pdf (2017)

2.90 OFS: LaserWave FLEX wideband optical fiber, http://fiber-optic-catalog.ofsoptics.com/Asset/LaserWave-FLEX-WideBand-Fiber-163-web.pdf (2016)

2.91 Prysmian: WideCap-OM4 multimode fibre, https://www.prysmiangroup.com/sites/default/files/business_markets/markets/downloads/datasheets/WideCap-OM4_2014-09.pdf (2014)

2.92 Y. Sun, R. Lingle Jr., R. Shubochkin, A. McCurdy, K. Balemarthy, D. Braganza, J. Kamino, T. Gray, W. Fan, K. Wade, F. Chang, D. Gazula, G. Landry, J. Tatum, S. Bhoja: SWDM PAM4 transmission over next generation wide-band multimode optical fiber, J. Lightwave Technol. **35**(4), 690–697 (2017)

2.93 P. Winzer, D. Neilson: From scaling disparities to integrated parallelism: A decathlon for a decade, J. Lightwave Technol. **35**(5), 1099–1115 (2017)

2.94 T. Mori, T. Sakamoto, M. Wada, T. Yamamoto, F. Yamamoto: Few-mode fibers supporting more than two LP modes for mode-division-multiplexed transmission with MIMO DSP, J. Lightwave Technol. **32**(14), 2468–2479 (2014)

2.95 T. Hayashi, T. Taru, O. Shimakawa, T. Sasaki, E. Sasaoka: Uncoupled multi-core fiber enhancing signal-to-noise ratio, Opt. Express **20**(26), B94–B103 (2012)

2.96 T. Hayashi, Y. Tamura, T. Hasegawa, T. Taru: Record-low spatial mode dispersion and ultra-low loss coupled multi-core fiber for ultra-long-haul

transmission, J. Lightwave Technol. **35**(3), 450–457 (2017)
2.97 T. Mizuno, Y. Miyamoto: High-capacity dense space division multiplexing transmission, Opt. Fiber. Technol. **35**, 108–117 (2017)
2.98 P. Sillard, D. Molin, M. Bigot-Astruc, A. Amezcua-Correa, K. de Jongh, F. Achten: 50 μm multimode fibers for mode division multiplexing, J. Lightwave Technol. **34**(8), 1672–1677 (2016)
2.99 T. Sakamoto, S. Aozasa, T. Mori, M. Wada, T. Yamamoto, S. Nozoe, Y. Sagae, K. Tsujikawa, K. Nakajima: Randomly-coupled single-mode 12-core fiber with highest core density. In: *Proc. OFC* (2017), Paper Th1H.1
2.100 R. Ryf, J. Alvarado, B. Huang, J. Antonio-Lopez, S. Chang, N. Fontaine, H. Chen, R. Essiambre, E. Burrows, R. Amezcua-Correa, T. Hayashi, Y. Tamura, T. Hasegawa, T. Taru: Long-distance transmission over coupled-core multicore fiber. In: *Proc. ECOC* (2016), Paper Th.3.C.3
2.101 J. Knight, J. Broeng, T. Birks, P. Russell: Photonic band gap guidance in optical fibers, Science **282**, 1476–1479 (1998)
2.102 R. Cregan, B. Mangan, J. Knight, T. Birks, P. Russell, P. Roberts, D. Allan: Single-mode photonic band gap guidance of light in air, Science **285**, 1537–1540 (1999)
2.103 P. Roberts, F. Couny, H. Sabert, B. Mangan, D. Williams, L. Farr, M. Mason, A. Tomlinson, T. Birks, J. Knight, P. Russell: Ultimate low loss of hollow-core photonic crystal fibres, Opt. Express **13**(1), 236–244 (2005)
2.104 Y. Chen, Z. Liu, S. Sandoghchi, G. Jasion, T. Bradley, E. Fokoua, J. Hayes, N. Wheeler, D. Gray, B. Mangan, R. Slavik, F. Poletti, M. Petrovich, D. Richardson: Multi-kilometer long, longitudinally uniform hollow core photonic bandgap fibers for broadband low latency data transmission, J. Lightwave Technol. **34**(1), 104–113 (2016)
2.105 P. Barker: FTTH infrastructure BT pilot deployments. In: *Proc. 58th IWCS Conf.* (2009) pp. 133–137
2.106 J. Landos, O. Bouffant: Building cabling system field trial to connect residential customers to broadband access network. In: *Proc. 55th IWCS Conf.* (2006) pp. 378–383
2.107 N. Nakagawa, S. Hamaguchi, Y. Endo, K. Omoto, Y. Shibata, Y. Aoyagi: Development of ultra high-count and high-density optical fiber cable (2000 cores). In: *Proc. 64th IWCS Conf.* (2015) pp. 655–658
2.108 N. Itoh, M. Isaji, K. Osato, M. Yamanaka, N. Okada: Development of 2000-fiber ultra-high density underground cable. In: *Proc. 64th IWCS Conf.* (2015) pp. 664–668
2.109 Furukawa Electric: Ribbon fusion splicer S123M series, http://www.furukawa.co.jp/fitel/english/fusion/fusion_s123m.htm (2019)
2.110 AFL: Fujikura 70R fusion splicer, https://www.aflglobal.com/Products/Fusion-Splicing/Field-Fusion-Splicing-Equipment/Splicers-Ribbon-Fiber.aspx (2019)
2.111 Sumitomo Electric Industries: Mass fusion splicer TYPE-71M12, http://global-sei.com/sumitomo-electric-splicers/products/fusion-splicer/t-71m12.html (2017)
2.112 D. Kakuta, T. Yamamoto, S. Tetsutani, K. Shiraki: Development of small diameter 1000-fiber underground optical cable for effective utilization of underground infrastructure facilities. In: *Proc. 58th IWCS Conf.* (2009) pp. 289–292
2.113 K. Tomikawa, D. Takeda, E. Ino, K. Mitsuhashi, N. Okada: Development of small size 1000-fiber cable using BIF. In: *Proc. 58th IWCS Conf.* (2009) pp. 297–302
2.114 D. Collado, B. Overton, J. Ryan: Reduced diameter fiber optic cable family optimized for bend insensitive fiber. In: *Proc. 61st IWCS Conf.* (2012) pp. 42–49
2.115 T. Fukute, A. Namazue, M. Ogi, S. Shiobara, N. Okada: Development of the ultra high density loose tube cables using 200 μm coated fibers for microduct application. In: *Proc. 60th IWCS Conf.* (2011) pp. 56–60
2.116 B.J. Overton, D. Gharbi, O. Delwal, O. Tatat, L. Provost, L. de Montmorillon: Reduced diameter optical fiber and cable. In: *Proc. 59th IWCS Conf.* (2010) pp. 117–122
2.117 W. Stöcklein, H. Knoch: Development of a micro cable family with stranded micromodules for blown cable applications. In: *Proc. 58th IWCS Conf.* (2009) pp. 293–296
2.118 Y. Hashimoto, M. Ogi, F.L. Saputra, N. Okada: Development of the high fiber density dry microduct cable. In: *Proc. 57th IWCS Conf.* (2008) pp. 165–168
2.119 Y. Hashimoto, K. Osato, N. Okada: Development of the downsized and gel-free dry loose tube cable. In: *Proc. 56th IWCS Conf.* (2007) pp. 567–571
2.120 M.R. Santana: Gel-free versus gel-filled buffer tubes. In: *Proc. 54th IWCS Conf.* (2005) pp. 154–160
2.121 OFS: Loose tube fiber optic cables, http://fiber-optic-catalog.ofsoptics.com/viewitems/outdoor-fiber-optic-cables/loose-tube-fiber-optic-cables-1 (2019)
2.122 AFL: Non-armored single jacket dry loose tube cable, http://www.aflglobal.com/Products/Fiber-Optic-Cable/Loose-Tube/Non-Armored-Single-Jacket-Dry-Loose-Tube-Cable.aspx (2011)
2.123 P. Van Vickle, L. Alexander, S. Stokes, W. Aiguo: Central tube cable ribbon coupling. In: *Proc. 57th IWCS Conf.* (2008) pp. 498–503
2.124 OFS: Central core fiber optic cables, http://fiber-optic-catalog.ofsoptics.com/viewitems/outdoor-fiber-optic-cables/central-core-fiber-optic-cables-1 (2019)
2.125 SEL: Indoor/outdoor riser ribbon cable, https://www.sumitomoelectric.com/product/indooroutdoor-riser-ribbon-cable/ (2020)
2.126 M. Tsukamoto, Y. Hoshino: Development of small size single-fiber optical drop cable. In: *Proc. 63rd IWCS Conf.* (2014) pp. 866–869
2.127 Corning: ROC™ drop cable, https://www.corning.com/jp/jp/products/communication-networks/products/fiber-optic-cable/outdoor-cables/drop-cable.html (2012)
2.128 K. Yoneda, K. Kuramoto, H. Iino, K. Nakayachi, Y. Aoyagi, K. Takamizawa: Technology for improving

reliability and workability of drawing and wiring cable for home networks. In: *Proc. 63rd IWCS Conf.* (2014) pp. 862–865

2.129 F. Lissillour, O. Bouffant, N. Evanno, D. Lecoq, P. Guillas, E. Pénard: Innovative in-house cabling solutions for FTTH. In: *Proc. 63rd IWCS Conf.* (2014) pp. 870–875

2.130 M. Tsukamoto, Y. Hoshino, N. Okada: Development of low friction indoor cable. In: *Proc. 58th IWCS Conf.* (2009) pp. 30–35

2.131 M. Takami, I. Sakabe, H. Ishikawa, H. Sotome, K. Aihara, Y. Suzuki, M. Yamano, M. Seita, T. Yamauchi, M. Hamada: New design of thinner, low-friction indoor drop cable for MDUS. In: *Proc. 58th IWCS Conf.* (2009) pp. 36–41

2.132 K. Endoh, T. Sayama, D. Takeda, S. Tanaka, M. Ohno, S. Shiobara, N. Okada, M. Miyamoto: Development of the small size low-friction indoor cable. In: *Proc. 58th IWCS Conf.* (2009) pp. 303–308

2.133 S. Shiobara, M. Ohno, T. Fukute, D. Takeda, N. Ito, N. Okada: Development of high-flame retardance LSZH-compliant small size low-friction indoor cables. In: *Proc. 61st IWCS Conf.* (2012) pp. 517–521

2.134 L. Han, P. Shah, X. Wu, S. Schmid: Improvement of the precision (repeatability and reproducibility) of a test method to characterize microbending performance of optical fibers. In: *Proc. 60th IWCS Conf.* (2011) pp. 6–10

2.135 B. Arvidsson, P. Shah, S.R. Schmid, R. Alexandersson, A. Björk: Microbend evaluation of selected G652D & G657 fibers and ribbons before cabling. In: *Proc. 60th IWCS Conf.* (2011) pp. 61–68

2.136 IEC 60794-1-22: Basic optical cable test procedures – Environmental test methods (2012)

2.137 M. Sanghvi, H. Nee: Dielectric rodent resistance cable. In: *Proc. 61st IWCS Conf.* (2012) pp. 246–271

2.138 M. Ohno, D. Takeda, T. Sayama, S. Shiobara, N. Okada: Development of the new cicada resistant aerial optical drop cables. In: *Proc. 59th IWCS Conf.* (2010) pp. 164–169

2.139 K. Matsuhashi, Y. Uchiyama, T. Kaiden, M. Kihara, H. Tanaka, M. Toyonaga: Fault cases and countermeasures against damage caused by wildlife to optical fiber cables in FTTH networks in Japan. In: *Proc. 59th IWCS Conf.* (2010) pp. 274–278

2.140 K. Shiraishi, Y. Maehara, O. Inoue, K. Takamizawa: Development of economical cicada resistant optical drop cables. In: *Proc. 59th IWCS Conf.* (2010) pp. 279–283

2.141 M. Tsukamoto, E. Konda, Y. Hoshino, N. Okada, M. Aragaki, K. Mizuno: Development of cicada-resistant optical drop cable. In: *Proc. 59th IWCS Conf.* (2010) pp. 284–290

2.142 K. Fujimoto, Y. Suzuki, M. Yamano, Y. Nagao, H. Ishikawa, H. Sotome, K. Aihara, T. Fujii: Effectiveness of optical fiber drop cable with low friction and high abrasion-resistant feature. In: *Proc. 59th IWCS Conf.* (2010) pp. 291–296

2.143 T. Sato, K. Kawase, T. Sugano: Development of cicada proof optical drop cable. In: *Proc. 59th IWCS Conf.* (2010) pp. 297–302

2.144 C. Wei, S. Yang, R. Xia, W. Xiao, B. Xiao: A study on optimized design of an optical fiber cable metallic armoring layer for squirrel resistance. In: *Proc. 59th IWCS Conf.* (2010) pp. 303–308

2.145 Recommendation ITU-T L.161: Protection of telecommunication cables and plant from biological attack (2000)

2.146 R.H. Norris, H.M. Kemp, T.G. Goddard: The validity of emerging test techniques for the evolving outside plant cable design. In: *Proc. 56th IWCS Conf.* (2007) pp. 555–560

2.147 IEC 60794-1-21: Basic optical cable test procedures – Mechanical tests methods (2015)

2.148 G. Tosik, P. Kolodziej, M. Mirynowska: Flexible bend intensive riser cable for fast FTTH deployments. In: *Proc. 58th IWCS Conf.* (2009) pp. 303–308

2.149 J. Bör: Fire risk and safety philosophy in indoor cabling. In: *Proc. 63rd IWCS Conf.* (2014) pp. 490–494

2.150 M. Maritano, D. Ceschiat: Fire resistant optical cable. In: *Proc. 60th IWCS Conf.* (2011) pp. 52–55

2.151 J.C.V. da Silva, F.G. Corcini, R. Thomas, B.N. Aires, J.G.D. de Aguiar, C.F. Rodrigues: The evolution & challenges of dielectric optical microcables for high capacity optical fiber in Brazilian market. In: *Proc. 65th IWCS Conf.* (2016) pp. 581–587

2.152 J.L. Baucom, P. Tandon, L.A. Barrett: Installation and access of a novel stranded loose tube optical fiber outside plant micro cable design. In: *Proc. 64th IWCS Conf.* (2015) pp. 598–605

2.153 J. Quinn, O. Storaasli: 288 fiber ultra-high fiber density micro-duct cable with extreme operating performance. In: *Proc. 64th IWCS Conf.* (2015) pp. 606–611

2.154 S. Olszewski, J. Toughlian, M. Emmerich, K. Smith: High density micro cable family with next generation 200 micron fibers. In: *Proc. 64th IWCS Conf.* (2015) pp. 612–616

2.155 B.G. Risch, D.L. Collado, E.J. Bowman, G.S. Emeterio: Cable design, testing, and materials for enhanced fiber access MDU cables. In: *Proc. 63rd IWCS Conf.* (2014) pp. 883–888

2.156 O. Tatat, J.M. Testu, F. Chirita, W. Greven: Combination of innovative micromodules cable designs with different types of single mode fibers: Performance and applications. In: *Proc. 62nd IWCS Conf.* (2013) pp. 160–164

2.157 O. Tatat, J.M. Testu, A. Lavenne: A further step in diameter reduction of high fiber count duct cables for FTTH applications taking full benefit of reduced size bend resistant optical fibers. In: *Proc. 60th IWCS Conf.* (2011) pp. 391–396

2.158 P. Marelli, E. Consonni: High fibre count outdoor cables with very compact designs, dedicated to FTTH applications. In: *Proc. 58th IWCS Conf.* (2009) pp. 285–288

2.159 J. Jonker, B. van der Tuuk, P. Bindels, J. Barker: 216 fibers in a Ø8.3 mm micro duct cable. In: *Proc. 58th IWCS Conf.* (2009) pp. 308–312

2.160 A. Weiss, K. Nothofer, W. Griffioen: High fiber count microduct cables. In: *Proc. 55th IWCS Conf.* (2006) pp. 420–423

2.161 Corning: MiniXtend® micro duct cables, https://www.corning.com/jp/jp/products/communication-networks/products/fiber-optic-cable/outdoor-cables/microduct-cable.html (2019)

2.162 IEC 60794-5-10: Outdoor microduct optical fibre cables, microducts and protected microducts for installation by blowing (2014)

2.163 OFS: Microcables and blown fiber units, http://fiber-optic-catalog.ofsoptics.com/viewitems/outdoor-fiber-optic-cables/microcables-and-blown-fiber-units-2 (2019)

2.164 Recommendation ITU-T L.156: Air-assisted installation of optical fibre cables (2003)

2.165 Recommendation ITU-T L.108: Optical fibre cable elements for microduct blowing-installation application (2008)

2.166 AFL: eABF® Enterprise air-jetted fiber optic cable, http://www.aflglobal.com/Products/Fiber-Optic-Cable/eABF-Fiber-Optic-Cable/eABF-Enterprise-Air-Jetted-Fiber-Cable.aspx (2018)

2.167 IEC 60794-5-20: Outdoor microduct fibre units, microducts and protected microducts for installation by blowing (2014)

2.168 F. Sato, K. Okada, T. Hirama, K. Takeda, R. Oka, K. Takahashi: Ultra-high-fiber-count and high-density slotted core cable with pliable 12-fiber ribbons. In: *Proc. 65th IWCS Conf.* (2016) pp. 604–609

2.169 M. Kikuchi, J. Kawataka, H. Izumita, T. Kurashima, Y. Yamada: Advanced optical characteristics of SZ-stranded high-density 1000-fiber cable with rollable 80-optical-fiber ribbons. In: *Proc. 65th IWCS Conf.* (2016) pp. 588–593

2.170 T. Kaji, S. Moriya, A. Murata, K. Yamashiro, K. Osato: Development of wrapping tube cable with spider web ribbon using fiber based on ITU-T G652 D. In: *Proc. 65th IWCS Conf.* (2016) pp. 600–603

2.171 F. Sato, M. Takami, Y. Nagao, K. Takeda, H. Kato: Design of ultra-high-density 2000-optical fiber cable with pliable 8-fiber ribbons for underground deployment. In: *Proc. 64th IWCS Conf.* (2015) pp. 659–663

2.172 P. Dobbins, B. Villiger: New ultra-density fiber cable technology for FTtx and access markets using New Spider Web™ ribbon. In: *Proc. 64th IWCS Conf.* (2015) pp. 674–683

2.173 Y. Yamada, M. Kikuchi, H. Izumita, T. Kurashima: 5.2 fibers/mm^2 high-density 1000-fiber cable assembled from rollable optical fiber ribbon. In: *Proc. 62nd IWCS Conf.* (2013) pp. 592–596

2.174 AFL: Wrapping tube cable with SWR®, http://www.aflglobal.com/Products/Fiber-Optic-Cable/Wrapping-Tube-Cable/Wrapping-Tube-Cable-(WTC)-with-SWR.aspx (2018)

2.175 Corning: Direct buried cables, https://www.corning.com/jp/jp/products/communication-networks/products/fiber-optic-cable/outdoor-cables/direct-buried-cable.html (2019)

2.176 SEL: Freeform Ribbon™ indoor armored RoHS riser, https://www.sumitomoelectric.com/product/freeform-ribbon-indoor-armored-rohs-riser/ (2020)

2.177 Recommendation ITU-T L.100: Optical fibre cables for duct and tunnel application (2015)

2.178 IEC 60794-3-10: Family specification for duct, directly buried and lashed aerial optical telecommunication cables (2015)

2.179 Recommendation ITU-T L.101: Optical fibre cables for buried application (2015)

Takashi Sasaki

Innovation Core SEI, Inc.
San Jose, CA, USA
sasaki@sei-innovation.com

Takashi Sasaki received his ME degree from Tohoku University in 1994. He joined Sumitomo Electric Industries, Ltd. in 1994 and has since been engaged in research and leading optical fiber and optical products development projects. Since 2016 he has been with Innovation Core SEI, Inc.

Takemi Hasegawa

Optical Communications Laboratory
Sumitomo Electric Industries, Ltd.
Yokohama, Japan
hase@sei.co.jp

Takemi Hasegawa received his ME degree in Electronic Engineering from the University of Tokyo in Japan, in 1999. In 1999, he joined Sumitomo Electric Industries, Ltd. Since then he has been engaged in R&D on optical fibers and photonic sensors. He is currently a group manager engaged in R&D of new optical fibers. He received the 2002 Hasunuma Prize from The Society of Instrument and Control Engineers.

Hiroki Ishikawa

Optical Communications Laboratory
Sumitomo Electric Industries, Ltd.
Yokohama, Japan
ishikawa-hiroki@sei.co.jp

Hiroki Ishikawa received his ME degree from Tohoku University in 1990. He joined Sumitomo Electric Industries, Ltd. in 1990 and has since been engaged in research and development of optical fiber cables. He is Senior Assistant General Manager of the Optical Communications Laboratory.

3. Optical Amplifiers

Michael Vasilyev, Stojan Radic

The principles, design, and operation of erbium-doped and Raman amplifiers, two of the most important classes used in modern lightwave communication, are described. Developed over two decades, erbium-doped fiber devices act as lumped optical gain elements in terrestrial, submarine, and access networks, underpinning nearly all commercial data traffic today. Raman amplifiers have allowed significant reach and capacity increases and, unlike erbium-doped devices, are not confined to a specific lightwave band. In contrast to alternatives such as parametric and semiconductor amplification technology, erbium-doped and Raman amplifiers have been commoditized and can be readily designed and constructed from a mature set of components that includes specialty fibers, semiconductor pumps, dedicated filters, and passive elements. The design of both types of amplifiers is described, along with the most important engineering rules that allow for optimal device construction. Mitigation of noise and distortion mechanisms is detailed for both types of amplifiers when operating with commercial fiber plants.

Similar to their electronic counterparts, optical amplifiers are intended to regenerate signal power and can be classified according to the role they play in lightwave networks. An end-of-the-link gain element, referred to as a preamplifier, is used to amplify a weak channel prior to its conversion from the photonic to electronic domain. This type provides high gain (> 30 dB), operates with a small-signal input ($P_{in} < -25$ dBm), and adds minimal excess noise. Repeatered optical transmission relies on in-line amplifiers, intended to regenerate signal power after each fiber span. An in-line amplifier must match span-induced and in-line element loss, typically providing gains greater than 20 dB. In contrast to a preamplifier, which can be dedicated to a single channel, an in-line amplifier is designed and operated as a multichannel gain block in wavelength-division-multiplexed (WDM) links. Finally, a booster (or power amplifier) is deployed at the beginning of the WDM transmission to launch elevated channel power and preserve signal integrity deep into the lightwave link. In general, these designations are reserved for lumped gain elements, characterized by an internal optical path that is much shorter than the typical transmission span. Conversely, distributed amplification takes place over distances that are comparable to the transmission span, as discussed in Sect. 3.2.

In addition to this functional distinction, optical gain elements can also be categorized according to the underlying physical platform used to achieve photon amplification. During the four decades that mark the evolution of modern optical networks, amplification using semiconductor, rare-earth-doped, Raman, Brillouin, and parametric platforms has been investigated. While these possess unique sets of distinguishing characteristics, only two types, viz. rare-earth doped and Raman amplifiers, have seen widespread deployment in commercial optical networks. Consequently, this chapter focuses on these two technologies and describes the principles, operating characteristics, design, and performance limits of Raman and erbium-doped fiber amplifiers (EDFA). Indeed, while semiconductor, Brillouin, and parametric amplifiers offer performance in terms of response time, band selectivity, and phase-sensitive processing that neither EDFA nor Raman amplifiers can match, they are not used at present, nor are they expected to be deployed in near-future optical networks.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_3

To proceed, we first define commonly used parameters applicable to any optical gain element, regardless of the underlying physical mechanism responsible for signal amplification. The optical gain (G) is defined as the ratio between the output and input optical signal powers

$$G = \frac{P_{\text{out}}}{P_{\text{in}}} . \tag{3.1}$$

This relation expresses the gain on a linear scale, while amplifier design often calls for a logarithmic (dB) scale – the remainder of chapter will make this distinction explicit in specific cases. In its simplest form, the gain definition in (3.1) implies a quasimonochromatic signal. However, when the signal has a nonnegligible spectral width, a more general expression must be used, including the gain spectral dependence, often referred to as gain ripple.

The noise figure (NF) is defined by the input and output signal-to-noise ratios (SNR), under the assumption that the input signal possesses the maximal (shot-noise limited) *SNR*

$$\text{NF} = 10 \log \left(\frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}} \right) . \tag{3.2}$$

In accordance with convention, noise figures expressed on a linear scale are denoted by F throughout Sect. 3.1. In Sect. 3.2, the notation *NF* will be used for both linear- and log-scale noise figures. In analogy with electrical amplifiers, the gain bandwidth represents the frequency (wavelength) range over which a predefined (minimal) gain is maintained. However, in contrast to electrical amplifiers, the optical gain elements of interest (EDFA and Raman) possess bandwidths measured on the THz rather than GHz scale. As an example, the spectrally equalized EDFA bandwidth is engineered between 1528 and 1620 nm (≈ 10 THz) – nearly two orders of magnitude wider than ultrawideband electrical amplifier responses. Finally, the gain saturation characterizes the compression mechanism that applies to all types of optical amplifier. This effect originates from the inability to amplify an arbitrarily large signal power, signifying the difference between small- and large-signal levels. Specifically, gain saturation occurs when the input signal is large enough to appreciably deplete the inversion level of the gain medium [3.1, 2]. Gain saturation is a direct function of the signal power and is defined by

$$g = \frac{g_{\text{s}}}{1 + (P/P_{\text{S}})} , \tag{3.3}$$

where g_{s} is the small-signal gain measured on the dB scale, while the saturation power P_{S} indicates the input signal power at which the amplifier gain is compressed by 3 dB. Note that the simple definition (3.3) also implies a lumped (two-port) amplifier but does not account for spectral or spatial gain variation in a finite-length amplifier device.

3.1 Erbium-Doped Fiber Amplifiers

EDFAs represent the dominant amplifier technology that drives all aspects of optical networks, from short-range (access/metro) to ultralong-haul links in submarine transmission. Commercially introduced more than quarter of a century ago, EDFAs currently power both uni- and bidirectional transmission architectures, as shown in Fig. 3.1.

Modern fiber amplifiers operate within the 1550-nm transparency band of silica and are the result of multiple technological advances that matured simultaneously during the 1990s. First, extensive research on rare-earth (lanthanide) emission led to Er^{3+}-, Nd^{3+}-, and Pr^{3+}-doped fibers [3.3–5] that combine high gain coefficients, simple interfacing with passive (transmission) fibers, and compatibility with semiconductor pump technology. The energy levels of Er^{3+}, shown in Fig. 3.2, with its inherent 1550-nm transition, have served as the impetus for EDFA development. Er^{3+} emission/absorption covers the minimum-loss transmission window in silica fibers almost entirely and currently defines the most important transmission bands, viz. C-band (1530–1562 nm) and L-band (1565–1610 nm). C/L-band amplification in Er-doped fibers is supported by both short- (980 nm) and long-wavelength (1480 nm) semiconductor pump diodes that match the $^4I_{11/2}$ and $^4I_{13/2}$ Er^{3+} transitions required for 1550-nm signal amplification. Multiple silica doping and EDF fabrication techniques [3.2] have been developed, enabling wide variation in the Er concentration which governs the emission (absorption) strength, variable mode confinement, and negligible passive (background loss). As a result, these advances have enabled the realization of high-gain devices in only a few meters of doped fiber and in any band spanning 1530–1610 nm. With the maturation of doped fiber fabrication technology, the EDFA designer can demand not only specific emission (absorption) levels but also tailored dispersive and birefringent characteristics. As a consequence, lightwave networks currently deploy a wide variety of isotropic (non-birefringent) and polarization-maintaining (PM) EDFAs, with gains that vary from 10 to 50 dB and output powers that range from a few mW to over a watt.

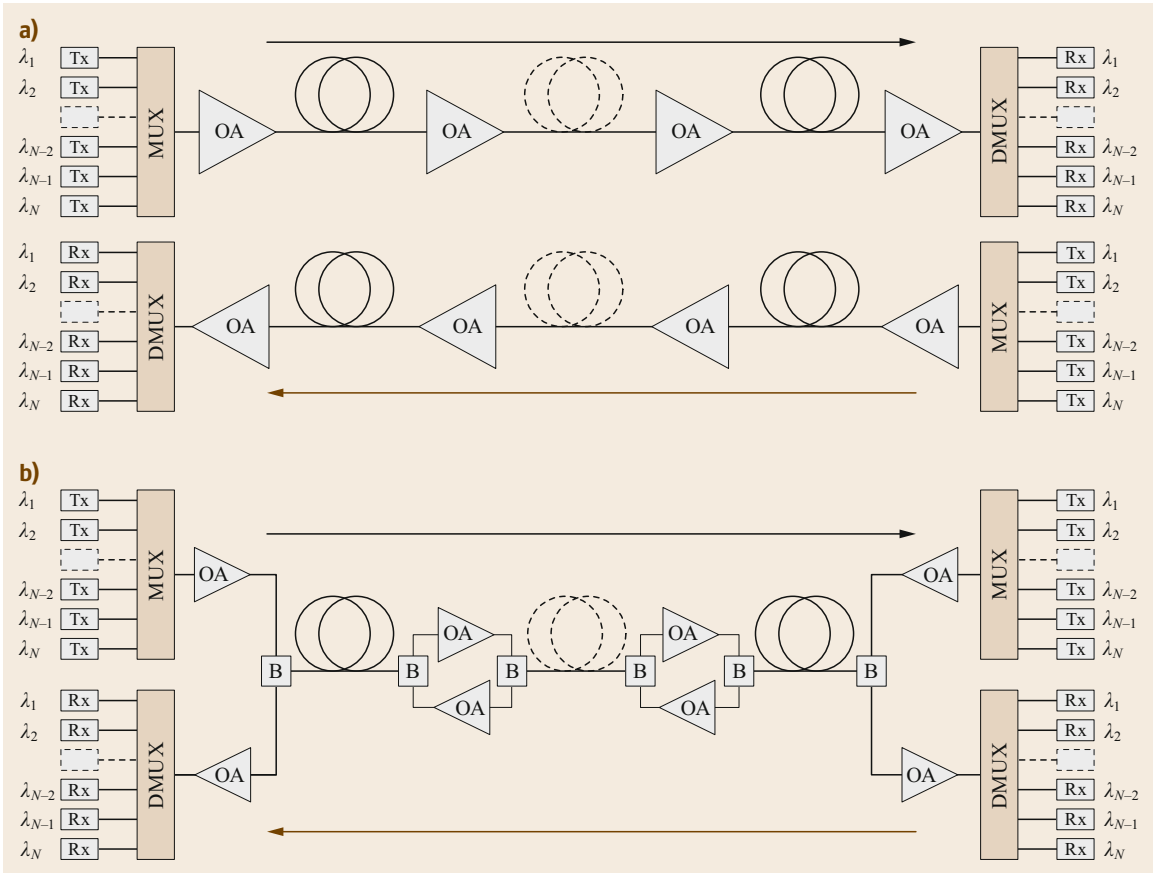


Fig. 3.1a,b Repeatered amplified WDM lightwave transmission: Channels are transmitted over a unidirectional architecture that uses a dedicated fiber for transmission in each direction **(a)** or over a bidirectional topology **(b)** that allows counterpropagating traffic to share a single fiber strand

3.1.1 EDFA Design and Characterization

In cases where an EDFA operates in the stationary regime, its gain and noise response can be accurately described by a two-level model, neglecting the third-level population [3.7–9], as illustrated in Fig. 3.3. The use of such a two-level inversion model is justified in most cases of practical interest and assumes, among other conditions [3.10], that the input signal and pump powers are static or fluctuate on faster than a microsecond scale. Under this approximation, the third-level ion population is negligible and decays rapidly, justifying the simple inversion model that accounts only for single upper (N_2) and lower (N_1) levels. Surprisingly, while this approximation appears to be simplistic, it is also remarkably accurate for nearly all WDM amplifier architectures of interest utilizing subwatt pump levels and Er-doped fibers with peak absorption below 30 dB/m.

The two-level inversion model allows an Er-doped waveguide to be considered as a distributed amplifier that can be easily described by balancing the local emission and absorption rates. In a stationary state, any mechanism leading to decay from the upper-level population (N_2) is offset by excitation from the ground level along the length of the fiber. The rate of each process varies along the fiber and changes with the local signal, pump, and noise levels. More importantly, the inversion level and photon count are closely coupled: to describe the longitudinal inversion variation, full knowledge of all three radiation types must be available; moreover, the longitudinal inversion function must be known to calculate the signal, pump, and noise levels. Figure 3.4 illustrates the relation between the spatially dependent inversion and distributed optical power levels. In the general case, the forward- and backward-propagating beams have temporal, spectral, and spatial dependence: $P_{\pm} = P_{\pm}(z, \lambda, t)$. In the sta-

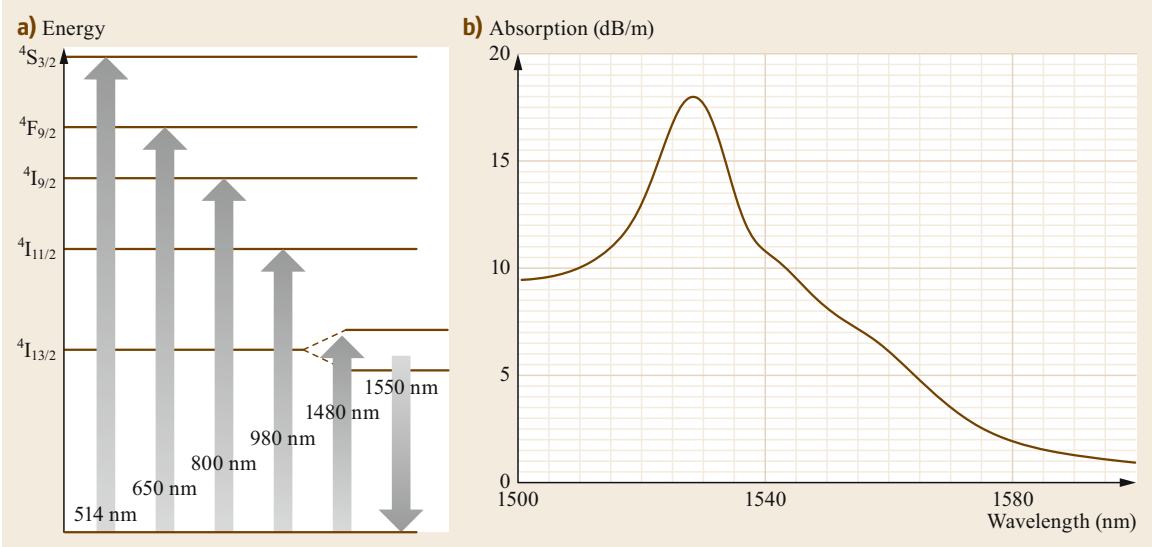


Fig. 3.2 **(a)** Er^{3+} energy level diagram, indicating viable pump and amplification bands. Fine (Stark) level splitting of the $^4I_{13/2}$ level is responsible for the large tolerance in 1480-nm pumping and allows for a wide emission band in the low-loss silica band (1550 nm); **(b)** absorption of a commercial Er-doped fiber [3.6]

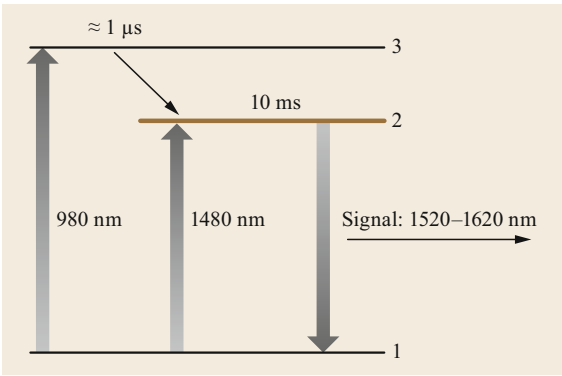


Fig. 3.3 In the two-level approximation, the significant discrepancy between lifetimes is used to justify the assumption of a vanishing upper-level population ($N_3 \to 0$), allowing the total ion population to be distributed between the ground and excited level: $N_T = N_1 + N_2$

tionary case that is of primary interest in EDFA design, the rate equation coupling the inversion level and counterpropagating beam powers is particularly simple

$$\frac{N_2(z)}{\tau_0} = -\frac{1}{\rho A}\sum_{i=1}^{N} u_i \frac{\mathrm{d}P_i(z)}{\mathrm{d}z} , \qquad (3.4)$$

where N_2 represents the normalized upper-level ion population, τ_0 is the spontaneous lifetime of the upper level, A is the waveguide effective cross-section area, and ρ is the ion density. This relation implies spectral beam dependency by channelizing propagating beams ($i = 1, N$), where the index corresponds to a specific wavelength (frequency). The ion population is normalized and is defined by scaling both by the total ion count: $N_1 + N_2 = 1$. Finally, the directional coefficients ($u_i = \pm 1$) account for the propagation direction along the z-axis, as defined in Fig. 3.4. Note that the left term in (3.4) represents decay from the upper level, while the right term accounts for signal- and noise-induced depletion and pump-induced repletion of the upper-level population.

The signal, pump, and noise power levels are directly governed by the local ion population in the ground and excited levels

$$\frac{\mathrm{d}P_n(z)}{\mathrm{d}z} = u_n\gamma_n N_2(z)P_n(z) - u_n\alpha_n N_1(z)P_n(z) , \qquad (3.5)$$

where γ_n and α_n are the emission and absorption constants of the EDF [3.10], often referred to as Giles coefficients. Note that (3.5) has a straightforward interpretation: The photon annihilation (generation) rate is directly proportional to the population of the ground (upper) level and the local photon flux. This relation is often written in a form that emphasizes only the upper-level dependence

$$\frac{\mathrm{d}P_n(z)}{\mathrm{d}z} = u_n(\gamma_n+\alpha_n)N_2(z)P_n(z) - u_n\alpha_n P_n(z) . \qquad (3.6)$$

By simultaneously solving (3.4) and (3.6), it is possible to obtain the inversion, signal, pump, and noise levels at any point along an Er-doped waveguide. While multiple

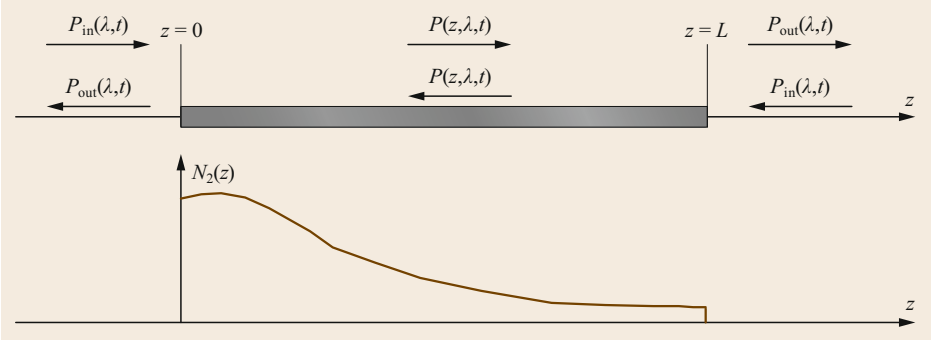


Fig. 3.4 Two-port, two-level bidirectional EDF model coupling signal, noise, and pump levels (P) and the inversion population (N_2)

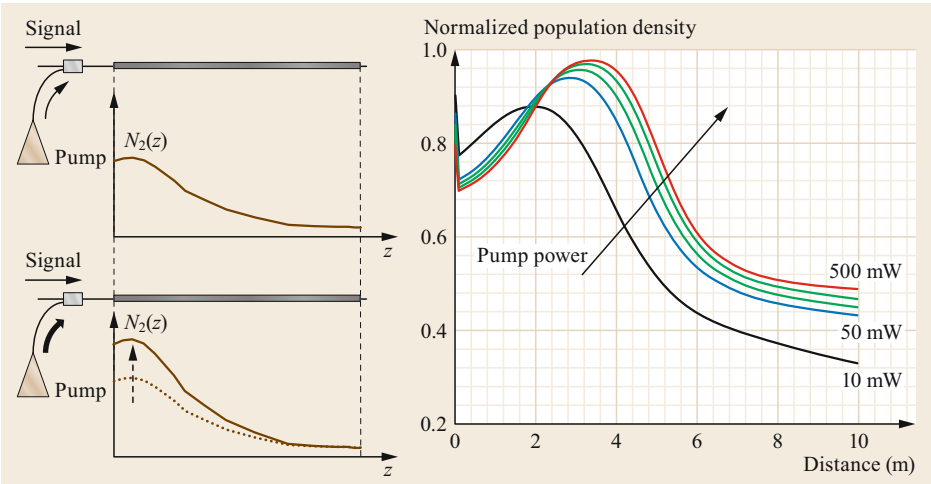


Fig. 3.5 Inversion dependence on pump power: An increase in pumping level results in a simultaneous increase of the inversion $N_2(z)$ peak and its longitudinal shift along the EDF section

techniques for analytic and semianalytic solutions have been developed in the past [3.8, 9, 11], rigorous numerical solutions are typically preferred during the EDFA design process.

Such two-level modeling of the localized inversion and photon flux provides important insights into EDFA operation regimes, which can then be used to create a design rule framework for the analysis and construction of complex, multistage amplifier subsystems. To develop an intuitive notion of EDFA behavior, consider the interdependence that exists between the pump/signal levels and the population inversion function $N_2(z)$. For the former, consider a pump level variation while the signal input is kept constant, as illustrated in Fig. 3.5. Since the pump provides the sole mechanism for ion excitation, it is naturally expected that the inversion will increase monotonically with the pump power.

While valid, this assumption is also somewhat simplistic: the ion inversion peak increase that occurs with an elevated pump power is also accompanied by a longitudinal shift of the peak along the fiber. This behavior cannot be accounted for by considering only the signal–pump interaction, as the amplitude spontaneous emission (ASE) in both propagation directions must also be included. However, Fig. 3.5 still provides one of the most important notions related to EDFA design: An increase in both the front-end and total inversion level must be guaranteed by a higher pumping level.

The second EDFA design notion relates to its response to an increased signal level, as illustrated in Fig. 3.6.

Assuming that the input signal is increased while the pump level is maintained at a constant level, it is reasonable to expect that the average inversion will decrease as stimulated emission grows. Again, this intuitive notion, while generally correct, is also an oversimplification; instead, an increase in the signal level will result in a total decrease of excited ions, but the local inversion level will be depleted much faster, shifting the peak of the inversion function peak closer to the EDF input. Similar to the case in which the pump is varied, an accurate description requires that the model for the signal–pump interaction also accounts for ASE generation.

A further insight into EDFA operation and response prediction can be gained by recognizing the unique role

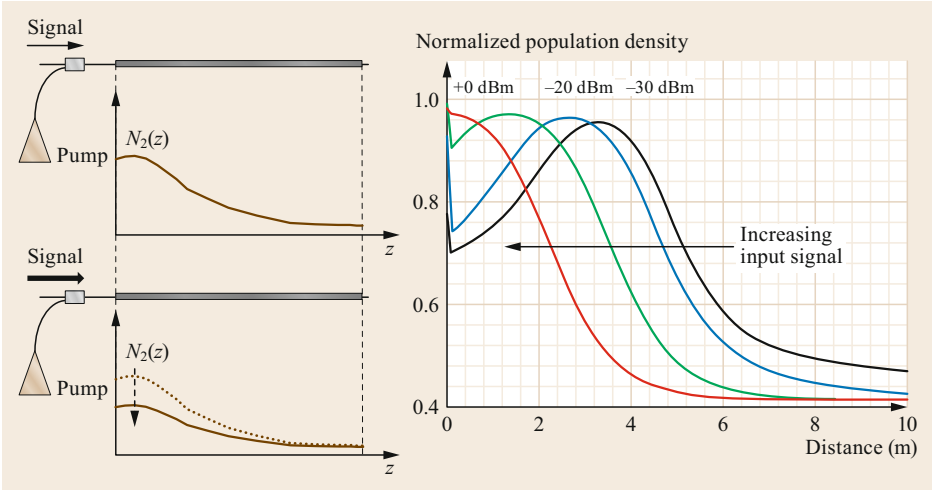


Fig. 3.6 Inversion dependence on input signal power: An increase in input level results in a decrease in the total inversion level N_2, resulting in rapid depletion along the EDF length

played by average rather than localized (spatially dependent) inversion [3.12]. This fact is counterintuitive and requires revisiting (3.4) and (3.6), after introducing the notion of the average inversion parameter $\overline{N}_2$

$$\overline{N}_2 = \frac{1}{l}\int_0^l N_2(z)\mathrm{d}z\,, \tag{3.7}$$

where l is the length of the Er-doped waveguide. Integration of (3.4) along the entire EDF length yields

$$\begin{aligned}\frac{1}{\tau_0}\int_0^l N_2(z)\mathrm{d}z &= -\frac{1}{\rho A}\int_0^l \sum_{i=1}^{N} u_i \frac{\mathrm{d}P_i(z)}{\mathrm{d}z}\mathrm{d}z \\ &= -\frac{1}{\rho A}\sum_{i=1}^{N}(P_{i,\mathrm{out}} - P_{i,\mathrm{in}})\,. \end{aligned} \tag{3.8}$$

Now, defining the average gain as

$$\overline{g}_n = \frac{1}{l}\ln\frac{P_{i,\mathrm{out}}}{P_{i,\mathrm{in}}}\,,$$

it is possible to relate the average inversion level to the EDFA gain and input optical powers by the simple relation

$$\frac{1}{\tau_0}\overline{N}_2 = -\frac{1}{l\rho A}\sum_{i=1}^{N} P_{i,\mathrm{in}}(\mathrm{e}^{\overline{g}_n l} - 1)\,. \tag{3.9}$$

The average gain, also referred to as the global (port-to-port) gain, can be measured easily to provide the EFFA average inversion level. However, during the design process, the output power levels are not known in advance and must be replaced by physical parameters tied to the EDF. This can be accomplished by integrating (3.6) along the entire EDF length

$$\int_0^l \frac{\mathrm{d}P_n(z)}{\mathrm{d}z}\mathrm{d}z = u_n\int_0^l [(\gamma_n + \alpha_n)N_2(z) - \alpha_n]P_n(z)\mathrm{d}z\,, \tag{3.10}$$

which when combined with the definition of the average gain $\overline{g}_n$, results in the simple inversion–gain relationship

$$\overline{g}_n = (\gamma_n + \alpha_n)\overline{N}_2 - \alpha_n\,. \tag{3.11}$$

This allows a single, master EDFA equation to be written only in terms of the average inversion [3.12]

$$\overline{N}_2 = -\frac{\tau_0}{lA\rho}\sum_{i=1}^{N} P_{i,\mathrm{in}}(\mathrm{e}^{[(\gamma_i+\alpha_i)\overline{N}_2 - \alpha_i]l} - 1)\,. \tag{3.12}$$

It is difficult to overstate the importance of (3.12), since its solution not only allows for complete EDFA analysis but also provides important insights that aid the amplifier design process. With the emission (absorption) parameters, uniquely defined by the selected Er doping level and specified input signal/pump powers, one can, at least in principle, calculate the average inversion level in any EDFA operating regime. While $\overline{N}_2$ is a single, scalar parameter, its knowledge uniquely defines the EDFA response across the entire spectral range. Remarkably, even though the above derivation discards the information on *local* inversion in order to calculate the *global* ion inversion, the amplifier gain, spectral response, and noise characteristics can still be calculated in an unambiguous manner.

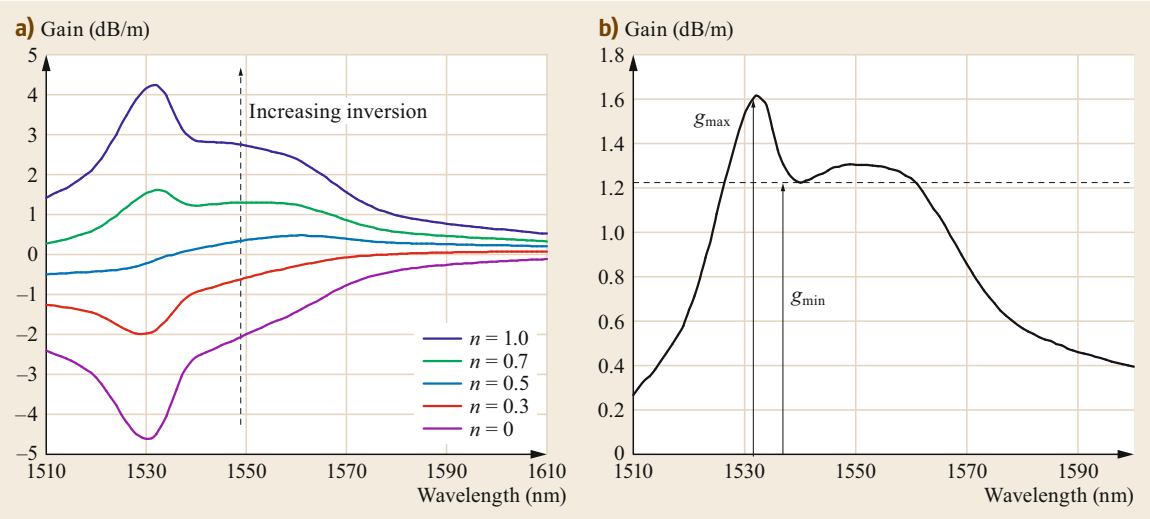


Fig. 3.7 **(a)** EDF gain dependence on average inversion level. **(b)** The selection of a specific inversion value uniquely defines the gain ripple and total attenuation that must be applied to equalize the EDFA spectral response $g_{max} - g_{min}$

Knowledge of $\overline{N}_2$ also allows the calculation of the EDFA gain at any wavelength of interest. Indeed, it is sufficient to note that (3.11) implies an ability to calculate the gain at an arbitrary wavelength as

$$G(\lambda_n) = \{[\gamma(\lambda_n) + \alpha(\lambda_n)]\overline{N}_2 - \alpha(\lambda_n)\}l\,. \qquad (3.13)$$

A direct and practical implication is that one only needs to measure the gain at a single wavelength in order to calculate the EDFA gain at any other wavelength, commonly referred to as gain ripple. Indeed, assuming that the gain is measured at λ_n, then the gain at any other wavelength λ_m is given by

$$G(\lambda_m) = \frac{[\gamma(\lambda_m) + \alpha(\lambda_m)]}{[\gamma(\lambda_n) + \alpha(\lambda_n)]} \times [G(\lambda_n) + \alpha(\lambda_n)l] - \alpha(\lambda_m)l\,. \qquad (3.14)$$

However, one also needs to be cautioned that the predictive accuracy of (3.8) and (3.9) greatly depends on that of the emission and absorption parameters. The last relation is also often used for the characterization of black-box EDFA devices, where no knowledge about the EDF length is available and one only has access to the input/output ports with a single, fixed-wavelength laser. To illustrate its utility, let us assume that a gain increase (decrease) has occurred, resulting in a change in the average inversion of the amplifier. In this case, the gain change measured at the probe wavelength (λ_p) is

$$\Delta G(\lambda_p) = [\gamma(\lambda_p) + \alpha(\lambda_p)]\Delta\overline{N}_2 l\,. \qquad (3.15)$$

The gain difference at any other wavelength can then be calculated as [3.12]

$$\Delta G(\lambda) = \frac{[\gamma(\lambda) + \alpha(\lambda)]}{[\gamma(\lambda_p) + \alpha(\lambda_p)]}\Delta G(\lambda_p) = T(\lambda)\Delta G(\lambda_p)\,, \qquad (3.16)$$

where $T(\lambda)$ is referred to as the gain tilt function. Consequently, if the EDFA is regarded as a black-box, two-port module, it is only necessary to measure the port-to-port gain response at two wavelengths to obtain the gain tilt function and subsequently determine the full gain spectrum.

Determining the average inversion plays a crucial role in any EDFA design, and its importance can be easily visualized by plotting the average (unit-length) gain as a function of the normalized inversion level, as shown in Fig. 3.7.

In a general EDFA design problem, a target gain, bandwidth, and noise response are given, subject to predetermined pump and signal budgets. The design process is initiated by selecting the specific inversion level, under the assumption that all the physical constraints can be met. To illustrate the iterative design methodology, let us assume that a line-amplifier element should generate 25 dB of gain in the C-band ($1530 < \lambda < 1560$ nm). In the first step, one chooses the average inversion level and uniquely defines a native (unequalized) spectral response of the amplifier, as illustrated in Fig. 3.7b. At this inversion level, the minimum gain ($g_{min} = 1.21$ dB/m) occurs at 1542 nm, dictating that the EDF length must be at least 25 dB/(1.12 dB/m) = 20.67 m. At the same time, the maximal unit gain

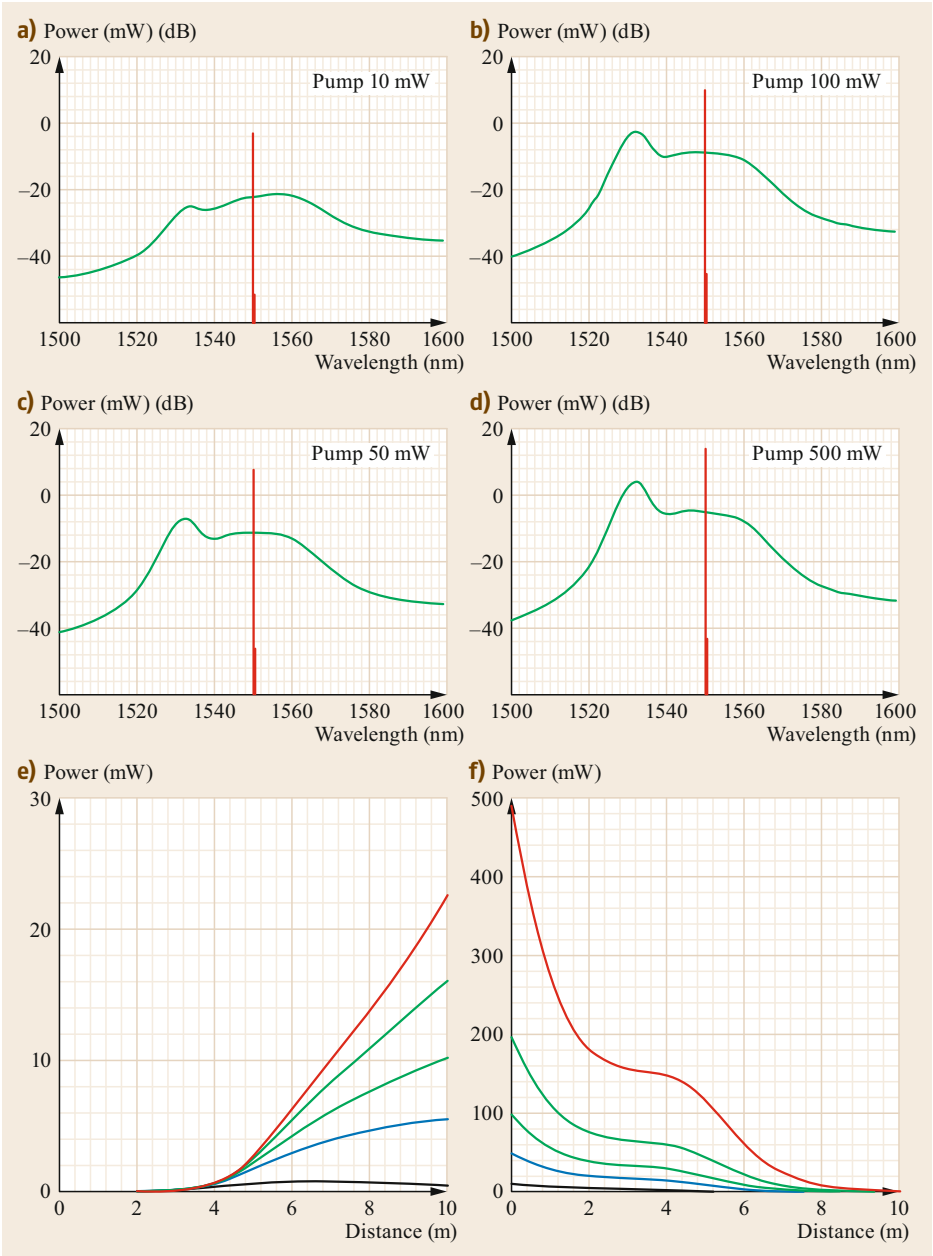


Fig. 3.8a–f Pump power and inversion: An increased pump level guarantees a higher signal gain **(a–d)** (signal at 1550 nm, *red curve*) and higher inversion level, as indicated by ASE (*green curve*). However, an increased pump level also changes the effective EDF length, as indicated by the signal **(e)** and pump **(f)** evolution

($g_{max} = 1.67\,dB/m$) will result in a peak amplifier gain of 34.52 dB, inducing a gain ripple of 14.52 dB.

The latter imposes the first decision during the design process; if such a ripple is not acceptable (because it exceeds the maximal loss of the gain-flattening filter or compromises the EDFA noise performance), a new inversion level is selected to reduce the gain ripple. Even if the average inversion level meets the ripple requirement, the design must still meet the available pump budget. Indeed, the desired inversion level also defines the required pump budget, which becomes progressively higher with increasing EDF length and signal power. To illustrate this challenge, consider a fixed EDF section ($l = 10\,m$), intended to amplify a typical signal level at the end of the transmission span ($P_{in} = -25\,dBm$). Figure 3.8 shows the progressively higher pump requirement as the average inversion is increased in order to meet the high gain condition. In this particular example, a gain increase from 20 to 35 dB mandates a 50-fold (17 dB) pump increase, indicating

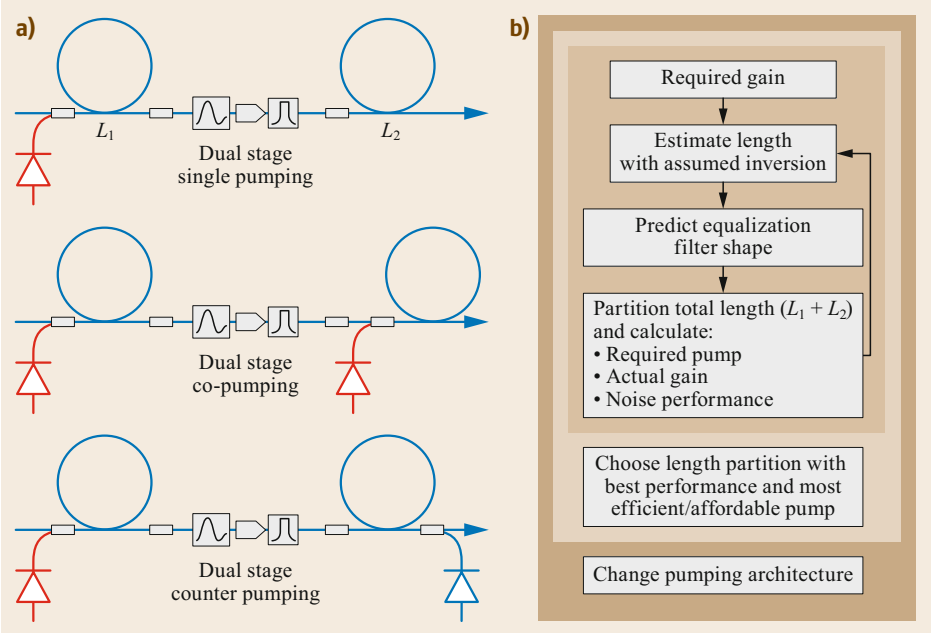


Fig. 3.9a,b Multistage EDFA design as iterative process: multiple pumping strategies and ASE management architectures **(a)** must be evaluated during a physical parameter search that include pump powers, EDF section lengths and partitioning, and feasibility of ASE filter implementation **(b)**

an approximate scaling law, when ASE generation is discounted.

Unfortunately, the design considerations discussed so far are applicable only to single-stage EDFA devices but do not correspond to practically useful gain blocks. Indeed, all three types of amplifier (preamplifier, line amplifier, and boosters) are designed and constructed using multiple EDF stages in order to manage ASE generation, impose local control over inversion levels, manage crosstalk and impairments, bury passive optical loss, and distribute pumps efficiently. As a result, the design process becomes considerably more complex, even in the case of simple, two-stage devices, as illustrated in Fig. 3.9. This flowchart illustrates a general design methodology, albeit still not incorporating crosstalk management or distortion suppression, that provides a systematic path to achieve the desired EDFA performance.

However, in practice, the design challenge is considerably greater, since amplifiers comprise more than two EDF stages. Figure 3.10 illustrates an example of a banded line amplifier capable of supporting 27 dB port-to-port gain along with buried (interstitial) loss in order to accommodate wavelength add/drop or dispersion-compensating modules. In this case, the parameter search space spans four distinct EDF segments that share co- and counterdirectional pumping and ASE management. While the detailed design strategy is considerably more complex than the procedure outlined in Fig. 3.9, it shares the common traits of first setting the average inversion, followed by iterative attempts to maintain the spectral equalization and gain targets.

3.1.2 Noise Characteristics

In addition to its spectral and gain response, an EDFA amplifier should be designed to generate minimal noise under given physical constraints. In the context of lightwave communications, this means that the signal integrity should not be substantially degraded during the amplification process, by either additive mechanisms such as ASE or intraamplifier nonlinear distortions such as cross-phase modulation or four-wave mixing. In nearly all practical cases of interest, EDFA noise is dominated by ASE generation and is uniquely defined by the input signal and the inversion level of the amplifier. Comprising spontaneously emitted photons, ASE manifests as forward- and backward-propagating fields that are both unpolarized (in a non-PM EDF) and uncorrelated to the input signal. ASE occurs due to the large number of Er ion emitters that, in the absence of cooperative effects [3.10, 13], can be modeled under the assumption of random phase and position. As a consequence, the distributed EDF ASE gives rise to nearly ideal, white noise that is easily characterized by measuring the spectral power density, distributed between the two polarizations in an isotropic (non-birefringent) EDF.

To characterize the impact of ASE on the signal, it is common to consider the field received by an ideal

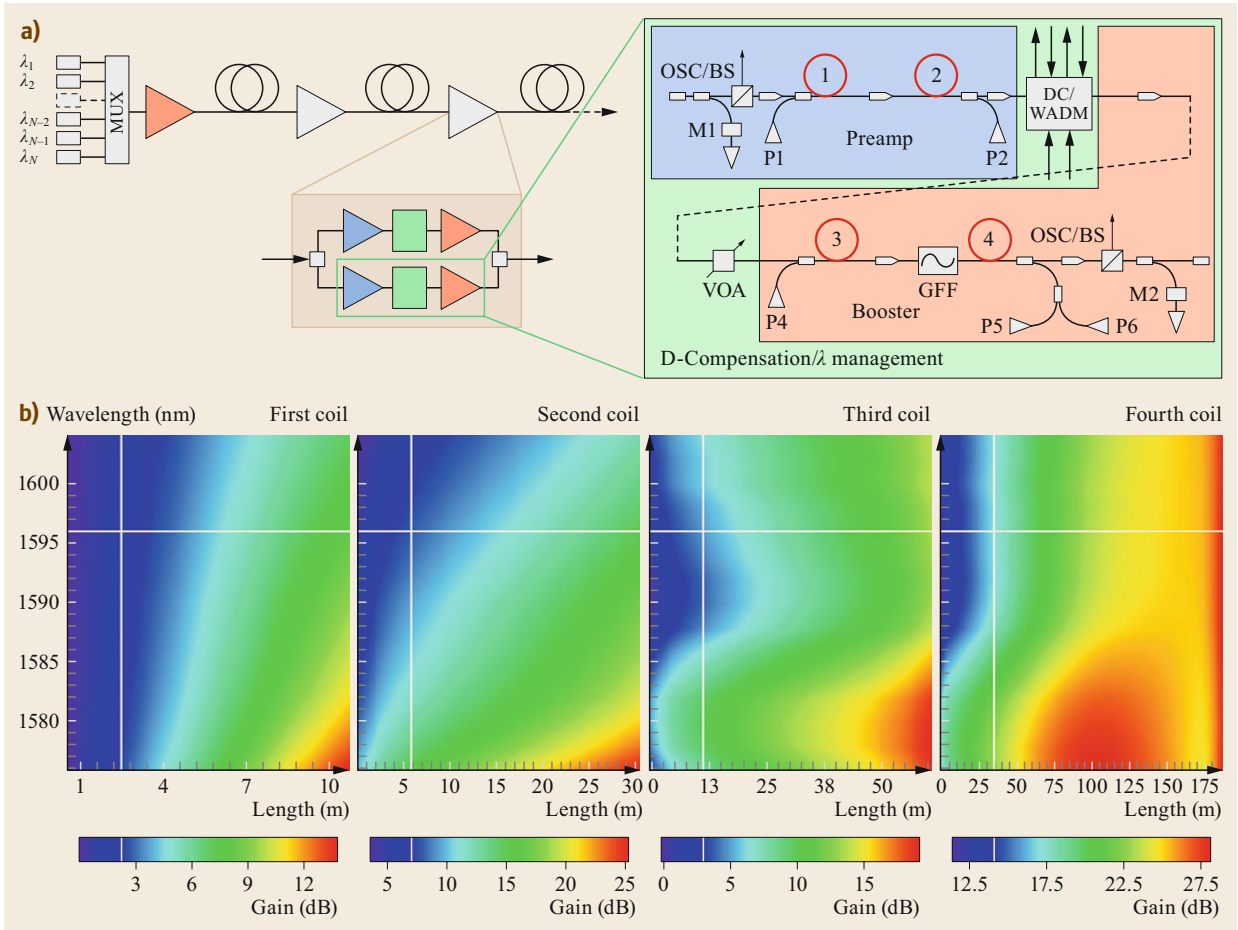


Fig. 3.10a,b Multistage EDFA with interstitial loss **(a)** and **(b)** optimal signal/pump evolution in each section

detector modeled as a band-limited, square-law opto-electronic converter. In this case, the received current can be expressed as

$$I_{\mathrm{det}}(t) = \eta \frac{1}{2T} \int_{-T}^{T} |\boldsymbol{E}_{\mathrm{sig}}(t+t') + \boldsymbol{E}_{\mathrm{ASE}}(t+t')|^2 \mathrm{d}t' \,, \quad (3.17)$$

where η is the detector efficiency, T is the detector response time, E_{sig} is the signal, and E_{ASE} is the ASE field contribution. Following the Wiener–Khintchine theorem, the spectral density of the detector current is defined by the Fourier transform of its correlation, giving rise to three distinct spectral components

$$S(f) = (P_{\mathrm{sig}} + P_{\mathrm{ASE}})^2 \delta(f) + P_{\mathrm{sig}} \rho_{\mathrm{ASE}}(f_{\mathrm{sig}}) + \int_{f_{\mathrm{sig}}-\Delta f}^{f_{\mathrm{sig}}+\Delta f} \rho_{\mathrm{ASE}}(f)^2 \mathrm{d}f \,, \quad (3.18)$$

where P_{sig} and P_{ASE} represent the average received powers, ρ_{ASE} is the ASE spectral power density, and f_{sig} is the signal frequency. The first term in (3.18) represents the total direct-current (DC) power received at the detector, the second term is readily recognized as the signal–spontaneous beat noise, while the last term is the spontaneous–spontaneous beat noise. In almost all cases of interest, channel reception is limited by the signal–ASE beat noise, allowing for a simplified quantification of the characteristics of the EDFA noise and its impact on the performance of the lightwave link. Following this approach, a general notion of the noise figure as the ratio of the input to output signal-to-noise-ratio (SNR) F can be defined as

$$F = \frac{\mathrm{SNR}_{\mathrm{in}}}{\mathrm{SNR}_{\mathrm{out}}} \,. \quad (3.19)$$

In contrast to (3.2), the noise figure is expressed on a linear rather than dB scale here. To make this defi-

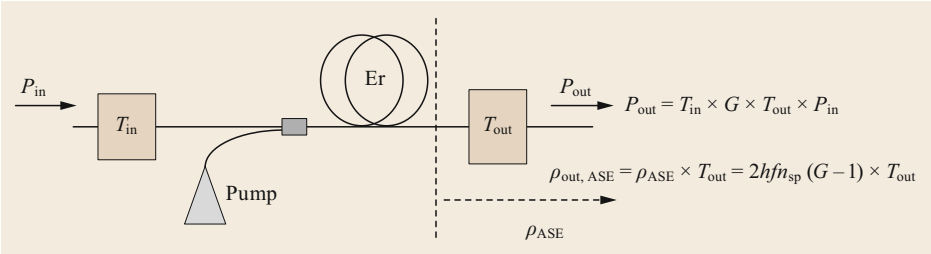


Fig. 3.11 Single-stage EDFA noise generation and noise figure, accounting for input and output insertion losses

nition valid, the input SNR must be referenced to the least-noisy (shot-noise-limited) signal

$$\mathrm{SNR}_{in} = \frac{P_{in}^2}{2P_{in}hfB_{el}}\,, \tag{3.20}$$

where h is the Planck constant, f is the signal carrier frequency, and B_{el} is the detector bandwidth. The output SNR, dominated by the signal–ASE beat, is defined as

$$\mathrm{SNR}_{out} = \frac{P_{out}^2}{P_{out}\rho_{ASE}B_{opt}}\,. \tag{3.21}$$

This last relation is written under the assumption that the bandwidth of the optical filter preceding the detector and its electrical bandwidth are matched as $B_{opt} = 2B_{el}$, avoiding mismatch correction factors. In this case, the ratio of (3.21) and (3.20) gives rise to a remarkably simple expression for the EDFA noise figure

$$F = \frac{\rho_{ASE}}{Ghf}\,. \tag{3.22}$$

Substituting the value of the ASE spectral power density ($\rho_{ASE} = 2n_{sp}(G-1)hf$), where n_{sp} represents the spontaneous emission factor [3.10], into the last equation yields the well-known expression that relates the noise figure and gain of an EDFA

$$F = 2n_{sp}\frac{G-1}{G}\,. \tag{3.23}$$

In the case when $G \gg 1$, and with near full inversion ($n_{sp} \approx 1$), the EDFA noise figure approaches 2 (or 3 dB in logarithmic terms), which is often referred to as the practical noise limit of a well-designed amplifier. Indeed, it can be shown that this limit cannot be violated for phase-insensitive amplifiers while any further reduction requires the introduction of a phase-sensitive physical process such as parametric interaction. However, in special circumstances, such as when the EDFA operates with low gain ($G < 10$), its noise figure can be lower than 2 (3 dB) without resorting to phase-sensitive amplification.

In practice, the noise figure of an EDFA must be expressed in terms of readily available physical parameters that can be measured using conventional instruments, such as the optical signal-to-noise ratio (OSNR). By expressing the OSNR of the EDFA in terms of the signal power and ASE spectral power density as

$$\mathrm{OSNR} = \frac{P_{sig}}{\rho_{ASE}\Delta f|_{B_{opt}}} \tag{3.24}$$

and eliminating the ASE spectral power density, a particularly simple expression for the output OSNR can be written

$$\mathrm{OSNR} = \frac{GP_{in}}{\rho_{ASE}\Delta f|_{B_{opt}}} = \frac{GP_{in}}{hfGF\Delta f|_{B_{opt}}}\,, \tag{3.25}$$

allowing the NF to be expressed as

$$\mathrm{NF} = P_{in} - \mathrm{OSNR} - 10\log\left(\frac{hc^2\Delta\lambda}{\lambda^3}\right). \tag{3.26}$$

A direct relationship between the OSNR, signal power, and noise figure is particularly useful when considering a practical EDFA device that must account for finite losses associated with signal insertion, pump multiplexing, and any passive elements required within the optical path. To illustrate its utility, consider a simple, single-stage EDFA amplifier, as shown in Fig. 3.11.

A composite noise figure for the gain element (the Er-doped fiber section) with accompanying passive losses can be calculated by noting that the output SNR can be expressed in terms of the amplified signal level and output ASE

$$\mathrm{SNR}_{out} = \frac{P_{out}}{2\rho_{ASE}B_{el}} = \frac{T_{in}GT_{out}P_{in}}{4hfn_{sp}(G-1)T_{out}B_{el}}\,. \tag{3.27}$$

Consequently, an aggregate noise figure can be readily calculated by combining (3.14) and (3.21)

$$F = \frac{\mathrm{SNR}_{in}}{\mathrm{SNR}_{out}} = \frac{1}{T_{in}}\frac{2n_{sp}(G-1)}{G}\,. \tag{3.28}$$

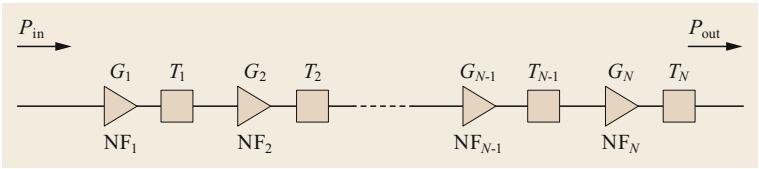


Fig. 3.12 EDFA cascade: Noise and signal power accounting

This last relation is particularly simple when written on the decibel scale

$$\mathrm{NF} = \mathrm{NF}_{\mathrm{EDFA}} - T_{\mathrm{in}}|_{\mathrm{dB}} \,, \tag{3.29}$$

where $T_{\mathrm{in}}|_{\mathrm{dB}} < 0$ represents the insertion loss on the dB scale. The last relation states that the aggregate EDFA noise figure does not depend on the output loss but does increase directly with addition of the input loss, regardless of the gain block architecture. Using a similar rationale, an aggregate noise characteristic of a cascade of EDFA gain blocks with interstitial losses, which corresponds to a typical transmission link, can be easily characterized, as shown in Fig. 3.12.

The transmitted signal power at the end of a gain-balanced ($G = 1/T$) link is $P_{\mathrm{out}} = G^N T^N P_{\mathrm{in}} \xrightarrow{G=1/T} P_{\mathrm{in}}$. The case in which each the gain block matches the subsequent loss exactly $G_{\mathrm{i}} = 1/T_{\mathrm{i}}$ is of particular interest. In such a link, the ASE spectral density emitted by the first EDFA

$$\rho_{\mathrm{ASE}}^{(1)} = 2hfn_{\mathrm{sp}}(G-1) \tag{3.30}$$

is attenuated by the first interamplifier loss (T_1) and then amplified by the second gain element (G_2)

$$\rho_{\mathrm{ASE}}^{(2)} = 2hfn_{\mathrm{sp}}(G-1)T \times G + 2hfn_{\mathrm{sp}}(G-1) \,. \tag{3.31}$$

By propagating the ASE spectral density along the entire link, the final (end-of-link) spectral density can be expressed as

$$\rho_{(\mathrm{out}),\mathrm{ASE}} = 2hfn_{\mathrm{sp}}(G-1)T \sum_{i=1,N} (GT)^N \xrightarrow{G=1/T} 2hfn_{\mathrm{sp}}(G-1)NT \,, \tag{3.32}$$

leading to a simple expression for the aggregate noise figure of the link

$$F = 2n_{\mathrm{sp}}(G-1)NT \,. \tag{3.33}$$

While simple, this last relationship has important implications: it states that the aggregate noise figure of a transparent transmission link grows directly with the number of gain elements, as expected intuitively.

3.1.3 Distortion Mechanisms

In addition to ASE noise, the integrity of the amplified signal is also impaired by scattering and nonlinear mechanisms along the optical path. The first of these, commonly referred to as multiple-path interference (MPI) [3.14], is illustrated in Fig. 3.13.

MPI is generated by reflection from optical interfaces along the optical path and can originate from both inside and outside the gain element. In practice, even if an EDFA is constructed using only all-fiber components, connectorization between dissimilar fiber types [such as a single-mode fiber (SMF) and an Er-doped fiber] will result in finite facet reflectivity. This effect is much stronger if fusion splicing cannot be used and the fiber optical path has to be closed using fiber connectors or interfaces with free-space optical elements such as gain-flattening filters, isolators, or filters. Even if such (discrete-type) reflection is eliminated, an in-line amplifier is still surrounded by long fiber spans that inherently act as distributed Rayleigh mirrors [3.15]. This is a particularly challenging impairment in bidirectional transmission links that must amplify signals in both directions. In this case, a signal traveling in one direction will be scattered by a distributed Rayleigh mirror, aligning its reflection with the counterpropagating signal flow. In nearly all cases of practical interests (i.e., when

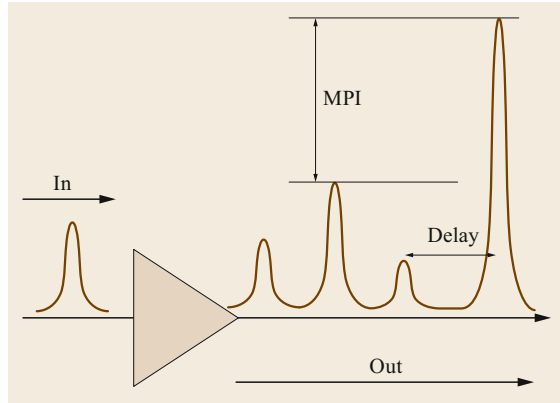


Fig. 3.13 Multiple-path interference (MPI) is generated by multiple reflections off discrete facets of distributed reflectors. The effect is amplified when reflectors are either embedded into or surround an amplifier device

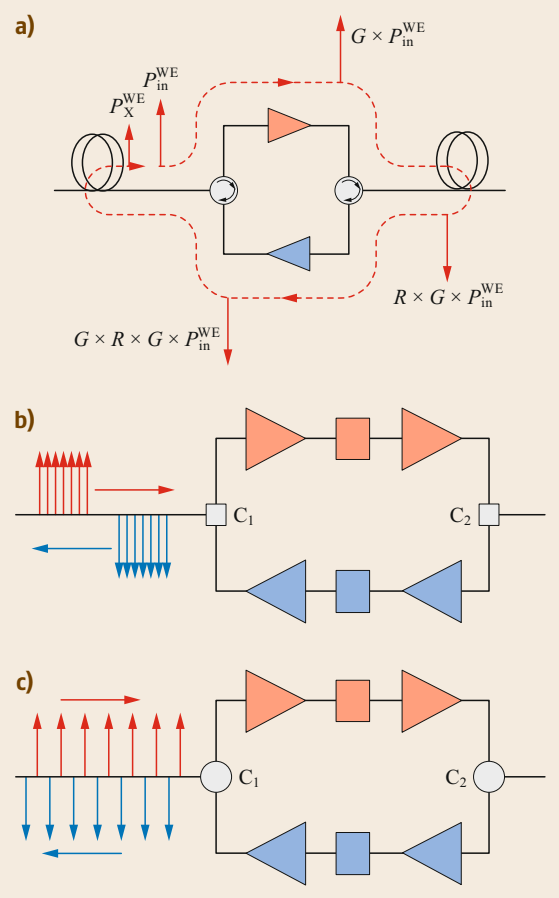


Fig. 3.14 **(a)** MPI induction path in bidirectional amplifier node. **(b)** Banded MPI-mitigation architecture relying on band-splitting (band-combining) elements $C_{1,2}$. **(c)** Interleaved MPI mitigation using interleaved optical filters

the fiber span is longer than 20 km), a Rayleigh mirror will return a reflection that is attenuated by approximately 30 dB with respect to the launched signal and subject it to an amplification pass. Consequently, a conventional in-line amplifier designed to offset a typical ($\approx 80-100$ km) span, will necessarily generate an in-band crosstalk that is only 10 dB lower than the signal power. This is clearly not acceptable, since the effective OSNR, now defined by the crosstalk, would be only 10 dB after traversing a single amplification node.

To address this problem in bidirectional amplified transmission links, counterpropagating channel plans must be frequency (wavelength) nondegenerate, as shown in Fig. 3.14. Banded bidirectional transmission, the preferred implementation for this architecture, has been widely deployed and provides an elegant solution that mitigates MPI induced by either discrete or distributed reflectors. In contrast, an interleaved channel plan can also achieve the same goal, allowing, at least in principle for higher channel density before the onset of a four-photon mixing penalty [3.16, 17]. However, it must be noted that the banded scheme requires only band-splitting (band-combining) elements, while interleaved bidirectional transmission relies on more complex optical filters that have larger insertion loss and lower spectral isolation [3.18].

When used in unidirectional lightwave transmission, an EDFA is a single-pass, traveling amplifier that does not generate internal optical powers comparable to those found within laser cavities with comparable output. An EDFA amplifier is also considered to be a lumped network element, since its effective length is considerably smaller than that of any transmission span of interest. Consequently, it is reasonable to assume that no intra-EDFA nonlinear effects will make a significant contribution to signal impairments, as neither its effective length nor the intradevice power levels appear sufficient to generate large nonlinear phase shifts [3.19].

Unfortunately, this is not always a valid assumption, particularly in the case of L-band EDFA devices [3.20]. Since L-band absorption (emission) is much weaker than in the C-band, the EDF must be correspondingly longer, increasing the effective power–length product that governs the efficiency of the nonlinear interaction in the fiber. This is particularly true in the case when a dense WDM channel plan is combined with L-band EDFA amplification, creating the potential for nonnegligible four-photon mixing (FPM) inside the amplifier. However, intra-EDFA FPM analysis is considerably more complex than that of a passive (transmission) fiber where the signal evolution is monotonically decreasing. This notion is illustrated in Fig. 3.15, showing the direct correspondence between a specific EDFA design that governs the signal evolution and the induced FPM penalty [3.21].

The intraamplifier signal evolution function $G(\omega, z)$, uniquely defined by the local inversion level (i.e., the specific EDFA design) also determines the local power for all the amplified channels. By segmenting the EDFA into a set of quasiuniform sections, it is possible to calculate the FPM efficiency for any combination of channels ω_p, ω_q, and ω_r. By integrating these contributions along the entire EDFA length, the aggregate FPM efficiency can be expressed in the simple form [3.21]

$$\eta = \Omega \left| \int_0^L e^{j\Delta\beta z} \sqrt{\frac{G_p(z)G_q(z)G_r(z)}{G_F(z)}} dz \right|^2 , \qquad (3.34)$$

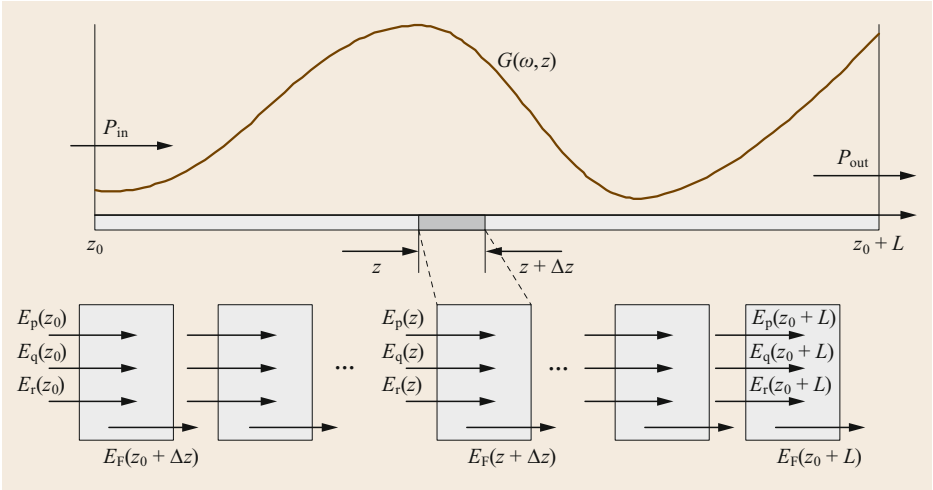


Fig. 3.15 The localized gain function $G(\omega, z)$ within an EDFA uniquely governs the FPM efficiency. By segmenting the EDF into sufficiently small sections, concatenated FPM analysis can be derived and applied to any intraamplifier signal evolution

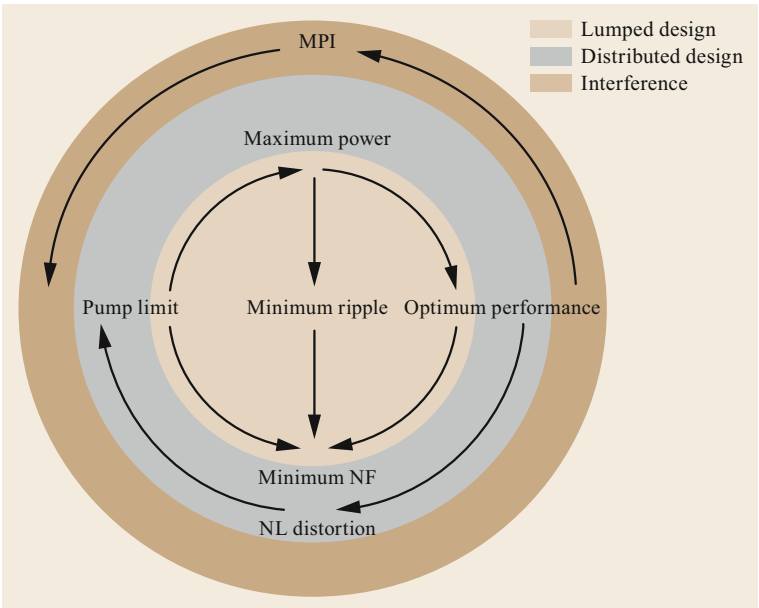


Fig. 3.16 Generalized EDFA design strategy combining conventional (gain/noise/ripple) and advanced network-imposed constraints (signal distortion)

where the scaling factor Ω and mismatch $\Delta\beta$ are given as

$$\Omega = \frac{1024\pi^2 D_\chi^2 P_p(0) P_q(0) P_r(0)}{n^4 \lambda^2 c^2 A_{\text{eff}}^2} G_F(L)$$

$$\Delta\beta \approx \frac{2\pi\lambda^2 \Delta f_{pr} \Delta f_{qr}}{c} \times \left[D(\lambda) + \frac{\lambda^2}{2c} (\Delta f_{pr} + \Delta f_{qr}) \frac{dD(\lambda)}{d\lambda} \right].$$

Consequently, to minimize intra-EDFA FPM distortion, it is necessary to increase the effective area (A_{eff}) and chromatic dispersion (D) of the EDF. Unfortunately, this is not always possible, since the EDF doping level imposes its own constraints on the fiber index profile, leaving FPM minimization to gain-function optimization. Indeed, (3.34) states that FPM can be effectively minimized without affecting the optical properties of the EDF by minimizing the path integral defined by the gain function $G(\omega, z)$.

As a consequence, the design and construction of an optimal EDFA requires much more than a determination of the optimal inversion level. Figure 3.16 illustrates this challenge by outlining an iterative design strategy that aims to produce an amplifier with desired gain and minimal spectral ripple and noise figure while simultaneously suppressing MPI and nonlinear distortions.

3.2 Raman Amplifiers

The phenomenon of scattering of light of frequency ν_0 on molecular vibrations of frequency ν_R, which has become known as *Raman scattering*, was originally discovered in 1928 independently by C.V. *Raman* and K.S. *Krishnan* in India [3.22] and L. *Mandelstam* and G. *Landsberg* in the Soviet Union [3.23]. Raman scattering produces new sidebands with frequencies $\nu_0 - \nu_R$ (the downshifted, or Stokes, wave) and $\nu_0 + \nu_R$ (the upshifted, or anti-Stokes, wave) around the *pump* frequency ν_0. From the quantum-mechanical point of view, the difference between the energy of the pump photon $h\nu_0$ and the Stokes photon $h(\nu_0 - \nu_R)$ is used to excite a phonon with energy $h\nu_R$. To generate the anti-Stokes photon with energy $h(\nu_0 + \nu_R)$, the pump photon energy has to combine with the energy of an existing phonon $h\nu_R$, which means that the anti-Stokes process is $\exp[-h\nu_R/(k_B T)]$ times less probable than the Stokes one, making it possible to carry out precise temperature measurements based on the anti-Stokes/Stokes intensity ratio. With the invention of the laser, it became possible to observe not only spontaneous Raman scattering (where the process starts with no Stokes and anti-Stokes waves at the input) but also stimulated Raman scattering (SRS) [3.24], where the pump wave provides gain for the input Stokes wave. Soon after the invention of the low-loss optical fiber, SRS was observed and characterized in fibers [3.25, 26]. The potential of SRS in fibers for distributed amplification within the transmission span was understood and demonstrated in the 1980s [3.27–29], but it was not until the mid-1990s that the availability of high-power semiconductor laser pumps thrust Raman amplification into the mainstream of optical fiber communications (see [3.30–33] for review). Distributed Raman amplifiers are now widely used both in repeaterless transmission (e.g., festoon-type submarine links) and in repeatered communication systems, while discrete (i.e., lumped) Raman amplifiers are used to overcome the loss of fiber-based components (e.g., dispersion-compensating modules) and to create optical gain at wavelengths outside of the EDFA gain window.

3.2.1 Advantages of Raman Amplifiers

Although Raman amplifiers have lower pump efficiency compared with EDFAs, which requires higher pump powers and/or longer gain fibers, they also offer a number of advantages, which can be summarized as follows:

- Raman gain can be provided to the signal within the transmission fiber span (distributed amplification), which can dramatically improve the signal-to-noise ratio.
- Raman gain can also be implemented in a lumped (discrete) amplifier by employing a relatively short length of a small-effective-area fiber, such as a dispersion-compensating fiber (DCF) or highly nonlinear fiber (HNLF).
- The Raman effect can amplify signals at any wavelength (e.g., outside the EDFA C- and L-bands), as long as the pump wavelength is available one anti-Stokes shift away from it.
- An arbitrarily small gain ripple can be achieved by wavelength-division multiplexing (WDM) of the pumps.
- By employing many fewer optical components in the signal path (essentially only the pump/signal WDM multiplexer), Raman amplifiers can yield lower polarization mode dispersion (PMD) and polarization-dependent gain (PDG) or loss (PDL) than EDFAs, which is important at high symbol rates.

The rest of Sect. 3.2 is organized as follows: Sect. 3.2.2 reviews the basic principles of Raman amplification and its leading noise mechanisms; Sect. 3.2.3 describes practical design issues such as flattening the gain and noise figure, improving the signal-to-noise ratio, avoiding transfer of pump noise, and using numerical modeling; and Sect. 3.2.4 provides a summary and outlook.

3.2.2 Fundamentals

Basic Theory

When a pump beam at frequency ν_0 and a signal beam at frequency $\nu_S = \nu_0 - \nu_R$ propagate in the same optical fiber, their evolution along the fiber is described by the following set of equations for the signal (here $z = 0$ and $z = L$ at the beginning and end of the fiber, respectively)

$$\frac{dP_S^+}{dz} = \left[\overbrace{-\alpha_S}^{\text{Fiber loss}} + \overbrace{\frac{g_R(\nu_R)}{A_{\text{eff}}}}^{\text{Raman gain}} P_P \right] P_S^+ + \underbrace{\alpha_{RBS}}_{\text{Rayleigh scattering}} P_S^- + \underbrace{2n_{sp} h\nu_S \Delta\nu \frac{g_R(\nu_R)}{A_{\text{eff}}} P_P}_{\text{Raman ASE}} ,$$

$$\frac{dP_S^-}{dz} = -\left[-\alpha_S + \frac{g_R(\nu_R)}{A_{\text{eff}}} P_P \right] P_S^- - \alpha_{RBS} P_S^+ - 2n_{sp} h\nu_S \Delta\nu \frac{g_R(\nu_R)}{A_{\text{eff}}} P_P \qquad (3.35)$$

and the pump $P_P(z) = P_P^+(z) + P_P^-(z)$

$$\frac{dP_P^+}{dz} = -\left[\overbrace{\alpha_P}^{\text{Fiber loss}} + \overbrace{\frac{g_R(\nu_R)}{A_{eff}}\frac{\nu_0}{\nu_S}(P_S^+ + P_S^-)}^{\text{Pump depletion}}\right] P_P^+ ,$$

$$\frac{dP_P^-}{dz} = \left[\alpha_P + \frac{g_R(\nu_R)}{A_{eff}}\frac{\nu_0}{\nu_S}(P_S^+ + P_S^-)\right] P_P^- , \quad (3.36)$$

where the subscripts S and P refer to the signal and pump, while the superscripts + and − denote the components of each beam propagating in the $+z$ (forward) and $-z$ (backward) direction, respectively [3.30]. In (3.35) and (3.36), $\alpha_{S(P)}$ represents the fiber attenuation coefficient at the signal (pump) wavelength, $g_R(\nu_R)$ is the Raman gain coefficient for the Stokes shift $\nu_R = \nu_0 - \nu_S$, A_{eff} is the effective area of the fiber, and α_{RBS} is the Rayleigh backscattering coefficient that couples the forward- and backward-propagating waves. A graph of the Raman coefficient $g_R(\nu_R)/A_{eff}$ in several optical fibers is shown in Fig. 3.17, demonstrating broadband gain with a peak near the Stokes shift $\nu_R = 13\,\text{THz}$, which around a signal wavelength of 1550 nm corresponds to a wavelength shift of $\approx 100\,\text{nm}$ (i.e., the pump wavelength has to be $\approx 1450\,\text{nm}$). The relatively small value of the Raman gain coefficient in optical fibers makes Raman amplifiers less power-efficient than EDFAs and requires kilometer-scale fiber lengths at reasonable pump power levels (hundreds of mW). The second term in the bracket of (3.35) describes the Raman gain, which depends on the total power of the forward- and backward-propagating pumps. The similar term in (3.36) corresponds to the pump depletion by the amplified signal. The Raman gain coefficient in a fiber is highly polarization dependent, with the highest gain for a copolarized and the smallest for a cross-polarized signal and pump. However, the fiber PMD along with the large wavelength separation ($\approx 100\,\text{nm}$) between the signal and pump, as well as the use of backward pumping, essentially result in averaging of the Raman gain coefficient over all relative states of polarization of the signal and pump. The use of polarization-multiplexed pumps and/or pump depolarizers makes the Raman gain polarization independent even for a closer-spaced signal and pump and for the forward-pumping configuration. Thus, in (3.35) and (3.36), it can be assumed that the pumps are unpolarized and that the Raman gain coefficient $g_R(\nu_R)$ is polarization independent and equal to approximately one-half of that for a copolarized signal and pump.

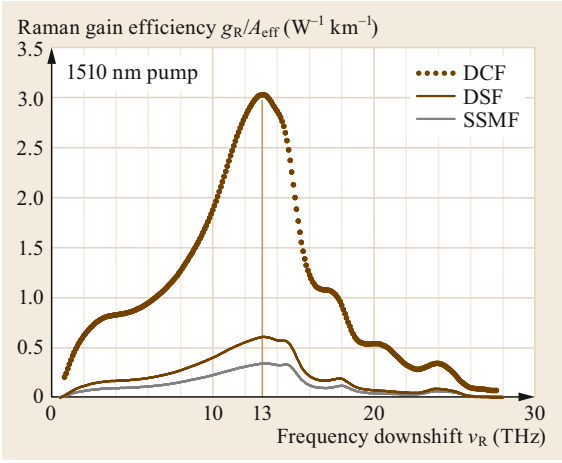


Fig. 3.17 Raman gain spectra in optical fibers. DCF – dispersion-compensating fiber, DSF – dispersion-shifted fiber, SSMF – standard single-mode fiber. (© 2001 IEEE. Adapted, with permission, from [3.31])

The final term in (3.35) represents the ASE power generated by the Raman gain in the optical bandwidth $\Delta\nu$. The factor of 2 describes the total ASE power contained in both polarizations. The spontaneous emission factor $n_{sp} = 1/(1 - e^{-h\nu_R/k_BT})$ has the same meaning as n_{sp} for an EDFA and represents the Raman equivalent of the extra noise due to incomplete inversion, arising from the nonzero phonon population at temperature $T > 0$ ($n_{sp} = 1$ for $T = 0$). At room temperature ($T = 300\,\text{K}$), $n_{sp} \approx 1.14$ for the Stokes shift $\nu_R = 13\,\text{THz}$ corresponding to the peak of the Raman gain in a silica fiber. In (3.36), the Rayleigh scattering of the pump and the pump depletion by the ASE noise are neglected.

Equations (3.35) and (3.36) can first be solved for the evolution of the signal and pump without the Rayleigh backscattering and ASE noise terms, which can be subsequently added to (3.35) as small perturbations of the obtained first-order solutions. For the first-order solution, the evolutions of the signal $P_S^\pm(z)$ and pump $P_P(z)$ are related as

$$P_S^-(z) = 0 ,$$
$$P_S^+(z) = P_S^+(0)g(z) , \quad (3.37)$$

where the signal evolution profile $g(z)$ is given by

$$g(z) = e^{-\alpha_S z}\exp\left[\frac{g_R(\nu_R)}{A_{eff}}\int_0^z P_P(z')\,dz'\right]$$
$$= e^{-\alpha_S z}G_R(z) , \quad (3.38)$$

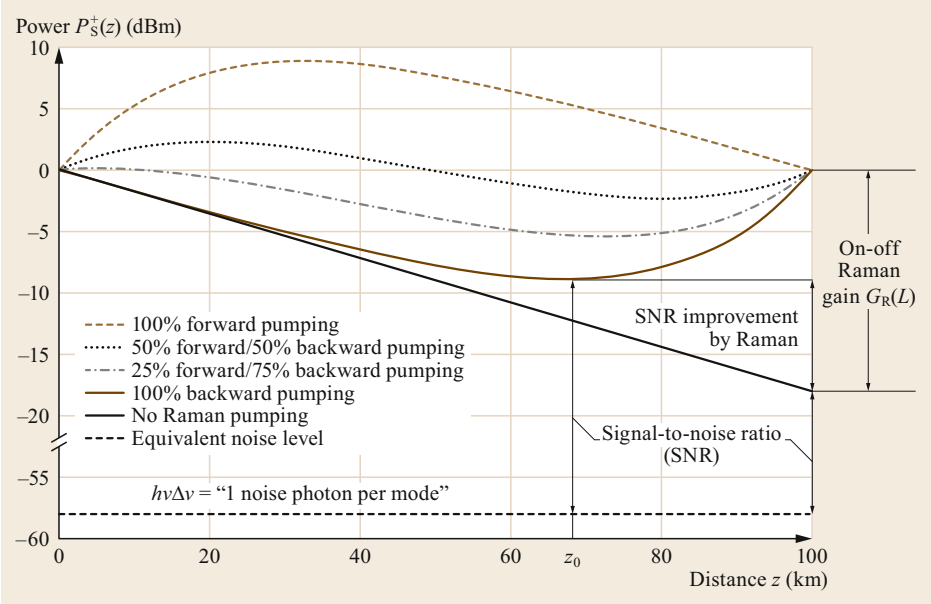


Fig. 3.18 Evolution of signal power over $L = 100$ km span of an SSMF fiber with $\alpha_S = 0.18$ dB/km, $\alpha_P = 0.22$ dB/km, under various Raman pumping configurations. The signal launch power is $P_S^+(0) = 0$ dBm, and the on–off Raman gain $G_R(L)$ of the entire fiber span is set to 18 dB (*transparent* span). The single-photon noise power $h\nu\Delta\nu = -58$ dBm corresponds to $\nu = 193.41$ THz ($\lambda = 1550$ nm) and $\Delta\nu = 12.5$ GHz ($\Delta\lambda = 0.1$ nm). The single-mode SNR at the distributed or lumped amplifier output is related to the ratio between the lowest signal power in the span and the single-photon power level representing the equivalent input noise of the amplifier. The optical signal-to-noise ratio (OSNR) at the amplifier output can be obtained by dividing the single-mode SNR by 2 to account for the ASE in two polarization modes

and $G_R(z)$ is the on–off Raman gain at location z. The on–off gain of the full fiber span is

$$G_R(L) = \exp\left[\frac{g_R(\nu_R)}{A_{\text{eff}}}\int_0^L P_P(z)\,dz\right] = \exp\left[\frac{g_R(\nu_R)}{A_{\text{eff}}}\overline{P_P(z)}L\right], \tag{3.39}$$

where $\overline{P_P(z)} = (1/L)\int_0^L P_P(z)\,dz$ is the path-averaged pump power. To obtain the signal evolution profile in the general case, where the pump depletion term is present in (3.36), one needs to solve (3.35) and (3.36) numerically. However, most frequently, one can neglect the pump depletion term (because the launched pump power is usually much higher than the amplified signal power), which yields

$$P_P(z) = P_{\text{in,P}}[x e^{-\alpha_P z} + (1-x)e^{-\alpha_P(L-z)}] \tag{3.40}$$

and reduces (3.38) to the analytical expression

$$G_R(z) = \exp\left\{\left[x(1-e^{-\alpha_P z}) + (1-x)(e^{\alpha_P z}-1)e^{-\alpha_P L}\right] \times \frac{g_R(\nu_R)P_{\text{in,P}}}{\alpha_P A_{\text{eff}}}\right\}, \tag{3.41}$$

where $P_{\text{in,P}} = P_P^+(0) + P_P^-(L)$ is the total launched pump power, and $x = P_P^+(0)/P_{\text{in,P}}$ and $(1-x) = P_P^-(L)/P_{\text{in,P}}$ are the forward- and backward-propagating fractions of it, respectively. The on–off Raman gain of the entire span in this undepleted pump approximation is given by

$$G_R(L) = \exp\left[\frac{g_R(\nu_R)}{A_{\text{eff}}}P_{\text{in,P}}L_{\text{eff,P}}\right], \tag{3.42}$$

where $L_{\text{eff,P}} = (1-e^{-\alpha_P L})/\alpha_P$ is the effective fiber length for the Raman interaction, and

$$G_R(L)[\text{dB}] = 4.343\frac{g_R(\nu_R)}{A_{\text{eff}}}P_{\text{in,P}}L_{\text{eff,P}}\,. \tag{3.43}$$

Fig. 3.18 shows the evolution of the signal power $P_S^+(z) = P_S^+(0)g(z)$ in a standard single-mode fiber (SSMF) for several values of x at $G_R(L) = 18$ dB.

Raman ASE

As can be seen from Fig. 3.18, the distributed Raman amplification reduces the signal power excursion along the transmission fiber span, and the minimum signal power in the span is increased compared with

that in a fiber with no Raman gain. As a result, one would expect the distributed Raman amplification to yield an improvement in the optical signal-to-noise ratio (OSNR) over lumped amplification. To quantify this improvement, one can compute the Raman ASE noise power from (3.35) by considering it to be a small perturbation of the first-order signal solution of (3.37), (3.38). This yields

$$P^{+}_{\mathrm{ASE}} = 2n_{\mathrm{sp}}h\nu\Delta\nu\left[g(L) - 1 + \int_0^L \frac{g(L)}{g(z)}\alpha_{\mathrm{S}}\,\mathrm{d}z\right], \tag{3.44}$$

which has an analytical solution in the undepleted pump approximation for the backward-pumped case [3.34]. For $\alpha_{\mathrm{P}}L \gg 1$, $G_{\mathrm{R}}(L) \gg 1$, and $\alpha_{\mathrm{S}} = \alpha_{\mathrm{P}}$, this solution for the number of ASE photons per mode reduces to [3.35]

$$N_{\mathrm{ASE}} = \frac{P^{+}_{\mathrm{ASE}}}{2h\nu\Delta\nu} \approx n_{\mathrm{sp}}\frac{G_{\mathrm{R}}(L)}{\ln G_{\mathrm{R}}(L)} \approx n_{\mathrm{sp}}\frac{\mathrm{e}^{\kappa}}{\kappa}\,, \tag{3.45}$$

where $\kappa = g_{\mathrm{R}}(\nu_{\mathrm{R}})P_{\mathrm{in}}/(A_{\mathrm{eff}}\alpha_{\mathrm{P}})$. Equation (3.45) is (up to a constant ≈ 1) equal to the ratio between the signal power at the span output and the lowest signal power in the span, which occurs at a distance $z = z_0 = L - \alpha_{\mathrm{P}}^{-1}\ln(\kappa\alpha_{\mathrm{P}}/\alpha_{\mathrm{S}})$. This leads to a simple physical interpretation of (3.45) as follows: "the generation of Raman ASE noise is equivalent to injecting one noise photon per mode at the point z_0, where the signal power is lowest in the span" [3.35]. (To be more precise, according to (3.45), the generation of Raman ASE noise is equivalent to injecting $en_{\mathrm{sp}} = 3.1$ noise photons per mode at point z_0.) Since the ASE of an ideal (quantum-limited) lumped amplifier at high gain ($N_{\mathrm{ASE}} = G - 1 \approx G$) is equivalent to injecting a single noise photon at its input (i.e., at the $z = L$ end of an unamplified fiber span), then the improvement of the signal-to-noise ratio (SNR) by using the distributed Raman amplifier instead of the ideal lumped one is proportional to the increase in the minimum signal power in the span (Fig. 3.18).

In the forward or bidirectional pumping cases, (3.45) is no longer valid so the ASE power, in general, must be obtained by numerical integration of (3.44). However, even in these cases, qualitative insight into the relative noise performance of various WDM channels or various pumping arrangements can still be obtained by comparing the lowest levels of the signal power in the span (the lower the lowest signal power, the lower the OSNR after the amplifier).

Trade-Off Between Noise Figure and Nonlinearity

From the system performance point of view, when analyzing the unit cell of an optical communication link (i.e., an amplifier and a transmission fiber span), one has to consider the trade-off between its total noise figure and the path-averaged signal power

$$\overline{P_{\mathrm{S}}(z)} = \frac{1}{L}\int_0^L P^{+}_{\mathrm{S}}(z)\,\mathrm{d}z$$

that is responsible for nonlinear impairments [3.36]. Indeed, there are three main possibilities of arranging the gain within the unit cell (i.e., within the amplified span): (a) having a lumped amplifier before the fiber span, (b) having a lumped amplifier after the fiber span, and (c) having the gain distributed within the fiber span. Assuming a transparent unit cell (where the amplifier gain G is equal to the fiber attenuation $\mathrm{e}^{\alpha_{\mathrm{S}}L}$) with input power $P^{+}_{\mathrm{S}}(0)$, one can obtain the noise figures (NFs) and path-averaged powers for these unit cell types, as given in Table 3.1. The lowest NF of the total span in case (a) comes at the expense of the highest path-averaged power (a high-power amplified signal is launched into the fiber). The distributed amplifier's NF and path-averaged power will lie somewhere between

Table 3.1 Performance comparison of types (a), (b), and (c) of amplified fiber spans described in the text. For each type, the total span noise figure, path-averaged signal power, and OSNR factor X from (3.46), (3.47) (as a function of either L or G) are provided, assuming $G = \mathrm{e}^{\alpha_{\mathrm{S}}L}$. For type (c), only the results for ideally distributed gain are given

Amplified span type	NF	$\overline{P_{\mathrm{S}}(z)}$	$X(L)$	$X(G)$
(a) $P^{+}_{\mathrm{S}}(0)$ — G — $1/G$	$3 - 2n_{\mathrm{sp}}(1 - 1/G)$	$P^{+}_{\mathrm{S}}(0)GL^{\mathrm{eff}}_{\mathrm{S}}/L$	$\mathrm{e}^{-\alpha_{\mathrm{S}}L}/(L^{\mathrm{eff}}_{\mathrm{S}})^2$	$1/[G(L^{\mathrm{eff}}_{\mathrm{S}})^2]$
(b) $P^{+}_{\mathrm{S}}(0)$ — $1/G$ — G	$1 + 2n_{\mathrm{sp}}(G - 1)$	$P^{+}_{\mathrm{S}}(0)L^{\mathrm{eff}}_{\mathrm{S}}/L$	$\mathrm{e}^{-\alpha_{\mathrm{S}}L}/(L^{\mathrm{eff}}_{\mathrm{S}})^2$	$1/[G(L^{\mathrm{eff}}_{\mathrm{S}})^2]$
(c) $P^{+}_{\mathrm{S}}(0)$ — $1/G$ — Ideally distributed gain G	$1 + 2n_{\mathrm{sp}}\ln G$	$P^{+}_{\mathrm{S}}(0)$	$1/L^2$	$\alpha_{\mathrm{S}}^2/(\ln G)^2$

the two extreme cases (a) and (b), depending on the signal evolution profile $g(z)$. To compare all of these cases properly, one should consider their OSNR at a fixed path-averaged power, i.e., at a fixed value of the nonlinear phase shift in the span $\Phi_{\mathrm{NL}} = \gamma \overline{P_{\mathrm{S}}(z)} L$, where γ is the nonlinear constant of the fiber. For the general distributed amplifier case (c), this OSNR can be obtained from (3.44) after setting $g(0) = g(L) = 1$

$$\mathrm{OSNR} = \frac{P_{\mathrm{S}}^{+}(0)}{P_{\mathrm{ASE}}} = \frac{\Phi_{\mathrm{NL}}}{2 n_{\mathrm{sp}} h \nu \Delta\nu \gamma \alpha_{\mathrm{S}}} \times X \,, \tag{3.46}$$

where

$$X = \frac{1}{\int_0^L \int_0^L \frac{g(z)}{g(z')} \,\mathrm{d}z \,\mathrm{d}z'} \tag{3.47}$$

is a functional whose value is determined by the signal evolution profile $g(z)$. By setting the functional derivative of (3.47) to zero, it is easy to find that the global maximum of the OSNR for a given value of the nonlinear phase shift corresponds to the signal evolution profile satisfying $g(z) = 1$ for all z, i.e., an *ideal* distributed amplifier (described by row (c) in Table 3.1), in which the fiber loss is compensated by the gain at each point in the span [3.37]. It is worth noting that (3.46), (3.47) also cover cases (a) and (b) when using $g(z) = \mathrm{e}^{\alpha_{\mathrm{S}}(L-z)}$ and $g(z) = \mathrm{e}^{-\alpha_{\mathrm{S}} z}$, respectively. For both of these cases, the OSNR factor X has identical values of $\mathrm{e}^{-\alpha_{\mathrm{S}} L}/(L_{\mathrm{S}}^{\mathrm{eff}})^2$, where $L_{\mathrm{S}}^{\mathrm{eff}} = (1 - \mathrm{e}^{-\alpha_{\mathrm{S}} L})/\alpha_{\mathrm{S}}$. This means that spans with lumped amplification do not result in optimum OSNR performance for a given Φ_{NL}, as illustrated by the fast exponential decay of the OSNR versus their span length L. In contrast, the ideal distributed amplifier exhibits an OSNR with a much slower $1/L^2$ dependence on the span length.

The NF of the distributed amplifier is the total NF of the amplified fiber span, which is subject to the quantum NF limit of $2 - 1/G$. For practical considerations of comparing the distributed and lumped amplified spans, however, it is convenient to separate the passive loss of the fiber and describe the distributed amplifier by a so-called "effective NF" [3.30], which is the noise figure of a fictitious lumped amplifier at the end of the passive fiber span, which provides the same gain and same number of ASE photons as the distributed amplifier. In the general case of a not necessarily transparent span, the total span noise figure is $\mathrm{NF} = (1 + 2N_{\mathrm{ASE}})/g(L)$, whereas the effective Raman NF is

$$\mathrm{NF}_{\mathrm{eff}} = \frac{1 + 2N_{\mathrm{ASE}}}{G_{\mathrm{R}}(L)} \,, \tag{3.48}$$

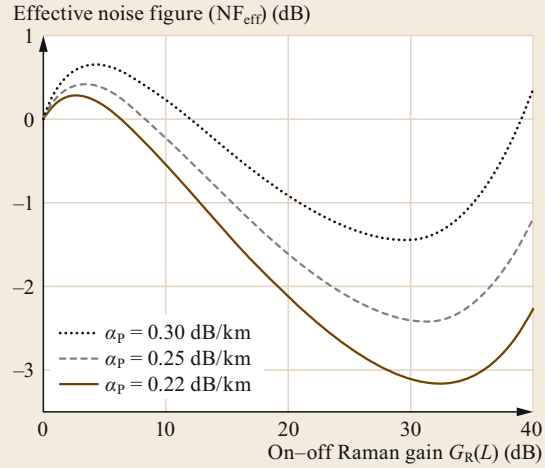


Fig. 3.19 Effective Raman noise figure $\mathrm{NF}_{\mathrm{eff}}$ for a 100-km-long SSMF with $\alpha_{\mathrm{S}} = 0.18\,\mathrm{dB/km}$ and $\alpha_{\mathrm{RBS}} = 5.5 \times 10^{-5}\,1/\mathrm{km}$ (i.e., $R = -31.8\,\mathrm{dB}$) for several values of α_{P}. The increase of $\mathrm{NF}_{\mathrm{eff}}$ at high Raman gains occurs due to Rayleigh backscattered ASE

which is smaller than the total span NF by a factor $g(z)/G_{\mathrm{R}}(L) = \mathrm{e}^{-\alpha_{\mathrm{S}} L}$, i.e., by the transmittance of an unamplified fiber span. The effective NF for a 100-km span of SSMF is shown in Fig. 3.19 as a function of the on–off Raman gain. One can see that $\mathrm{NF}_{\mathrm{eff}}$ can be less than 2 and even less than 1 (i.e., $< 0\,\mathrm{dB}$) at large Raman gains, and the fact that it is below the quantum NF limit simply means that such a fictitious lumped amplifier with the same noise and gain as the distributed Raman amplifier is not physically realizable. The benefit of using $\mathrm{NF}_{\mathrm{eff}}$ is that it permits quick estimation of the OSNR advantage of the distributed amplified span over its lumped counterpart by directly comparing the $\mathrm{NF}_{\mathrm{eff}}$ of the distributed and the NF of the lumped amplifier (in the standard formula $\mathrm{OSNR} = P_{\mathrm{S}}(0)/[h\nu\Delta\nu N(G\,\mathrm{NF} - 1)]$, the lumped NF is simply replaced by $\mathrm{NF}_{\mathrm{eff}}$). From Fig. 3.19, one can also see that this advantage decreases for larger values of α_{P}, because the higher pump attenuation reduces the length of pump penetration into the fiber span and makes the amplifier less distributed.

Impairments due to Rayleigh Scattering

The effect of Rayleigh scattering on the ASE noise and signal can be estimated from (3.35) by treating the Rayleigh terms as small perturbations around the solutions given by (3.44) and (3.37), respectively [3.30]. The first of these effects occurs when the Raman ASE propagating in the $-z$-direction is Rayleigh backscattered in the $+z$-direction, which gives rise to an

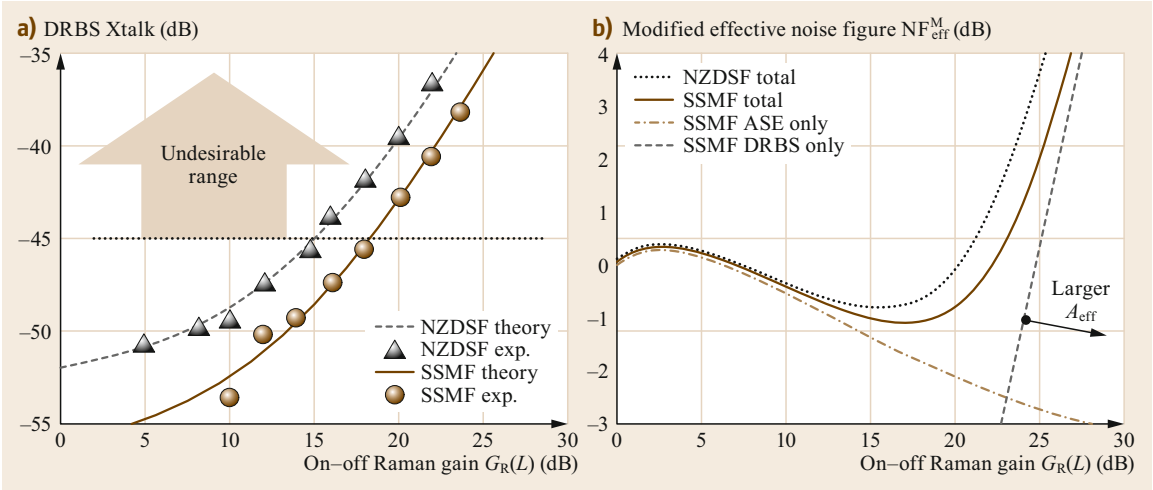


Fig. 3.20 **(a)** DRBS crosstalk, both experimentally measured and theoretically predicted by (3.50), and **(b)** effective Raman noise figure, modified to account for DRBS as in (3.51), for SSMF fiber with $\alpha_S = 0.18\,\mathrm{dB/km}$, $\alpha_P = 0.22\,\mathrm{dB/km}$, $\alpha_{RBS} = 5.5\times10^{-5}\ 1/\mathrm{km}$ ($R = -31.8\,\mathrm{dB}$), $A_{eff} = 80\,\mu\mathrm{m}^2$ and NZDSF fiber with $\alpha_S = 0.19\,\mathrm{dB/km}$, $\alpha_P = 0.23\,\mathrm{dB/km}$, $\alpha_{RBS} = 8.5\times10^{-5}\ 1/\mathrm{km}$ ($R = -30.1\,\mathrm{dB}$), $A_{eff} = 55\,\mu\mathrm{m}^2$, with 100% backward pumping. In **(a)** the length is 80 km for the SSMF and 87 km for the NZDSF, while in **(b)** both fibers have 100 km length. In **(b)**, the NF asymptotes corresponding to ASE-only and DRBS-only contributions are also shown for the SSMF case. To compute the DRBS contribution to the NF, receiver electrical bandwidth $B_{el} = 8\,\mathrm{GHz}$, signal launch power into the fiber $P_S^+(0) = -3\,\mathrm{dBm}$, and signal wavelength $\lambda = 1550\,\mathrm{nm}$ are assumed

additional ASE term referred to as "backscattered ASE" [3.33]

$$P_{\mathrm{ASE}}^{\mathrm{backscatt.}} = 2n_{\mathrm{sp}}h\nu\Delta\nu \times \int_0^L \left(\frac{g^2(L)}{g^2(z)} - \frac{g(L)}{g(z)} + \int_z^L \frac{g(L)g(z')}{g^2(z)}\alpha_S \mathrm{d}z' \right) \times \alpha_{RBS}\mathrm{d}z\,. \tag{3.49}$$

In Fig. 3.19, the backscattered ASE of (3.49) has been added to the original ASE of (3.44) to compute NF_{eff}. The backscattered ASE rises rapidly and monotonically with the on–off Raman gain, and its effect is clearly manifested in Fig. 3.19 by the rise of the NF at high gain values.

The second effect takes place when the forward-propagating signal is backscattered in the $-z$-direction, and this backscattered signal is subsequently backscattered again, now in the $+z$-direction. This double Rayleigh backscattered signal (DRBS) P_{DRBS}^+ represents a crosstalk term that can cause undesirable multiple-path interference with the original signal; this crosstalk is given by [3.38]

$$\mathrm{Xtalk} = \frac{P_{\mathrm{DRBS}}^+(L)}{P_S^+(L)} = \int_0^L \int_z^L \frac{g^2(z')}{g^2(z)}\alpha_{RBS}^2\,\mathrm{d}z\,\mathrm{d}z'\,. \tag{3.50}$$

Since the DRBS makes two passes through the Raman gain region, while the signal only makes one, the DRBS grows much faster with the Raman on–off gain than the signal and, as a result, becomes the main limitation on the Raman gain in the fiber span. Although DRBS cannot be spectrally distinguished from the signal, its crosstalk can be measured either directly in the electrical domain [3.39] or by a time-domain extinction method in the optical domain [3.40], and is shown in Fig. 3.20a for a backward-pumped SSMF and nonzero dispersion-shifted fiber (NZDSF). The higher DRBS crosstalk in the NZDSF is due to this fiber's higher Rayleigh backscattering coefficient α_{RBS}. This coefficient is given by $\alpha_{RBS} = 2R\alpha_S$, where R is the fraction of the light returned (backscattered) by an infinite fiber. R is proportional to $\mathrm{NA}^2 \propto \lambda^2/A_{eff}$ (smaller effective area translates into a larger acceptance angle of the mode, leading to a larger fraction of the Rayleigh scattering being recaptured into the fiber mode). As the DRBS crosstalk keeps increasing with the Raman gain, at some point it enters the range labeled *undesirable* in Fig. 3.20a, where it becomes comparable to the inverse OSNR, i.e., starts causing noticeable performance degradations that limit the maximum usable Raman gain. For the purpose of finding the optimum Raman gain, it is helpful to include the effect of the DRBS crosstalk into a modified effective NF of the Raman amplifier

as [3.33, 41, 42]

$$\mathrm{NF}_{\mathrm{eff}}^{\mathrm{M}} = \mathrm{NF}_{\mathrm{eff}} + \frac{5}{9}\frac{P_{\mathrm{S}}^{+}(L)}{h\nu B_{\mathrm{el}}}\frac{\mathrm{XTalk}}{G_{\mathrm{R}}(L)}\,, \tag{3.51}$$

where the factor 5/9 is the DRBS fraction copolarized with the signal [3.43] and B_{el} is the electrical bandwidth of the receiver ([3.42] provides a more accurate bandwidth expression, suitable for various modulation formats).

The modified noise figure $\mathrm{NF}_{\mathrm{eff}}^{\mathrm{M}}$ is plotted in Fig. 3.20b for backward-pumped NZDSF and SSMF fibers. The presence of DRBS leads to a rapid rise of the NF at high Raman gains that are well below those leading to turnaround in Fig. 3.19. The optimum operation is achieved at the on–off Raman gain at which the modified NF has the lowest value. It can be seen from Fig. 3.20b that the larger effective area of the SSMF pushes the NF of the SSMF toward a later turnaround and lower values than the NF of the NZDSF.

3.2.3 Design Considerations

In the design of practical Raman amplifiers, one needs to consider the following main issues:

- Choosing between a lumped and distributed Raman amplifier, as well as between a Raman-only and hybrid Raman–EDFA amplified span; this choice also sets the target Raman gain.
- Designing a flat gain spectrum for WDM signals by using WDM pumping.
- Flattening and optimizing the NF of the amplifier, including the noise–nonlinearity trade-off.
- Minimizing the potential impact of pump relative intensity noise (RIN).

The rest of this section elaborates on these design steps. The semianalytical theory described in Sect. 3.2.2 provides the necessary insights into these procedures. For accurate quantitative predictions, however, a more complete Raman amplifier model is described at the end of the section, which is typically solved numerically using either home-made or commercial modeling software.

Choosing the Type of Raman Amplifier

The type of Raman amplifier to use is primarily determined by its intended application. Let us first distinguish between distributed and lumped (also known as discrete) Raman amplifiers. Although both types of amplifier use substantial (kilometer-scale) lengths of optical fiber, distributed amplification takes place within the transmission fiber span, whereas discrete amplification uses a piece of fiber specifically designated as the gain medium, so that all the gain occurs within a discrete amplifier box. The primary applications of distributed Raman amplification include SNR improvement by distributed gain, extending the span length between lumped amplifiers, and building communication links in wavelength ranges lying outside conventional EDFA bands. Since a discrete amplifier does not offer the NF advantages of a distributed amplifier, to justify its use instead of a more pump-power-efficient EDFA the application would typically either involve operation outside of the EDFA band or require gain control (e.g., ultraflat gain) beyond what is achievable with an EDFA, or employ a gain medium that has some additional functionality (e.g., a dispersion-compensating fiber). On the other hand, the discrete amplifier has greater design flexibility than the distributed one: it can use specialty fiber with high Raman gain coefficient, split the gain among multiple stages, employ optical isolators to reduce DRBS and backscattered ASE, etc. In what follows, we focus on the design issues that are common to both distributed and discrete amplifiers. More details specific to the discrete amplifier design can be found in [3.44].

The second distinction worth mentioning is whether the distributed Raman amplifiers are designed to compensate the loss of the fiber span on their own (all-Raman communication links) or as low-noise preamplifiers working in combination with EDFA boosters (hybrid Raman–EDFA links). The high gain required for all-Raman links comes at the price of having to deal with DRBS crosstalk, pump depletion, higher path-averaged signal power, etc. All-Raman systems might require higher pump power, but they use fewer optical components in the signal path than hybrid Raman–EDFA systems, which reduces PMD, PDL, PDG, etc. On the other hand, by splitting the gain between the distributed Raman and lumped EDFA stages in hybrid Raman–EDFA systems, one can realize almost the entire NF advantage of distributed Raman amplification without suffering from the aforementioned impairments (DRBS, pump depletion, etc.) associated with high Raman gain. This is because, in such a two-stage amplifier design, the total noise figure $\mathrm{NF}_{\mathrm{total}} = \mathrm{NF}_1 + (\mathrm{NF}_2 - 1)/G_1$ is mostly determined by the NF of the distributed Raman preamplifier (the first stage with $\mathrm{NF}_1 = \mathrm{NF}_{\mathrm{eff}}$ and gain $G_1 = G_{\mathrm{R}}(L)$). The hybrid Raman–EDFA combination is also a simple way to accommodate a longer-than-typical fiber span in an otherwise EDFA-only system. For optimum hybrid Raman–EDFA performance, however, it is important to have an EDFA power amplifier (booster) stage that is designed to handle the higher input signal powers coming from the Raman-amplified span compared with an unamplified span. For example, simple addition of distributed Raman preamplifiers to standard in-line ED-

FAs in an attempt to improve the OSNR in an existing EDFA-only link is of very limited benefit: the reduction of gain in the in-line EDFA, needed to accommodate the higher input signal power, comes with a large increase in the EDFA NF (because the gain adjustment is typically done by a variable optical attenuator inside the EDFA, if the gain flatness is to be preserved), which could considerably reduce the distributed Raman NF advantage while not offering any cost savings. On the other hand, replacing a high-gain in-line EDFA with a low-gain EDFA booster stage permits its operation with the higher input signal power coming from the Raman-amplified span without major degradation of the EDFA NF, while at the same time eliminating the cost of an EDFA preamplifier stage.

Once the application scenario for the Raman amplifier is known, the considerations above lead to the choice between a lumped or distributed, as well as all-Raman or hybrid Raman–EDFA, amplifier. For an all-Raman system, the required Raman gain is determined by the span loss (plus, possibly, the loss of any lumped end-of-span components, e.g., pump-signal WDM multiplexer and isolator). For a hybrid Raman–EDFA system, the Raman gain can be varied over a wide range. If the EDFA stage has been designed to accommodate the higher power of the Raman-amplified signal, then the optimum value of the Raman gain can be found from the trade-off among $\mathrm{NF}_{\mathrm{eff}}$, DRBS, and the nonlinearity of the Raman amplifier, as discussed below. Otherwise, i.e., if the EDFA stage exhibits strong NF dependence on the input signal power, a trade-off among $\mathrm{NF}_{\mathrm{total}}$, DRBS, and nonlinearity is required.

Shaping the Gain Spectrum by WDM Pumping

One of the advantages of Raman amplification is its flexibility in shaping the gain spectrum by the use of multiple pump wavelengths (WDM pumping), which can yield wideband amplification with very low gain ripple [3.31, 45]. Indeed, in the presence of multiple pump wavelengths, (3.35) takes the form

$$\frac{\mathrm{d}P_{\mathrm{S}}^{\pm}}{\mathrm{d}z} = \pm\left[-\alpha_{\mathrm{S}} + \sum_i \frac{g_{\mathrm{R}}(\nu_{\mathrm{P}}^i - \nu_{\mathrm{S}})}{A_{\mathrm{eff}}} P_{\mathrm{P}}^i\right] P_{\mathrm{S}}^{\pm} , \quad (3.52)$$

where ν_{P}^i, $i = 1, \ldots, N$ are the center frequencies of the individual pump lasers. The on–off Raman gain is given by

$$G_{\mathrm{R}}(L) = \mathrm{e}^{\sum_i \frac{g_{\mathrm{R}}(\nu_{\mathrm{P}}^i - \nu_{\mathrm{S}})}{A_{\mathrm{eff}}} \int_0^L P_{\mathrm{P}}^i(z)\mathrm{d}z} , \quad (3.53)$$

which can be conveniently rewritten on a decibel scale as

$$G_{\mathrm{R}}(L)[\mathrm{dB}] = 4.343 \sum_i \frac{g_{\mathrm{R}}(\nu_{\mathrm{P}}^i - \nu_{\mathrm{S}})}{A_{\mathrm{eff}}} \overline{P_{\mathrm{P}}^i(z)} L , \quad (3.54)$$

where $\overline{P_{\mathrm{P}}^i(z)}$ is the path-averaged power of the i-th pump. Equation (3.54) enables synthesis of the desired gain spectrum from the gain spectra $g_{\mathrm{R}}(\nu_{\mathrm{P}}^i - \nu_{\mathrm{S}})$ contributed by the individual pump wavelengths. A typical design procedure starts with the definition of constraints (fiber type, number of pump lasers, and maximum power of each) and desired specifications (gain bandwidth and flatness). Then, design parameters such as the wavelengths and powers of the individual pumps, are varied to achieve the target specifications as close as possible. For rough initial estimates, (3.54) can be used, but for the final amplifier design, the full amplifier model including pump depletion and pump–pump interactions (described below) should be employed. Figure 3.21, adapted from [3.31], presents the results of such synthesis for a >80-nm-wide gain spectrum in an SSMF fiber from five pump wavelengths, yielding gain ripple of 0.3 dB.

Since the Raman gain spectrum is easily shifted by shifting the pump wavelength, the conventional Fabry–Pérot laser-diode pumps used for EDFAs are not well suited for Raman pumping. To stabilize the central wavelengths of the pump lasers, fiber Bragg gratings (FBGs) are frequently used, and such a combination is referred to as FBG-stabilized laser diodes. Various pump wavelengths can be combined (multiplexed) using either cascaded Mach–Zehnder interferometer couplers on a planar lightwave circuit platform, or thin-film-based couplers, or fused fiber couplers, with the latter type being most popular owing to their high power-handling capabilities. For a given pump wavelength, a polarization beam combiner is often used to multiplex two orthogonally polarized lasers, which simultaneously doubles and depolarizes the pump power at this wavelength. If a single-diode pump power is sufficient, then an alternative depolarizing solution can be used, such as coupling the laser equally into the two axes of a polarization-maintaining fiber of sufficient length to introduce a delay between the two polarizations that is greater than one coherence length of the laser. Commercial Raman amplifier modules are usually available in the form of Raman pump units (RPUs), an example of which is shown in Fig. 3.22, and are capable of operating in both backward and forward pumping configurations.

Flattening and Improving the NF

Although WDM pumping can arbitrarily flatten the Raman gain spectrum, the spectrum of the Raman NF is not flat, with short-wavelength signals accumulating larger amounts of noise than longer-wavelength signals. This tilt in the NF spectrum originates from several factors:

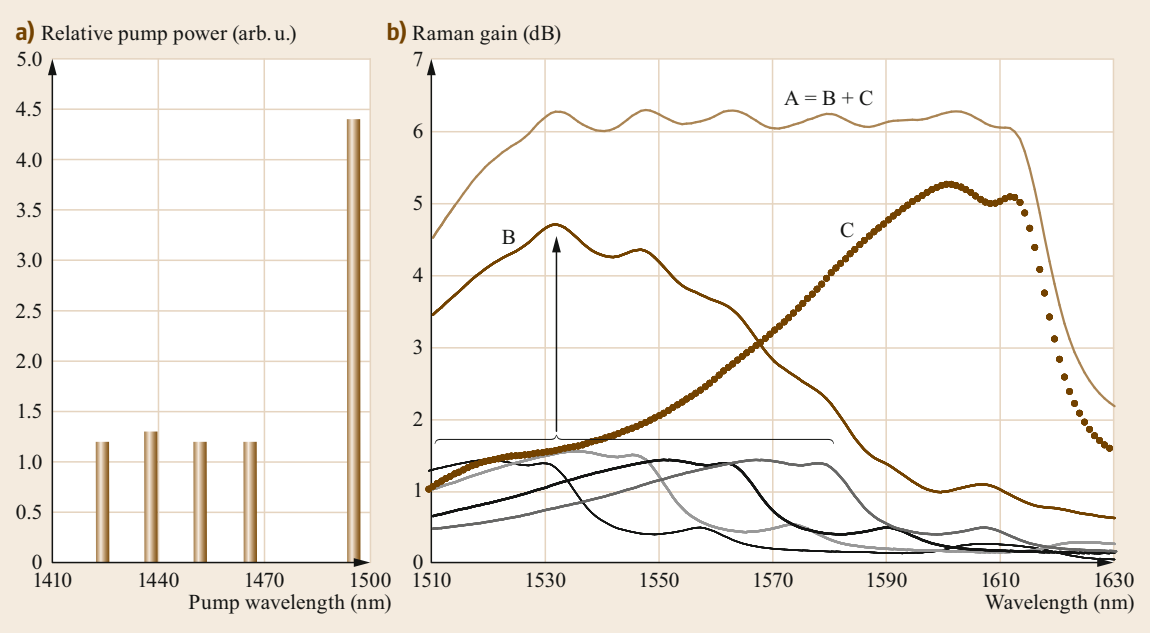


Fig. 3.21a,b Synthesis of broadband and flat Raman gain spectra by WDM pumping in SSMF fiber. **(a)** Relative powers of five WDM pumps. **(b)** Gain spectra contributed by the first four pumps (*lower traces, adding up to trace B*) and the last pump (C), adding up to the flattened net gain spectrum A. (© 2001 IEEE. Adapted, with permission, from [3.31])

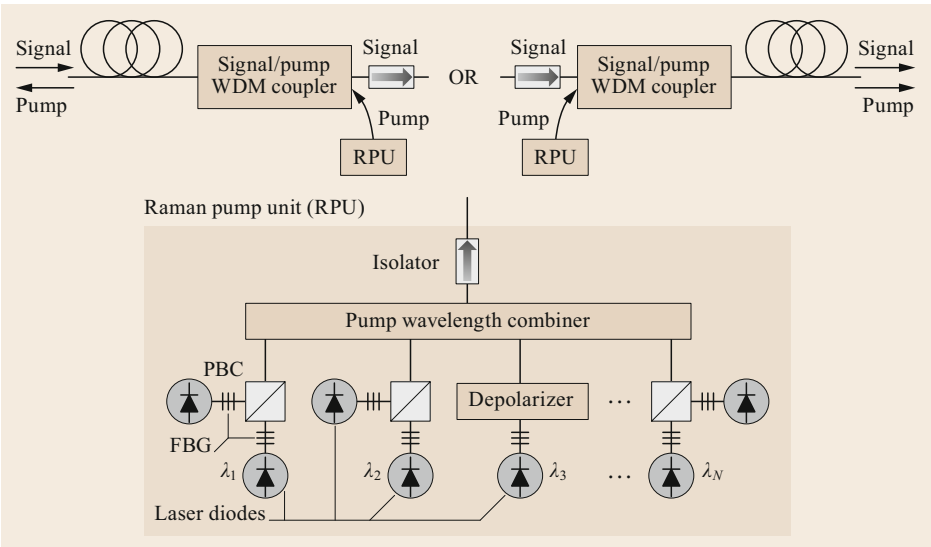


Fig. 3.22 Typical Raman pump unit with *N* pump wavelengths, which can be used in either counterpropagating (backward pumping) or copropagating (forward pumping) signal–pump configuration (PBC – polarization beam combiner; FBG – fiber Bragg grating)

(a) Shorter-wavelength signals are pumped primarily by shorter-wavelength pumps that have higher attenuation coefficients, i.e., do not penetrate as deep into the fiber span as longer-wavelength pumps.
(b) Shorter-wavelength pumps also lose some power to the longer-wavelength ones through pump–pump Raman interactions.
(c) Similar signal–signal Raman interaction leads to additional power loss of short-wavelength signals (SRS tilt).

Thus, the gain for shorter-wavelength signals is less distributed, permitting the signal power to drop to lower levels, which in turn leads to worse NF performance than that of longer-wavelength signals.

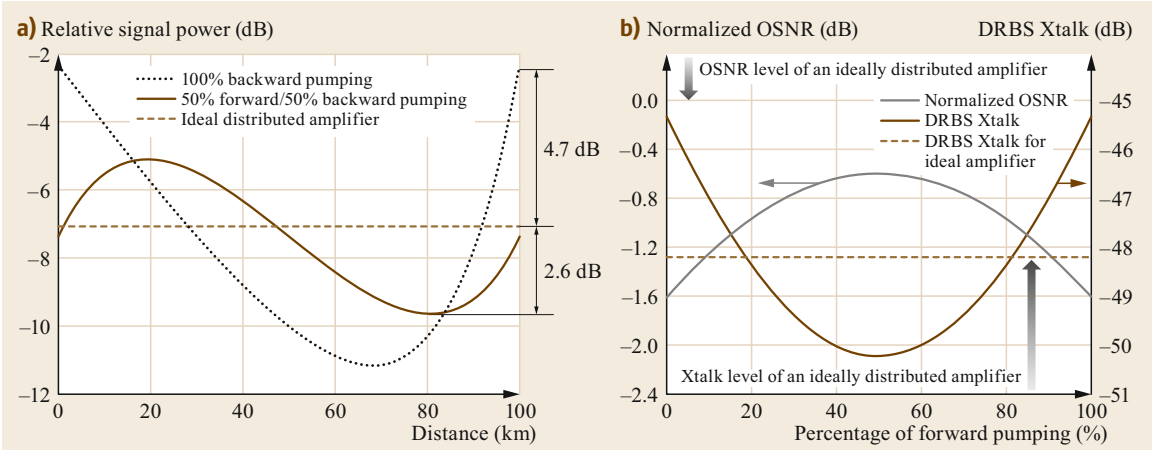


Fig. 3.23 **(a)** Evolution profiles of Fig. 3.18, normalized to yield the same path-averaged power as an unamplified fiber span. **(b)** OSNR, normalized by the OSNR of an ideally distributed amplifier (*left scale*), and DRBS crosstalk (*right scale*) of different pumping configurations for an SSMF span with parameters of Fig. 3.18

To counter this effect, one needs to make the Raman gain more distributed over the fiber span for these short-wavelength signal channels. In this section, three techniques for making the Raman gain more uniformly distributed are considered: (a) bidirectional (i.e., combination of forward and backward) pumping, (b) effective-area management in the span, and (c) second- and higher-order Raman pumping.

Although complementing the backward pumping of all signal channels with forward pumping of shorter-wavelength ones appears to be quite a simple solution, it needs to be accompanied by a significant reduction of the signal launch power into the span. Indeed the signal evolution profiles of Fig. 3.18, corresponding to different forward/backward pumping configurations, have very different path-averaged powers. The proper way to compare their performance is to adjust their signal launch power in such a way that all of these configurations yield the same path-averaged power (Fig. 3.23a). Under these conditions, the best performance is achieved for the 50/50 backward/forward combination, whose evolution profile over a 100-km-long SSMF deviates by no more than 2.6 dB from the ideally distributed amplifier profile, whereas the backward-pumped signal deviates by 4.7 dB. The resulting OSNR of the 50/50 combination is within 0.6 dB of that of an ideally distributed amplifier (Fig. 3.23b). It has also been shown that the 50/50 case is optimum for reduction of the DRBS crosstalk [3.38]. This can be understood by noting that (3.50) implies that, to minimize the DRBS, one needs to avoid generating large amounts of Raman gain in any particular section of the fiber. Indeed, 50/50 pumping splits the Raman gain between the two end regions of the span, with the loss in the middle effectively serving as an isolator to prevent circulation of DRBS between them. The DRBS performance of the 50/50 configuration even outshines that of the ideally distributed amplifier, whose distributed gain yields better DRBS than either backward-only or forward-only pumping, but is still about 2 dB worse than the DRBS of the 50/50 case.

However, since forward pumping has a number of drawbacks as outlined in the next section, alternative approaches have been found and explored. One particularly attractive method uses spans consisting of multiple fibers with judiciously chosen effective areas. The pump initially enters a fiber with large effective area (usually a silica-core fiber with large positive dispersion) and, after experiencing attenuation, subsequently enters a smaller-effective-area fiber (usually with large negative dispersion), so that the pump intensity again reaches high enough values to provide gain deeply into the fiber span (Fig. 3.24a). Spans consisting of $+D$ and $-D$ fibers are frequently used to perform accurate and temperature-independent dispersion and dispersion-slope compensation right within the fiber span, and such dispersion-managed fibers (DMFs) are ideal candidates for effective-area-managed Raman amplification [3.46, 47]. For maximum performance, the large-effective-area $+D$ fiber sections need to be positioned near the ends of the span, with small-effective-area $-D$ fiber in the middle ($+D/-D/+D$, or *symmetric DMF*). A DMF can employ a $-D$ fiber with one of several types of $x:1$ dispersion compensation ratios (x km of $+D$ fiber compensated by 1 km of $-D$ fiber). Figure 3.24b shows that, even with backward-only pumping, the NFs of both 1 : 1 and 2 : 1 DMFs outperform the NF of NZDSF by more than 2.5 dB [3.33, 46],

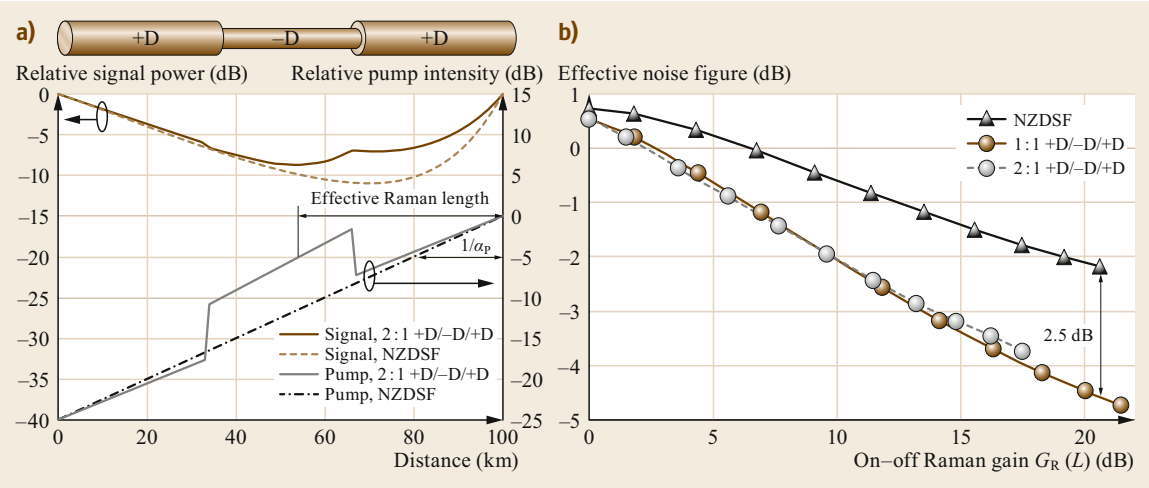


Fig. 3.24 **(a)** Comparison of signal and pump evolutions in 100-km-long NZDSF ($A_{eff} = 72\,\mu m^2$, $\alpha_S = 0.21\,dB/km$, $\alpha_P = 0.25\,dB/km$) and 2 : 1 DMF; $+D$ fiber of the DMF has $A_{eff} = 110\,\mu m^2$, $\alpha_S = 0.19\,dB/km$, $\alpha_P = 0.22\,dB/km$, and $-D$ fiber has $A_{eff} = 27.5\,\mu m^2$, $\alpha_S = 0.25\,dB/km$, $\alpha_P = 0.29\,dB/km$. **(b)** Comparison of experimentally measured effective Raman noise figures NF_{eff} between NZDSF ($L = 100\,km$) and two DMFs, with dispersion ratios 1 : 1 ($25\,km + 50\,km + 25\,km = 100\,km$) and 2 : 1 ($30\,km + 27\,km + 25\,km = 82\,km$). (After [3.33])

which has enabled all-backward-pumped Raman system demonstrations [3.48, 49]. It has been shown [3.50] that the normalized OSNR of the backward-pumped symmetric DMF is within 0.7 dB of the ideal distributed amplifier, and with optimum 30% forward pumping, it can be within 0.3 dB from the ideal. Equations (3.44), (3.49), and (3.50) are written in a form suitable for multiple-fiber spans, if one assumes α_S and α_{RBS} to be functions of distance (and in (3.50) one α_{RBS} should be a function of z, and the other of $-z'$) [3.33, 51–53].

Another approach that similarly achieves deeper penetration of the Raman gain into the fiber span is to use second- or higher-order Raman pumping [3.54, 56]. The second-order pump is placed one anti-Stokes shift higher in frequency from the conventional first-order pump and serves as a pump wave that provides gain for the first-order Raman pump. Third- and even higher-order pumps can be placed at additional anti-Stokes shifts away. As the first-order pump enters the fiber span, it starts being amplified by the second-order pump and reaches its peak power some distance away from the span's end. This pushes the Raman gain deeper into the span, i.e., makes it more distributed and closer to ideal. Figure 3.25, adopted from [3.54], shows that the combination of second-order and bidirectional pumping can yield very small (±0.4 dB) signal power excursions over the length of an 80-km SSMF span. Figure 3.25b shows a typical NF_{eff} tilt for first-order backward pumping, which is flattened by combining it with forward pumping, and further flattened and reduced by second-order pumping.

Challenges of Forward Pumping and Effects of Pump RIN

Forward pumping brings about a new challenge, as the copropagation of the pump and signal makes it necessary to consider time-domain effects such as the transfer of relative intensity noise (RIN) from the pump to signal and interchannel crosstalk mediated by cross-gain modulation (pump depletion). Indeed, pump RIN causes temporal variations of the Raman gain per unit length, which can accumulate exponentially and lead to signal RIN greatly exceeding that of the pump. Fortunately, this degradation is limited by the dispersive walk-off between the signal and pump, which averages the pump fluctuations and effectively acts as a low-pass filter of the pump noise. Figure 3.26 shows the RIN transfer function from the pump to signal in NZDSF and DCF fibers [3.55], demonstrating the much narrower bandwidth of the transferred noise in the more-dispersive DCF. In the counterpumped (backward-pumped) configuration, the averaging takes place over the time of flight over $2L_{eff}$, regardless of the fiber dispersion, which reduces the transferred RIN bandwidth to less than 100 kHz. To avoid Q-factor degradations due to the signal RIN, the total RIN from pump lasers in a chain of amplified spans should be under −120 dB/Hz for forward pumping and under −70 dB/Hz for backward pumping [3.55], which translates into individual pump laser RIN requirements of roughly less than −140 dB/Hz and less than −90 dB/Hz, respectively.

While common FBG-stabilized laser diodes with RIN of about −120 dB/Hz are quite adequate for

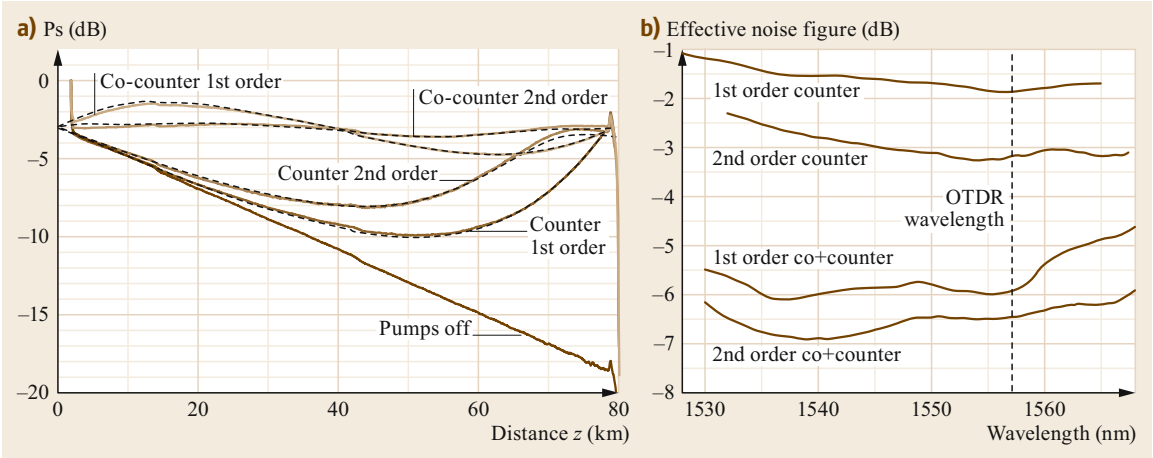


Fig. 3.25 **(a)** Signal evolution profiles measured by optical time-domain reflectometer (*solid lines*) and simulated (*dashed lines*). **(b)** Effective NFs measured for four Raman pumping configurations. (© 2002 IEEE. Adapted, with permission, from [3.54])

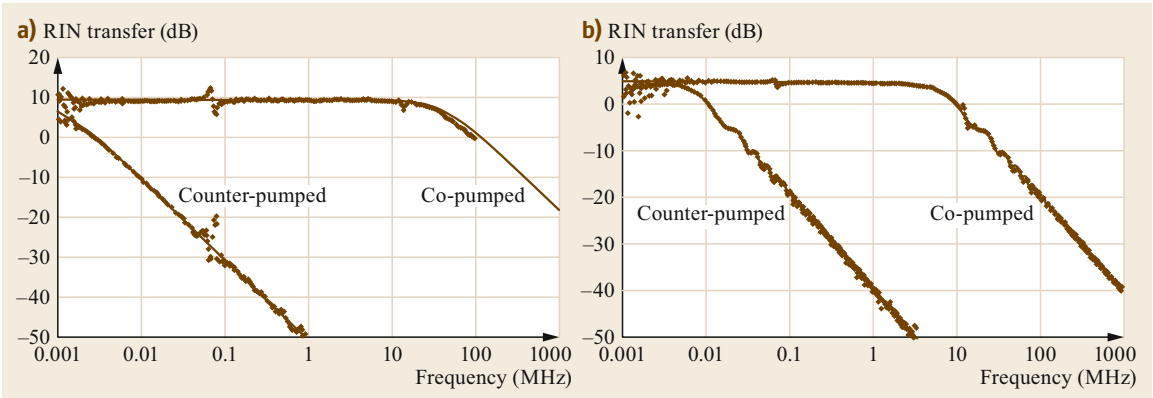


Fig. 3.26a,b RIN transfer characteristic for **(a)** 81 km of TrueWave-RS™ NZDSF ($\alpha_P = 0.25\,dB/km$ at 1455 nm, $D = 2.3\,(ps/nm)/km$ at 1500 nm, and on–off Raman gain = 12.6 dB) and **(b)** 6.8 km long dispersion-compensating fiber (DCF, $\alpha_P = 1.52\,dB/km$ at 1455 nm, $D = -97.6\,(ps/nm)/km$ at 1500 nm, and on–off Raman gain = 7.4 dB). *Dots* indicate measurements; *solid lines* are theoretical calculations. Pump wavelength is 1455 nm, signal wavelength is 1550 nm. (© 2001 IEEE. Adapted, with permission, from [3.55])

backward pumping, their RIN is too high for forward pumping. Fabry–Pérot lasers have lower high-frequency noise, but still unacceptably high low-frequency RIN (below 10–100 MHz). DFB lasers offer the best noise performance ($< -160\,dB/Hz$), but their narrow linewidths make them highly susceptible to stimulated Brillouin scattering (SBS). For forward-pumping applications, special versions of a DFB laser with a modified on-chip grating have been developed to allow multiple longitudinal modes to lase, resulting in significant suppression of SBS while retaining an RIN level close to $-160\,dB/Hz$ [3.57, 58]. Another type of low-RIN pump source for forward pumping is based on broadband ASE noise.

Further complications may occur if the zero-dispersion wavelength of the fiber λ_0 is located either within the pump band or between the pump and signal bands [3.59–62]. In the former case, significant pump–pump four-wave mixing needs to be considered. In the latter case, the pump and signal symmetrically located on the opposite sides of λ_0 copropagate with the same group velocity, and averaging of pump RIN does not take place. This leads to very strong Raman gain modulation and cross-phase modulation of the signal by the pump RIN, making forward pumping highly undesirable. Even in backward pumping, the backscattered ASE might experience gain enhanced by large pump RIN and become noticeable at RIN val-

ues of −115 dB/Hz and higher [3.61]. On the other hand, the same gain and cross-phase modulation by the pump RIN can have the beneficial effect of reducing the impact of DRBS crosstalk in the backward-pumped configuration: the modulation by the pump RIN dramatically broadens the DRBS spectrum and reduces its overlap with the signal [3.62].

Another type of impairment that is particularly important in forward-pumping schemes is the interchannel crosstalk mediated by the pump depletion. This happens when a fluctuation increasing intensity of one WDM channel depletes the pump and reduces the gain for another channel, causing crosstalk. This crosstalk is greatly reduced by the signal–pump walk-off and by increasing the number of WDM channels, which effectively perform temporal and ensemble averaging of the signal intensity fluctuations.

Modeling

Accurate modeling of wideband Raman amplifiers employing multiple WDM signals and WDM pumps requires taking into account not only the Raman gain of the signal, but also pump depletion and pump–pump and signal–signal Raman interactions. The evolution of the pumps, signals, and noise in such a Raman amplifier is described by the following equations [3.31]:

$$\pm\frac{\mathrm{d}P_\nu^\pm}{\mathrm{d}z} = \overbrace{-\alpha_\nu P_\nu^\pm + \alpha_{\mathrm{RBS}_\nu} P_\nu^\mp}^{\text{Fiber loss and Rayleigh backscattering}} + \overbrace{P_\nu^\pm \sum_{\mu>\nu} \frac{g_{\mu\nu}}{A_{\mu\nu}}(P_\mu^+ + P_\mu^-)}^{\text{Raman gain from shorter wavelengths}} + \overbrace{2h\nu\Delta\nu \sum_{\mu>\nu} \frac{g_{\mu\nu}}{A_{\mu\nu}}(P_\mu^+ + P_\mu^-)n_{\mu\nu}}^{\text{ASE noise}} - \overbrace{P_\nu^\pm \sum_{\mu<\nu} \frac{\nu}{\mu}\frac{g_{\nu\mu}}{A_{\nu\mu}}(P_\mu^+ + P_\mu^-)}^{\text{Raman loss to longer wavelengths}} - \overbrace{4h\nu P_\nu^\pm \sum_{\mu<\nu} \frac{g_{\nu\mu}}{A_{\nu\mu}} n_{\nu\mu}\Delta\mu}^{\text{Raman loss due to noise emission}} , \tag{3.55}$$

where the superscripts + and − denote forward- and backward-propagating (propagating in $+z$- and $-z$-direction) waves, respectively, $n_{\mu\nu} = \{1 - \exp[-h(\mu - \nu)/(k_\mathrm{B}T)]\}^{-1}$ is the spontaneous emission factor n_sp dependent on the Stokes shift $(\mu - \nu)$, and the subscripts μ and ν refer to two different frequencies of light. Equation (3.55) treats the signals and pumps in the same way: the wave at frequency ν is amplified by waves at frequencies $\mu > \nu$ and serves as a pump for waves with frequencies $\mu < \nu$. One should note that, in the design of wideband (> 100 nm wide) Raman amplifiers, one needs to consider that the Raman gain coefficient $g_{\mu\nu}$ not only depends on the Stokes shift $(\mu - \nu)$ but also varies slowly with the pump frequency μ. The scaling of $g_{\mu\nu}$ and the effective area $A_{\mu\nu}$ with μ is described in [3.63], but from a practical perspective, it is easier to measure the gain spectrum $g_{\mu\nu}/A_{\mu\nu}$ in a given fiber for several pump wavelengths across the desired wavelength range and then interpolate the results.

The system of coupled ordinary differential equations (3.55) describes a boundary-value problem with initial values known at $z = 0$ for forward- and at $z = L$ for backward-propagating waves. It can be solved numerically, e.g., by implementing algorithms from [3.64–66] or by using commercial software [3.67, 68]. The main challenges that require the numerical solution of (3.55) are the effects of pump depletion, pump–pump interactions, and signal–signal interactions. On the other hand, Rayleigh scattering and ASE noise can usually still be accounted for semianalytically as small perturbations around the evolution profile $g(z)$ of the signal at frequency ν. Such a semianalytical approach is helpful when using some numerical packages that do not otherwise compute the Rayleigh scattering and ASE noise terms. In that case, the DRBS crosstalk is still given by (3.50), whereas the ASE power needs to be rewritten to account for the varying Raman gain coefficients of the multiple pumps as follows

$$P_{\mathrm{ASE}_\nu}^+ = 2h\nu\Delta\nu \times \left\{ \int_0^L \frac{g(L)}{g(z)} \sum_{\mu>\nu} n_{\mu\nu}\frac{g_{\mu\nu}}{A_{\mu\nu}}[P_\mu^+(z) + P_\mu^-(z)]\,\mathrm{d}z + g(L)\int_0^L \frac{\alpha_{\mathrm{RBS}_\nu}(z)}{g^2(z)} \int_z^L g(z') \sum_{\mu>\nu} n_{\mu\nu}\frac{g_{\mu\nu}}{A_{\mu\nu}} \times [P_\mu^+(z') + P_\mu^-(z')]\,\mathrm{d}z'\,\mathrm{d}z \right\} , \tag{3.56}$$

where the first term generalizes the regular ASE of (3.44) and the second term generalizes the backscattered ASE of (3.49). Note also that (3.50), (3.55), and (3.56) are also applicable to multiple-fiber spans [3.33, 51–53], in which case α, α_RBS, $g_{\mu\nu}$, and $A_{\mu\nu}$ become z-dependent (and in (3.50) one α_RBS is a function of z, and the other α_RBS is a function of z').

3.2.4 Summary and Outlook

The low-noise properties of distributed Raman amplification have already made it indispensable for transmission over extralong spans as well as over ultralong-haul links. The extreme OSNR demands of advanced modulation formats, such as high-order quadrature amplitude modulation (QAM), might soon lead to the use of Raman amplifiers even in relatively short-reach systems. Raman amplification can also mitigate the effect of large component losses in fiber-based signal-processing devices and systems [3.69] and lead to the creation of ultralong distributed feedback fiber lasers [3.70].

3.3 Alternative Amplification Technologies

While EDFAs and Raman amplifiers clearly dominate the communication field today and are expected to retain this role in the foreseeable future, it is of interest to note the existence of alternative amplification technologies. These primarily include devices based on parametric optical interaction in nonlinear waveguides and semiconductor optical amplifiers (SOAs). The latter class of amplifiers has a long history that exceeds that of either EDFA and Raman development, dating back to the early 1960s [3.71–74]. One can also regard a semiconductor optical amplifier as a laser device with removed reflective facets, allowing for subthreshold operation. Indeed, it is natural that SOA technology had to be developed before the onset of EDFA and Raman technologies, since semiconductor pumps, which underlie both of these, inherently incorporate semiconductor gain. Depending on the facet reflectivity, an SOA represents either a pure traveling-wave device [3.75, 76] or a bandwidth-limited amplifier, often referred to as Fabry–Pérot SOA types [3.77, 78]. The underlying SOA gain mechanism rests on electron–hole recombination that provides the energy for optical emission. While SOAs can match multiple EDFA operational properties, an important set of differences distinguish the two types and are readily recognized in deployment. In contrast to an EDFA or Raman amplifier, an SOA is pumped electrically and is more efficient. the gain characteristics of SOA material platforms allow for amplification not only within the standard telecommunications bands (1550 and 1300 nm) but also in distant spectral windows.

Despite their long development history, SOAs are not regarded as viable gain elements in conventional lightwave communication links. The incompatibility of SOAs with WDM transmission schemes is the greatest reason for the distinctively niche role that these devices still play. The SOA excited state is orders of magnitude shorter than that of EDFA, resulting in gain that can change rapidly with the signal level. While tolerable in single-channel amplification, in the presence of multiple (independently modulated) channels, this fast SOA dynamics leads to significant impairments that need to be carefully mitigated [3.79, 80]. However, this ultrafast response, which is undesirable in the case of WDM gain elements, has been successfully leveraged in a host of signal processing devices, providing an efficient means for wavelength conversion, signal regeneration, and switching [3.81, 82].

Parametric amplification is the second class of optical amplification that has been widely investigated in multiple material platforms. Similar to Raman amplification, the parametric interaction in silica fiber is not limited to a specific, ion-defined spectral band. This feature was recognized early [3.83, 84] and has justified significant effort on the development of specialty fiber types specifically designed to maximize the parametric gain [3.85, 86]. Parametric devices are also capable of selective quadrature amplification, a unique feature that has been successfully used in signal processing and quantum field manipulation [3.87]. While parametric amplifiers can provide gain–bandwidth products that cannot be matched by any comparable technology [3.88], their use in WDM transport is also limited by interchange crosstalk that must be strictly managed. Consequently, it is expected that parametric devices will find much wider use in signal processing roles [3.84, 89] than in ordinary amplification applications.

References

3.1 P.W. Milonni, J.H. Eberly: *Lasers* (Wiley, New York 1988)

3.2 J.R. Armitage: Three-level fiber amplifier: a theoretical model, Appl. Opt. **27**, 4831 (1988)

3.3 Y. Ohishi, T. Kanamori, T. Kitagawa, S. Takahashi, E. Snitzer, G.H. Sigel: Pr-doped fluoride fiber amplifier operating at 1.31 mm. In: *Opt. Fiber Conf.* (1991), Paper PD 2

3.4 T. Sugawa, Y. Miyajima, T. Komukai: 10 dB gain and high saturation power in a Nd doped fluorozirconate fibre amplifier, Electron. Lett. **26**, 2042 (1990)
3.5 R.J. Mears, L. Reekie, I.M. Jauncey, D.N. Payne: Low noise erbium-doped traveling-wave fiber amplifier, Electron. Lett. **23**, 1026 (1987)
3.6 ofs: https://fiber-optic-catalog.ofsoptics.com/Products/Optical-Fibers/Rare-Earth-Doped-Optical-Fibers/Erbium-Doped-Optical-Fiber-3100100889
3.7 C.R. Giles, D. DiGiovanni: Spectral dependence of gain and noise in erbium-doped fiber optical amplifiers, IEEE Photonics Technol. Lett. **2**, 797 (1990)
3.8 C.R. Giles, E. Desurvire: Modeling erbium-doped fiber amplifiers, J. Lightwave Technol. **9**, 271 (1991)
3.9 T. Georges, E. Delevaque: Analytic modeling of high-gain erbium-doped fiber amplifiers, Opt. Lett. **17**, 1113 (1992)
3.10 E. Desurvire: *Erbium-doped Fiber Amplifiers: Principles and Applications* (Wiley, New York 2002)
3.11 A.A.M. Saleh, R.M. Jopson, J.D. Evankow, J. Aspell: Modeling of gain in erbium-doped fiber amplifiers, IEEE Photonics Technol. Lett. **2**, 714 (1990)
3.12 Y. Sun, J.L. Zyskind, A.K. Srivastava: Average inversion level, modeling, and physics of erbium-doped fiber amplifiers, J. Sel. Areas Quantum Electron. **3**, 991 (1997)
3.13 E. Desurvire, J.L. Zyskind, J.R. Simpson: Spectral gain hole burning at 1.53 microns in erbium-doped fiber amplifiers, IEEE Photonics Technol. Lett. **2**, 246 (1990)
3.14 M.O. Deventer: *Fundamentals of Bidirectional Transmission over a Single Optical Fibre* (Springer, Berlin, Heidelberg 1996)
3.15 S. Radic, S. Chandrasekhar: Limitations in dense bidirectional transmission in absence of optical amplification, IEEE Photonics Technol. Lett. **14**, 95 (2002)
3.16 S. Radic, S. Chandrasekhar, A. Srivastava, H. Kim, L. Nelson, S. Liang, K. Tai, N. Copner: Dense interleaved bidirectional transmission over 5×80 km of non-zero dispersion shifted fiber, IEEE Photonics Technol. Lett. **14**, 218 (2002)
3.17 S. Radic, S. Chandrasekhar, A. Srivastava, H. Kim, L. Nelson, S. Liang, K. Tai, N. Copner: Limitations imposed by Rayleigh backscattering in closely interleaved, bidirectional WDM transmission systems, IEEE Photonics Technol. Lett. **15**, 150 (2003)
3.18 M. Oguma, T. Kitoh, K. Jinguji, T. Shibata, A. Himeno, Y. Hibino: Flat-top and low-loss WDM filter composed of lattice-form interleave filter and arrayed-waveguide gratings on one chip. In: *Opt. Fiber Conf* (2001), Paper WB3
3.19 K.O. Hill, D.C. Johnson, B.S. Kawasaki, R.I. MacDonald: CW three-wave mixing in single-mode optical fibers, J. Appl. Phys. **49**, 5098 (1978)
3.20 A.K. Srivastava, S. Radic, C. Wolf, J.C. Centanni, J.W. Sulhoff, K. Kantor, Y. Sun: Ultradense WDM transmission in L-band, IEEE Photonics Technol. Lett. **12**, 1570 (2000)
3.21 S. Radic, G. Pendock, A. Srivastava, P. Wysocki, A. Chraplyvy: Four-wave mixing in optical links using quasi-distributed optical amplifiers, IEEE J. Lightwave Technol. **19**, 636 (2001)
3.22 C.V. Raman, K.S. Krishnan: A new type of secondary radiation, Nature **121**(3048), 501 (1928)
3.23 G. Landsberg, L. Mandelstam: Eine neue Erscheinung bei der Lichtzerstreuung in Krystallen, Naturwissenschaften **16**(28), 557–558 (1928)
3.24 E.J. Woodbury, W.K. Ng: Ruby laser operation in the near IR, Proc. IRE **50**, 2347–2348 (1962)
3.25 R.H. Stolen, E.P. Ippen, A.R. Tynes: Raman oscillation in glass optical waveguide, Appl. Phys. Lett. **20**, 62 (1972)
3.26 R.H. Stolen, E.P. Ippen: Raman gain in glass optical waveguides, Appl. Phys. Lett. **22**(6), 276–281 (1973)
3.27 Y. Aoki, S. Kishida, K. Washio, K. Minemura: Bit error rate evaluation of optical signals amplified via stimulated Raman process in optical fibre, Electron. Lett. **21**, 191–192 (1985)
3.28 L.F. Mollenauer, R.H. Stolen, M.N. Islam: Experimental demonstration of soliton propagation in long fibers: loss compensated by Raman gain, Opt. Lett. **10**(5), 229–231 (1985)
3.29 J. Hegarty, N.A. Olsson, L. Goldner: CW pumped Raman preamplifier in a 45-km-long fibre transmission system operating at 1.15 μm and 1 Gbit/s, Electron. Lett. **21**(7), 290–292 (1985)
3.30 P.B. Hansen, L. Eskildsen: Remote amplification in repeaterless transmission systems, Opt. Fiber Technol. **3**, 221–237 (1997)
3.31 S. Namiki, Y. Emori: Ultrabroad-band Raman amplifiers pumped and gain-equalized by wavelength-division-multiplexed high-power laser diodes, IEEE J. Sel. Top. Quantum Electron. **7**(1), 3–16 (2001)
3.32 K. Rottwitt, A.J. Stentz: Raman amplification in lightwave communication systems. In: *Optical Fiber Telecommunications IVA: Components*, ed. by I.P. Kaminow, T. Li (Academic Press, San Diego 2002) pp. 213–257, Chapter 5
3.33 A.F. Evans, A. Kobyakov, M. Vasilyev: Distributed Raman transmission: applications and fiber issues. In: *Raman Amplifiers for Telecommunications 2: Sub-Systems and Systems*, ed. by M.N. Islam (Springer, New York 2004) pp. 383–412, Chapter 12
3.34 S.R. Chinn: Analysis of counter-pumped small-signal fibre Raman amplifiers, Electron. Lett. **33**(7), 607–608 (1997)
3.35 R.G. Smith: Optical power handling capacity of low loss optical fibers as determined by stimulated Raman and Brillouin scattering, Appl. Opt. **11**, 2489–2494 (1972)
3.36 V.E. Perlin, H.G. Winful: Optimizing the noise performance of broad-band WDM systems with distributed Raman amplification, IEEE Photonics Technol. Lett. **14**, 1199–1201 (2002)
3.37 E. Desurvire: *Erbium-Doped Fiber Amplifiers: Principles and Applications* (Wiley, New York 1994)
3.38 M. Nissov, K. Rottwitt, H.D. Kidorf, M.X. Ma: Rayleigh crosstalk in long cascades of distributed unsaturated Raman amplifiers, Electron. Lett. **35**(12), 997–998 (1999)
3.39 C.R.S. Fludger, R.J. Mears: Electrical measurements of multipath interference in distributed Raman amplifiers, J. Lightwave Technol. **19**, 536–545 (2001)

3.40 S.A.E. Lewis, S.V. Chernikov, J.R. Taylor: Characterization of double Rayleigh scatter noise in Raman amplifiers, IEEE Photonics Technol. Lett. **12**, 528–530 (2000)

3.41 P.P. Wan, J. Conradi: Impact of double Rayleigh backscatter noise on digital and analog fiber systems, J. Lightwave Technol. **14**, 288–297 (1996)

3.42 R.-J. Essiambre, P. Winzer, J. Bromage, C.H. Kim: Design of bidirectionally pumped fiber amplifiers generating double Rayleigh backscattering, IEEE Photonics Technol. Lett. **14**, 914–916 (2002)

3.43 M.O. van Deventer: Polarization properties of Rayleigh backscattering in single-mode fibers, J. Lightwave Technol. **11**, 1895–1899 (1993)

3.44 S. Namiki, Y. Emory, A. Oguri: Discrete Raman amplifiers. In: *Raman Amplification in Fiber-Optic Communication Systems*, ed. by C. Headley, G.P. Agarwal (Elsevier, San Diego 2005) pp. 169–213, Chapter 4

3.45 K. Rottwitt, H.D. Kidorf: A 92-nm bandwidth Raman amplifier. In: *Proc. Opt. Fiber Commun. Conf.* (1998), Paper PD6

3.46 M. Vasilyev, B. Szalabofka, S. Tsuda, J.M. Grochocinski, A.F. Evans: Reduction of Raman MPI and noise figure in dispersion-managed fibre, Electron. Lett. **38**, 271–272 (2002)

3.47 R. Hainberger, J. Kumasako, K. Nakamura, T. Terahara, H. Osaka, T. Hoshida: Comparison of span configurations of Raman-amplified dispersion-managed fibers, IEEE Photonics Technol. Lett. **14**, 471 (2002)

3.48 S.N. Knudsen, B. Zhu, L.E. Nelson, M.Ø. Pederson, D.W. Peckham, S. Stulz: 420 Gbit/s (42 × 10 Gbit/s) WDM transmission over 4000 km of ultra wave fibre with 100 km dispersion-managed spans and distributed Raman amplification, Electron. Lett. **37**, 965–967 (2001)

3.49 M. Mehendale, M. Vasilyev, A. Kobyakov, M. Williams, S. Tsuda: All-Raman transmission of 80 × 10 Gb/s WDM signals with 50 GHz spacing over 4160 km of dispersion-managed fiber, Electron. Lett. **38**, 648–649 (2002)

3.50 M. Vasilyev: Raman-assisted transmission: toward ideal distributed amplification. In: *Opt. Fiber Comm. Conf. 2003*, Technical Digest, Vol. 1 (OSA, Washington, D.C. 2003) pp. 303–305, Paper WB1

3.51 A. Kobyakov, S. Gray, M. Vasilyev: Quantitative analysis of Rayleigh crosstalk in Raman amplifiers, Electron. Lett. **39**, 732–733 (2003)

3.52 A. Kobyakov, M. Vasilyev, S. Tsuda, G. Giudice, S. Ten: Analytical model for Raman noise figure in dispersion-managed fibers, IEEE Photonics Technol. Lett. **15**, 30–32 (2003)

3.53 A. Kobyakov: Prospects of Raman-assisted transmission systems, Proc. SPIE **5246**, 174–188 (2003)

3.54 J.-C. Bouteiller, K. Brar, C. Headley: Quasi-constant signal power transmission. In: *Proc. ECOC*, Vol. 3 (2002) pp. 1–2

3.55 C.R.S. Fludger, V. Handerek, R.J. Mears: Pump to signal RIN transfer in Raman fiber amplifiers, J. Lightwave Technol. **19**, 1140 (2001), correction: J. Lightwave Technol. **20**, 316 (2002)

3.56 S.B. Papernyi, V.I. Karpov, W.R.L. Clements: Third-order cascaded Raman amplification. In: *Proc. Opt. Commun. Conf.* (2002), Post-deadline paper FB4

3.57 S. Kado, Y. Emori, S. Namiki, N. Tsukiji, J. Yoshida, T. Kimura: Broadband flat-noise Raman amplifier using low-noise bi-directionally pumping sources. In: *Proc. Eur. Conf. Opt. Commun.*, Vol. 6 (2001) pp. 38–39, Paper PD.F.1.8

3.58 R.P. Espindola, K.L. Bacher, K. Kojima, N. Chand, S. Srinivasan, G.C. Cho, F. Jin, C. Fuchs, V. Milner, W. Dautremont-Smith: Penalty-free 10 Gbit/s single-channel co-pumped distributed Raman amplification using low RIN 14xx nm DFB pump, Electron. Lett. **38**, 113–115 (2002)

3.59 R.E. Neuhauser, P.M. Krummrich, H. Bock, C. Glingener: Impact of nonlinear pump interactions on broadband distributed Raman amplification. In: *Proc. Opt. Fiber Commun. Conf.*, OSA Technical Digest, Vol. 54 (OSA, Washington, D.C. 2001), Paper MA4

3.60 J. Bromage, P.J. Winzer, L.E. Nelson, C.J. McKinstrie: Raman-enhanced pump-signal four-wave mixing in bidirectionally-pumped Raman amplifiers. In: *Conf. Opt. Amplif. Appl., July 2002, Vancouver, Canada* (2002), Paper OWA5

3.61 M. Vasilyev, S. Gray, V.M. Ricci: Pump intensity noise and ASE spectrum of Raman amplification in non-zero dispersion-shifted fibers. In: *Conf. Opt. Amplif. Appl., July 2001, Stresa, Italy* (2001), Paper OMC3

3.62 M. Vasilyev, S. Gray, K. Jepsen: Spectral broadening of double Rayleigh backscattering in a distributed Raman amplifier. In: *Proc. Opt. Fiber Commun. Conf.*, OSA Technical Digest, Vol. 54 (OSA, Washington, D.C. 2001), Paper MA2

3.63 N.R. Newbury: Pump-wavelength dependence of Raman gain in single-mode optical fibers, J. Lightwave Technol. **21**(12), 3364–3373 (2003)

3.64 P. Xiao, Q. Zeng, J. Huang, J. Liu: A new optimal algorithm for multipump sources of distributed fiber Raman amplifier, IEEE Photonics Technol. Lett. **15**, 206–208 (2003)

3.65 X. Liu, H. Zhang, Y. Guo: A novel method for Raman amplifier propagation equations, IEEE Photonics Technol. Lett. **15**, 392–394 (2003)

3.66 X. Liu, B. Lee: Effective shooting algorithm and its application to fiber amplifiers, Opt. Express **11**, 1452–1461 (2003)

3.67 VPIComponentMaker™: Fiber optics software, http://www.vpiphotonics.com/Tools/FiberOptics/

3.68 Synopsys: RSoft OptSim™ software, https://www.synopsys.com/optical-solutions/rsoft/system-network-optsim.html

3.69 L. Li, P.G. Patki, Y.B. Kwon, V. Stelmakh, B.D. Campbell, M. Annamalai, T.I. Lakoba, M. Vasilyev: All-optical regenerator of multi-channel signals, Nat. Commun. **8**, 884 (2017)

3.70 S.K. Turitsyn, S.A. Babin, A.E. El-Taher, P. Harper, D.V. Churkin, S.I. Kablukov, J.D. Ania-Castañón, V. Karalekas, E.V. Podivilov: Random distributed feedback fibre laser, Nat. Photon. **4**, 231–235 (2010)

3.71 W. Dumke: Interband transitions and maser action, Phys. Rev. **127**, 1559–1563 (1962)

3.72 M.G. Bernard, G. Duraffourg: Laser conditions in semiconductors, Phys. Stat. Sol. **1**(7), 699–703 (1961)
3.73 N. Holonyak, S.F. Bevacqua: Coherent (visible) light emission from $Ga(As_{1-x}P_x)$ junctions, Appl. Phys. Lett. **1**, 82 (1962)
3.74 N.G. Basov, O.N. Krokhin, Y.M. Popov: Production of negative-temperature states in p-n junctions of degenerate semiconductors, J. Exp. Theor. Phys. **13**(6), 1320 (1961)
3.75 T. Mukai, K. Inoue, T. Saitoh: Homogeneous gain saturation in 1.5-micrometer InGaAsP travelling-wave semiconductor laser amplifier, Appl. Phys. Lett. **51**, 382–383 (1987)
3.76 T. Saitoh, T. Mukai: 1.5 micrometer GaInAsP traveling-wave semiconductor laser amplifier, IEEE J. Quantum Electron. **23**(6), 1010 (1987)
3.77 Y. Yamamoto: Characteristics of AlGaAs Fabry–Perot cavity type laser amplifiers, IEEE J. Quantum Electron. **16**, 1047 (1980)
3.78 M.J. O'Mahony: Semiconductor laser optical amplifiers for use in future fiber systems, J. Lightwave Technol. **6**, 531 (1988)
3.79 G.P. Agrawal, N.A. Olsson: Self-phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers, IEEE J. Quantum Electron. **25**, 2297 (1989)
3.80 W. Freude: Linear and nonlinear semiconductor optical amplifiers. In: *Proc. 12th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2010), Paper We.D4.1
3.81 Z. Zhu, M. Funabashi, P. Zhong Pan, L. Paraschis, D. Harris: High-performance optical 3R regeneration for scalable fiber transmission system applications, J. Lightwave Technol. **25**, 504–511 (2007)
3.82 O. Qasaimeh: Characteristics of cross-gain (XG) wavelength conversion in quantum dot semiconductor optical amplifiers, IEEE Photonics Technol. Lett. **16**, 542 (2004)
3.83 N. Shibata, R. Braun, R. Waarts: Phase-mismatch dependence of efficiency of wave generation through FWM in SMF, J. Lightwave Technol. **23**, 1205–1210 (1987)
3.84 M.E. Marhic: *Fiber Optical Parametric Amplifiers, Oscillators and Related Devices* (Univ. Press, Cambridge 2007)
3.85 B.P.-P. Kuo, S. Radic: Highly nonlinear fiber with dispersive characteristic invariant to fabrication fluctuations, Opt. Express **20**(7), 7716–7725 (2012)
3.86 B.P.-P. Kuo, M. Hirano, S. Radic: Continuous-wave, short-wavelength infrared mixer using dispersion-stabilized highly-nonlinear fiber, Opt. Express **20**(16), 18422–18431 (2012)
3.87 C.J. McKinstrie, S. Radic, M.G. Raymer, L. Schenato: Unimpaired phase-sensitive amplification by vector four-wave mixing near the zero-dispersion frequency, Opt. Express **15**(5), 2178–2189 (2007)
3.88 S. Moro, A. Peric, N. Alic, A.J. Anderson, C.J. McKinstrie, S. Radic: Continuous-wave parametric amplifier with 6.9-ExaHz gain-bandwidth product, IEEE Photonics Technol. Lett. **23**(20), 1532–1534 (2011)
3.89 S. Radic: Parametric amplification and processing in optical fibers, Laser Photonics Rev. **2**(6), 498–513 (2008)

Michael Vasilyev

Dept. of Electrical Engineering
University of Texas at Arlington
Arlington, TX, USA
vasilyev@uta.edu

Michael Vasilyev received his PhD from Northwestern University (Evanston, IL) in 1999. After working at Corning Inc. (Somerset, NJ), he joined the University of Texas at Arlington in 2003, where he is currently a professor. He works on nonlinear optical signal processing for classical and quantum communications. He is a Fellow of the OSA and SPIE, and a 2008 DARPA Young Faculty Award recipient.

Stojan Radic

Qualcomm Institute
University of California San Diego
La Jolla, CA, USA
sradic@ucsd.edu

S. Radic graduated from The Institute of Optics (Rochester) in 1995 and subsequently served in Corning and Bell Laboratories. He holds a BSE in electrical engineering, an MA in physics, and a PhD in optical engineering. He is chaired Professor of Electrical Engineering and Director of the Photonics System Laboratory at Qualcomm Institute. He is a Fellow of the Optical Society of America and IEEE.

4. Optical Transponders

Gabriella Bosco, Jörg-Peter Elbers

The first commercial 10-Gb/s transponders, deployed in the mid 1990s, were based on a very simple modulation technique, i.e., a binary light intensity modulation with envelope detection by a single photodiode. To extend the fiber capacity, bandwidth-efficient modulation techniques such as duobinary line coding and multilevel intensity-modulation formats gained popularity in optical communications in the late 1990s. In the following years, the use of differential phase modulation in combination with interferometric detection allowed the transponder data rates to be increased up to 40 Gb/s. However, despite all improvements, the system performance of these 40-Gb/s solutions was still not on par with state-of-the art 10-Gb/s systems at that time. With the advent of coherent detection, things suddenly changed and transmission rates of 100 Gb/s and beyond could soon be achieved, thanks to the use of high-order modulation formats and advanced digital signal-processing techniques.

In this chapter, the configuration and performance of the most common transmitter and receiver combinations that are currently used in optical transmission systems will be described, including an overview of transponder types and their hardware architectures. Finally, relevant standards will be discussed and pluggable optical transceiver modules used in modern transponder implementations will be explained.

Transponders are essential building blocks in any optical communication system. The term *transponder* stems from an amalgamation of two words *trans(mitter)* and *(res)ponder*, first coined in about 1940 to describe radio and radar systems [4.1]. In optical communications, a transponder is a receiver-transmitter combination, which translates an optical signal from one wavelength into another optical signal at a different wavelength by means of optical-electrical-optical (O/E/O) conversion [4.2]. During the conversion process, the received signal is reamplified and normally also reshaped and retimed, i.e., 3R regenerated. Optical transponders commonly operate in full-duplex mode and comprise transmitter and receiver functions on each of their ports. The ports can support different optical interface classes, which are characterized by properties such as the fiber medium, link distance, wavelength range, signaling rate, modulation format, forward error correction, line coding, and/or signal type.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_4

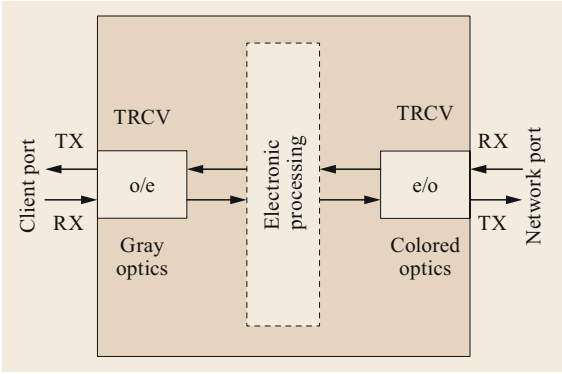


Fig. 4.1 Basic block diagram of an optical transponder (TX: transmitter, RX: receiver, o/e: opto-electric conversion, e/o: electro-optic conversion, TRCV: transceiver)

Network (or line) ports are connected to a multichannel wavelength-division multiplexing (WDM) networks. They are also called *colored interfaces*, as they are available in or can be tuned to multiple WDM colors. Client ports typically connect to subtended equipment such as routers or switches which act as clients of the optical transport network. These ports are termed *gray interfaces* if they only support a single channel per fiber (as opposed to colored interfaces, which can be wavelength multiplexed onto the same fiber). Wavelength grids for dense (DWDM) and coarse wavelength-division multiplexing (CWDM) are defined in the ITU-T G.694.1 and G.694.2 recommendations [4.3, 4]. In this chapter, the focus will be on colored DWDM interfaces.

Another very similar device to the transponder, and might be confused with it, is the *transceiver* (TRCV) which is a device that can simultaneously transmit and receive signals. As a portmanteau of *trans(mitter)* (TX) and *(re)ceiver* (RX), the term *transceiver* first appeared in the early 1920s and was used in descriptions of military communication equipment [4.5]. An optical transceiver comprises an optical transmit and receive interface on the so-called line side and a bidirectional electrical interface on the so-called host side. Transceivers can be realized as pluggable optical modules or as fixed host board implementations. An optical transponder can then be split into a client and line side optical transceiver function with an electronic processing block in between (Fig. 4.1).

In most cases, optical transponders are delivered as part of a complete DWDM network solution by one vendor. In other cases, transponders and the corresponding DWDM systems are supplied independently by different vendors. Colored interfaces may also be integrated into third-party switches or routers on the client side, thereby eliminating the need for separate transponders. When colored interfaces are separated from a WDM system and externally provided, then the scenario is referred to as an alien wavelength or black link approach [4.6] and has been used in commercial networks for many years. More recently, WDM systems supporting alien wavelengths have also been termed *open line systems* [4.7]. A line system is not restricted to just a point-to-point link in this context, but also includes optical networks in ring or mesh topologies.

A typical WDM system is depicted in Fig. 4.2. For simplicity, only one direction is shown. Colored interfaces are connected to a WDM multiplexer, which combines all of the WDM transmit channels onto one fiber before they are optically amplified by a booster amplifier and sent over a fiber link. The fiber link comprises one or more fiber spans of between 50 and 120 km in length (80 km is a typical span length in terrestrial systems). Erbium-doped fiber amplifiers (EDFAs) and/or distributed Raman amplifiers periodically compensate the span loss. At the receiving end, the WDM channels are optically preamplified, demultiplexed, and terminated by colored receiver interfaces. If access to the WDM channels is required at intermediate sites then reconfigurable optical add-drop multiplexers (ROADMs) are used on the transmission link. Multidegree ROADMs (MD-ROADMs) can offer additional wavelength-selective cross-connect functionality for dynamic routing of channels across a meshed network.

The colored interface performance is limited by intersymbol interference (ISI), thermal noise (THN), cross-talk (XTK), laser phase noise (LPN), laser sidemode suppression ratio (SMSR), and optical reflections (RFL). Transmission impairments such as amplified spontaneous emission noise (ASE), spectral loss/gain tilt and ripples, polarization-dependent loss/gain (PDL/PDG), group velocity dispersion (GVD), and polarization-mode dispersion (PMD) limit the optical performance of the channels. With higher optical power levels then fiber nonlinearities will be triggered which will limit the transmission performance further. Nonlinearities include stimulated Raman scattering (SRS) leading to Raman tilt and cross-talk, stimulated Brillouin scattering (SBS) which imposes an upper limit on optical channel carrier powers, and the Kerr effect which introduces intrachannel and interchannel nonlinearities. The Kerr nonlinearities are caused by signal–signal or signal–noise interactions and can be separated into a number of manifestations as shown below in Table 4.1. More specifically, the intrachannel nonlinearities are self-phase modulation (SPM)-induced nonlinear phase noise (NLPN), modulation instability (MI), isolated pulse SPM, intrachannel cross-phase modulation (IXPM), and intrachannel four-wave mixing (IFWM). The interchannel nonlinearities can be

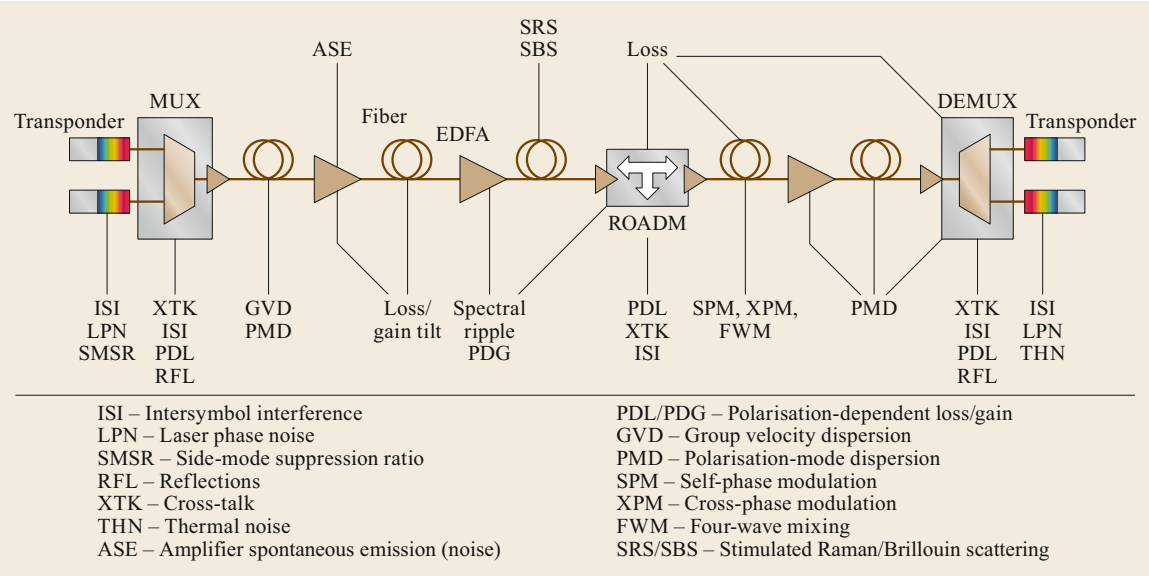


Fig. 4.2 A typical WDM line system and related transmission impairments. MUX: WDM multiplexer, DEMUX: WDM demultiplexer

Table 4.1 Transmission impairments due to fiber Kerr nonlinearity [4.8] (SPM: self-phase modulation, XPM: cross-phase modulation, FWM: four-wave mixing, NLPN: nonlinear phase noise, MI: modulation instability, IXPM: intra-channel cross-phase modulation, IFWM: intrachannel four-wave mixing)

Intrachannel nonlinearities					**Interchannel nonlinearities**		
Signal–noise interaction		Signal–signal interaction			Signal–noise interaction	Signal–signal interaction	
NLPN	Parametric amplification	SPM			NLPN	WDM!nonlinearities	
SPM-induced NLPN	MI	Isolated pulse SPM	IXPM	IFWM	XPM-induced NLPN	XPM	FWM

split into XPM-induced NLPN and interchannel XPM and FWM.

The choice of colored interface technology depends on the network application and is typically determined by such considerations as cost, performance, spectral efficiency, power consumption, and equipment footprint. Point-to-point enterprise and data-center interconnect (DCI) links can be distinguished from ring-based metro-access and metro-core networks as well as meshed regional and long-haul networks on the typical reaches involved, the number of ROADMs and the fiber cost, as shown in (Fig. 4.3). Submarine sys-

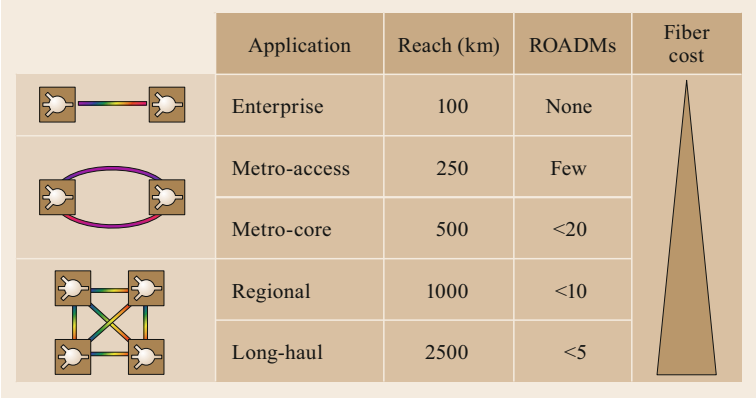


Fig. 4.3 Terrestrial WDM applications and main characteristics

tems can be separated into unrepeated point-to-point systems with passive wet plant over link distances of less than 600 km and repeated ultralong haul systems over link distances of several thousand kilometers. The larger the network reach, the higher is the fiber infrastructure cost and the more important is the spectral efficiency, but the most challenging are the transmission impairments. It is important to note that cascading of ROADMs also causes impairments such as insertions losses, intersymbol interference and cross-talk, so that a metro network with many ROADMs can show a similar performance to a long-haul network with few or no ROADMs despite the difference in network size.

4.1 Evolution of Modulation Formats and Symbol Rates

Because of its simplicity and cost efficiency, nonreturn-to-zero (NRZ) modulation with direct detection has long been the solution of choice for terrestrial DWDM systems. On the basis of a binary light on/off approach, this modulation technique is also known as on–off keying (OOK) [4.9]. OOK signals can be generated using an intensity-modulation (IM) transmitter, in which the intensity of an optical source is modulated by an RF signal, and demodulated using a direct-detection (DD) receiver, which converts the optical power into an electrical signal current using a photodetector. For this reason, optical systems based on OOK modulation are also classified as IMDD systems. Commercial 10-Gb/s NRZ transponders have been available since 1995 and were designed to guarantee a bit error ratio (BER) of 10^{-15} without forward error correction. They were deployed in 32-channel DWDM systems with 100-GHz channel spacing, yielding a total system capacity of 320 Gb/s over distances beyond 500 km of standard single-mode fiber (SSMF) in the C-band (1530−1565 nm) [4.10]. Motivated by the higher tolerance to fiber-nonlinearities, return-to-zero (RZ) modulation was introduced as an alternative to NRZ modulation. RZ modulation was first used in submarine systems [4.11] but also found commercial adoption in ultralong-haul terrestrial networks [4.12]. Aided by forward error correction (FEC) and Raman amplification, terrestrial DWDM systems with 10-Gb/RZ modulation could achieve C-band capacities of 800 Gb/s (80 channels at 50-GHz channel spacing) over distances of more than 3000-km SSMF [4.13]. By adding another 80 channels in the L-band (1565−1625 nm), this capacity could be doubled, albeit with some impact on reach [4.14]. Many terrestrial 10-Gb/s DWDM systems continued to use NRZ modulation, though. As enhanced forward error correction (EFEC) helped to increase the transmission distance up to 1500 km with EDFA-only amplification and more than 2500 km with Raman amplification, this was deemed to be sufficient for most practical applications.

To extend the fiber capacity, more bandwidth-efficient modulation techniques such as duobinary line coding and multilevel IMDD formats gained popularity in optical communications in the late 1990s [4.15]. Duobinary line coding (DB) is a partial-response format [4.9] and was first proposed in the 1960s for radio communications [4.16]. Comprehensive review papers on the advantages and disadvantages of optical duobinary (ODB) modulation have been published [4.17]. It has been pointed out that ODB, besides achieving higher bandwidth efficiency, also features a high resilience to GVD [4.18–21]. Further research has been published showing that improvements in the sensitivity of duobinary receivers is possible through careful optimization of both the electric and optical filter bandwidths in the system [4.22–26]. At 10-Gb/s line rate, ODB found only little commercial adoption apart from single-span transmission over extended link distances (for example 150 km) without optical dispersion compensation. At 40 Gb/s, though, ODB was found to be much more applicable, as it allows transmission of 40-Gb/s channels using 50-GHz filters and therefore a quadrupling of the system capacity compared to 10-Gb/s channels. It then substituted earlier realizations based upon NRZ or RZ modulation. In 2001, ODB modulation was successfully employed to transmit 80 channels at 40 Gb/s over 300-km SSMF [4.27]. Commercial deployments over several hundred kilometers followed in around 2005 [4.28]. Though ODB was by far not the only option for 40-Gb/s transmission.

The use of differential phase-shift keying (DPSK) with interferometric demodulation and direct detection was investigated in the late 1990s. DPSK demonstrated a sensitivity advantage over OOK [4.29, 30] and ODB as well as a high resilience to linear and nonlinear fiber propagation effects [4.25, 31–35]. Both binary (DPSK) and quaternary formats (DQPSK) were extensively studied in the following years, leading to their commercial deployment in around 2006 [4.36, 37]. While 40-Gb/s DQPSK could fit into a 50-GHz channel grid, this proved challenging initially for DPSK. Partial DPSK was able to overcome this limitation by reducing the differential delay of the Mach–Zehnder demodulator [4.38] and was commercially deployed from 2007

onwards under the name adaptive DPSK (ADPSK). Despite all improvements, the system performance of these 40-Gb/s solutions was still not on a par with state-of-the-art 10-Gb/s systems at that time. Stringent requirements on GVD and PMD tolerance required careful network planning and led to operational challenges.

With the advent of coherent detection, the transponder landscape changed dramatically. Earlier work on coherent optical communication systems in the 1980s had been hampered by a lack of digital signal processing (DSP) and large laser linewidths (compared to the Mb/s data rates used at that time). With much higher data rates and advances in CMOS technology now, coherent optical communications suddenly became a very practical option. Polarization-multiplexed quaternary phase-shift keying (PM-QPSK) reduces symbol rates by a factor of four compared to the information bit rate, while at the same time offering a 6-dB sensitivity gain over OOK. Furthermore, powerful DSP performs all the required GVD and polarization mode dispersion compensation, frequency and phase locking, and polarization demultiplexing in the electronic domain [4.39–45]. As a result, coherent transmission can deliver comparable system performance and reach as 10-Gb/s OOK and 40-Gb/s PM-QPSK signals. Furthermore, optical dispersion compensation is no longer required, a major operational advantage. Commercial 40-Gb/s PM-QPSK systems were first introduced in 2007 [4.46].

With the development of the 100-Gb Ethernet, a modulation format for its onward transmission across an optical DWDM network was urgently required. This triggered the development of 100-Gb/s PM-QPSK, which rapidly superseded 40-Gb/s technology. The higher capacity of a 100-Gb/s optical carrier allowed much better cost per bit performance compared to what was achievable at 40 Gb/s. The first commercial single-carrier 100-Gb/s PM-QPSK solutions were introduced in 2010 [4.47]. Competing solutions which achieved 100-Gb/s capacity by bonding two 50-Gb/s PM-QPSK carriers together (2C-PM-QPSK) [4.48, 49] had only limited market success.

Coherent transponder technology was first introduced in long-haul transmission and subsequently penetrated other areas such as metro and submarine networks as well. For shorter metro links and interdata center interconnects, the performance and spectral efficiency of coherent solutions is not always required, where the cost and power consumption of this technology can be prohibitive. As a consequence, alternative solutions based on direct-detect technology have emerged to address this market segment. A 4×25-Gb/s multicarrier ODB metro DWDM solution was introduced in 2011 to fulfill these needs for lower cost and lower power consumption and can operate on a 50 GHz or smaller channel grid to maximize spectral efficiency [4.50]. Using multilevel modulation formats in which each constellation symbol carries more than one bit, the spectral efficiency can be further increased. The transmission bandwidth of an M-ary signal (i.e., a modulation format whose constellation is composed of M points) scales at a factor of $1/\log_2(M)$ compared to a binary signal operating at the same bit rate. Multilevel IMDD systems are based on the use of the pulse-amplitude modulation (PAM), where only the amplitude of the optical field is modulated. Since the constellation points for this modulation format are *unipolar* (i.e., only zero or positive values can be encoded), a non-negligible loss in sensitivity is experienced. For this reason, multilevel IMDD formats have mainly been used in shorter reach applications. Utilizing the PAM4 modulation format to support the IEEE802.3bs and the OIF CEI56 standards, a dual-carrier 25-GBd PAM4 solution (two carriers occupying 100 GHz of bandwidth) was demonstrated in 2016 for DWDM data-center interconnects over 80-km distance [4.7], offering twice the fiber capacity compared to earlier 25-GBd NRZ or ODB OOK implementations.

Higher order modulation has also improved the fiber capacity and cost per bit in coherent communications employing quadrature-amplitude modulation (QAM). Maintaining a symbol rate of about 30 GBd, 200-Gb/s PM-16QAM [4.51] (and 150-Gb/s PM-8QAM [4.52]) was first commercially introduced in 2014 providing 100% (or 50%) more channel and fiber capacity at the cost of roughly 25% (or 50%) of the 100-Gb/s PM-QPSK reach. 200-Gb/s 8QAM [4.53] and 400-Gb/s 32QAM [4.54] have since then been presented as intermediate evolution steps. The next step change, though, is expected to be to 60-GBd technology, which has already been chosen as the basis for the 400G-ZR standard currently under development by the OIF [4.55]. 60-GBd technology doubles the symbol rate per optical carrier and can be used with a variety of modulation formats, for example 200-Gb/s PM-QPSK, 300-Gb/s PM-8QAM, 400-Gb/s 16-QAM, 500-Gb/s 32-QAM, and 600-Gb/s 64 QAM. Real-time 400-Gb/s PM-16QAM was first demonstrated in 2015 [4.56]. While doubling the symbol rate when maintaining the modulation format does not yield spectral efficiency gains for a given link distance, the use of fewer optical carriers results in cost, footprint, and power savings. Moving forwards, future developments will concentrate on (a) increasing the symbol rate even further and (b) providing more flexibility by adapting the modulation format to the requirements and available margins on a given optical path. Figure 4.4 graphically summarizes the historical development of WDM transponder symbol rates and modulation formats in commercial systems.

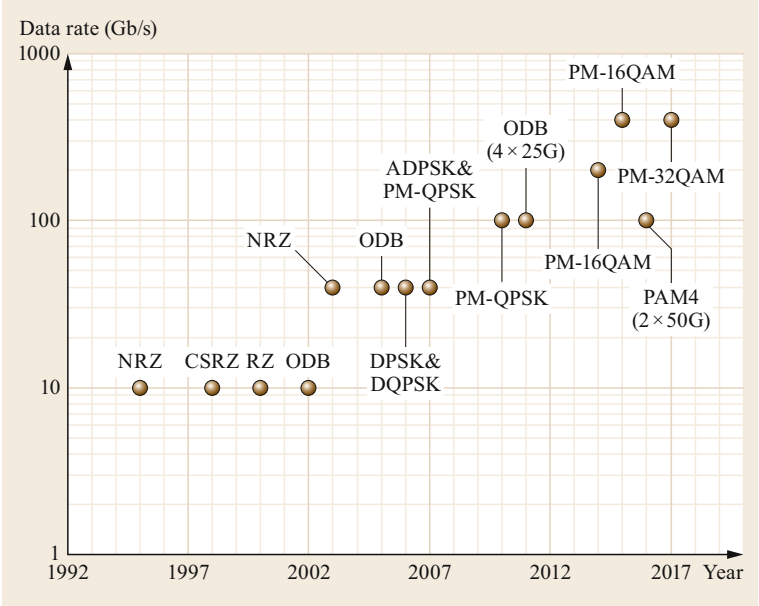


Fig. 4.4 Historical development of WDM transponder symbol rates and modulation formats

In the following sections, the architecture and performance of DWDM transponder line interfaces employing different modulation formats and detection schemes will be discussed. Section 4.2 starts with IMDD, to be followed by ODB in Sect. 4.3, then differential phase modulation in Sect. 4.4 and coherent interfaces in Sect. 4.5. Different transponder types and hardware architectures will then be explained in Sect. 4.6.

4.2 Intensity Modulation with Direct Detection

This section reviews the transmitter and receiver architectures required to generate and detect different types of IMDD optical signals. The intensity-modulation formats are introduced in Sect. 4.2.1, together with the most common line coding techniques used in IMDD transponders. The structures of several types of binary intensity-modulation transmitters are detailed in Sect. 4.2.2, while Sect. 4.2.3 describes the architecture of a receiver based on direct detection (DD). Section 4.2.4 shows the back-to-back performance of binary IMDD systems, including the optimization of the TX and RX filters and the impact of subsystem components implementation. Tolerance to GVD and PMD is also analyzed. Finally, Sect. 4.2.5 discusses multilevel IMDD formats.

4.2.1 Intensity Modulation

The complex envelope of an IM signal at the output of the modulator can be written as [4.9]

$$s_{\mathrm{T}}(t) = \sum_k a_k q_{\mathrm{T}}(t - kT_{\mathrm{s}}) \,, \tag{4.1}$$

where a_k are the symbols carrying the information, T_{s} is the symbol time interval (one symbol is sent every T_{s} seconds) and $q_{\mathrm{T}}(t)$ is the optical pulse shape in the time domain. The symbols a_k may assume $M = 2^{n_{\mathrm{b}}}$ values, where n_{b} is the number of information bits carried by each symbol. The symbol rate R_{s} is defined as the number of symbols that are transmitted in each time unit. It is equal to the inverse of the symbol time ($R_{\mathrm{s}} = 1/T_{\mathrm{s}}$) and is measured in baud or symbols per second. The bit rate R_{b} is defined as the number of bits that are conveyed per unit of time. It is measured in bit per second and is related to the symbol rate by $R_{\mathrm{b}} = n_{\mathrm{b}} R_{\mathrm{s}}$. In the case of binary IM, also known as on-off keying (OOK), $n_{\mathrm{b}} = 1$ and $a_k \in \{0, 1\}$. In each bit slot, one of two alternative waveforms can be transmitted, depending on the value of the information bit: $s_0(t) = 0$, $s_1(t) = q_{\mathrm{T}}(t)$. These two analog waveforms are represented in the signal space [4.9] by the two points shown in Fig. 4.5a. The constellation of quaternary IM, also known as PAM-4 (pulse amplitude modulation), is shown, as well. In practice, a *constellation diagram* shows all the possible symbols that can be transmitted by the system as a *constellation* of points,

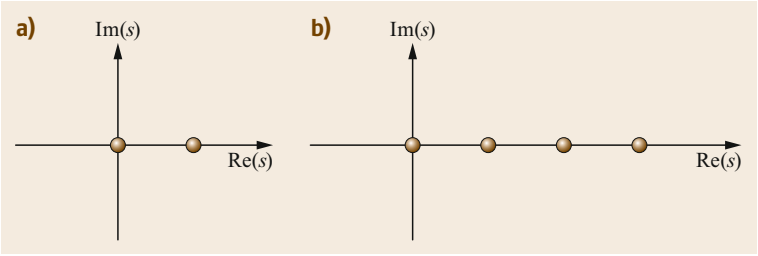


Fig. 4.5 (a) OK and (b) PAM-4 constellations

each characterized by an amplitude (distance from the origin) and a phase. The energy of each symbol is proportional to the square distance of the corresponding constellation point from the origin. For IM signals, only the intensity (amplitude) varies among different symbols, while the phase is constant, thus the constellation points lie on a single straight line.

Depending on the shape of $q_T(t)$ in (4.1), the modulation format is defined as either nonreturn-to-zero (NRZ), when the time duration of $q_T(t)$ is close to the symbol interval T_s, or return-to-zero (RZ), when the duration of the pulse is lower than T_s and the power returns to zero between two successive "1" bits.

The power spectral density of the signal $s_T(t)$ is given by [4.9]

$$G_s(f) = \frac{|Q_T(f)|^2}{T_s}\sigma_a^2 + m_a^2 \sum_n \frac{|Q_T(nR_s)|^2}{T_s^2}\delta(f - nR_s)\,, \qquad (4.2)$$

where $Q_T(f)$ is the Fourier transform of the pulse $q_T(t)$ and m_a and σ_a^2 are the mean and variance of the random variable a_k, respectively. For instance, in case of OOK with equi-probable bits, $m_a = \sigma_a^2 = 1/2$. The power spectrum of IM signals is composed of two terms: a continuous spectrum, which contains the information to be transmitted, and a discrete spectrum, which does not carry any information and whose lines are placed at the harmonics of the bit rate.

Figure 4.6 shows two examples of noiseless spectra of NRZ and RZ signals. The time duration of RZ pulses is shorter, therefore their bandwidth occupation can be significantly larger. The *eye diagram*, i.e., the synchronized superposition of all possible realizations of a signal viewed within a particular signaling interval (equal to $2T_s$ in Fig. 4.6), is also shown in the right upper part of the figures. The eye diagram can be easily displayed on an oscilloscope, and is a useful tool for the evaluation of the combined effects of channel noise and intersymbol interference on the transmitted or received signals [4.9].

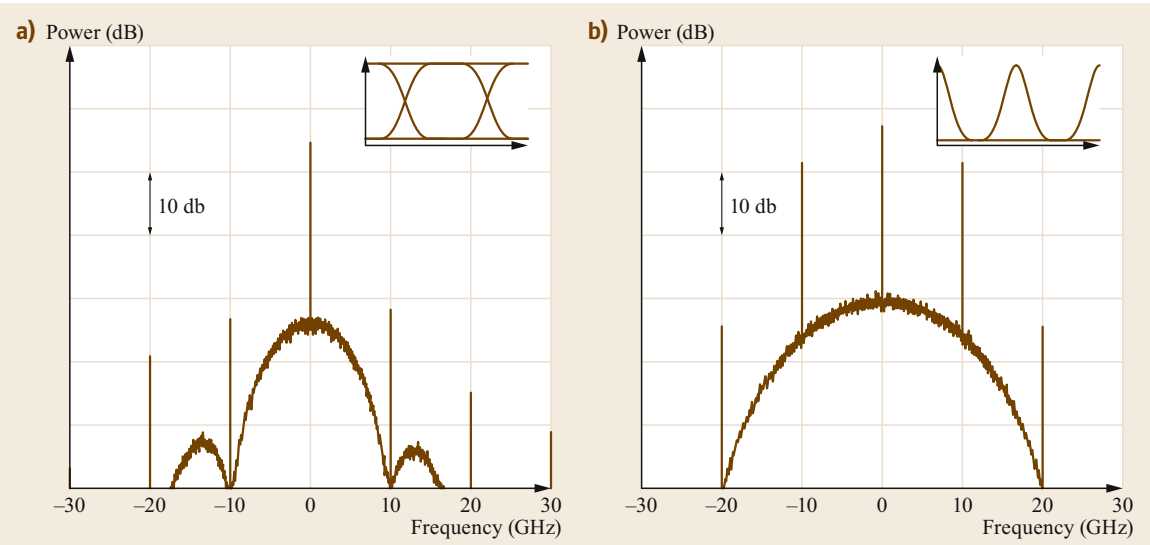


Fig. 4.6a,b Examples of NRZ (a) and RZ (b) optical spectra at 10 GBd. The corresponding eye diagrams are shown in the *insets*

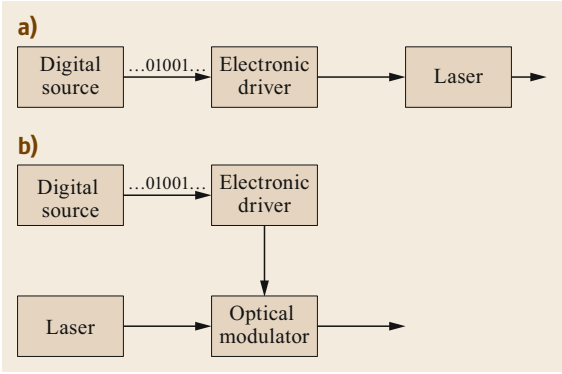


Fig. 4.7a,b Scheme of an OOK transmitter using (a) direct modulation and (b) external modulation

4.2.2 On–Off Keying Optical Transmitter

At the transmitter, there are two alternative ways to convert electrical data into optical form, known as *direct modulation* and *external modulation* [4.57]. In the case of direct modulation, the output power of the light source depends directly on the input drive current (or voltage), e.g., light is emitted from the device when a binary one is being transmitted and no light is emitted when a binary zero is being transmitted. The schematic of an OOK transmitter employing direct modulation is shown in Fig. 4.7a. In the case of external modulation, the light source is always on and an external modulator, driven by the electrical data signal, acts as a switch in order to control the intensity/phase of the generated optical signal. The schematic of an OOK transmitter employing external modulation is shown in Fig. 4.7b.

Since transmitters employing directly modulated lasers (DMLs) do not require the use of external modulators, they are simpler and cheaper, but their transmission performance is strongly limited by the *chirp* phenomenon [4.58, 59] (a residual phase modulation superimposed on the desired intensity modulation). Interacting with chromatic dispersion, chirp may introduce a relevant performance penalty, limiting the application of direct modulation to short-reach transponders [4.60–62]. In this chapter, we focus on external modulation by chirp-free Mach–Zehnder modulators (MZMs).

NRZ-OOK Transmitter

An NRZ-OOK signal can be generated using the transmitter scheme shown in Fig. 4.8, by properly driving an MZM with an RF signal carrying the information bits.

The electrical field and power transmission of an MZM as a function of the applied voltage V_d are,

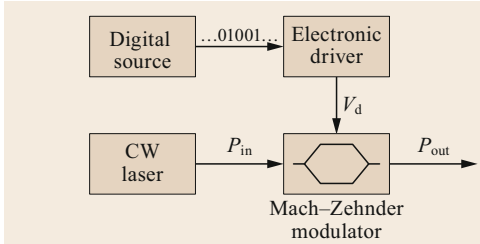


Fig. 4.8 Scheme of an NRZ-OOK transmitter

respectively

$$\frac{E_{out}}{E_{in}} = \sin\left(\frac{\pi}{2}\frac{V_d - V_0}{V_\pi}\right) - j\frac{1}{ER}\cos\left(\frac{\pi}{2}\frac{V_d - V_0}{V_\pi}\right), \quad (4.3a)$$

$$\frac{P_{out}}{P_{in}} = \sin^2\left(\frac{\pi}{2}\frac{V_d - V_0}{V_\pi}\right) + \frac{1}{ER}\cos^2\left(\frac{\pi}{2}\frac{V_d - V_0}{V_\pi}\right), \quad (4.3b)$$

where V_0 is the input voltage corresponding to minimum transmission, V_π is the input voltage difference between maximum and minimum transmission, and ER is the extinction ratio of the MZM, defined as the ratio between the maximum and the minimum output power. A noninfinite value of ER means that the minimum transmission is not zero and a performance penalty can be experienced, as shown in Sect. 4.2.4.

The ideal power transmission of an MZM is shown in Fig. 4.9, where $V_0 = 0$ and $ER = \infty$. The circles indicate the 50% transmission points, also referred to as *quadrature points*. In order to generate OOK pulses,

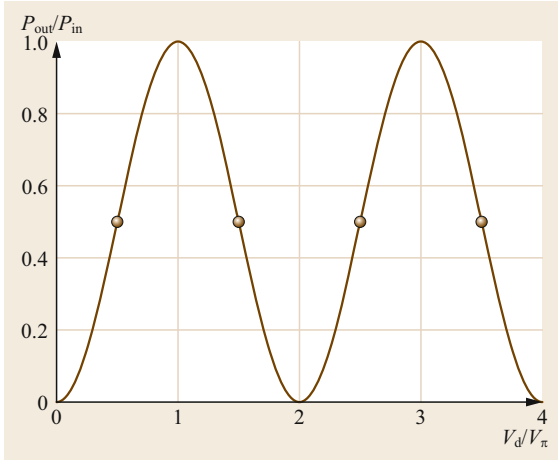


Fig. 4.9 Ideal MZM power transmission. The *circles* indicate the quadrature points

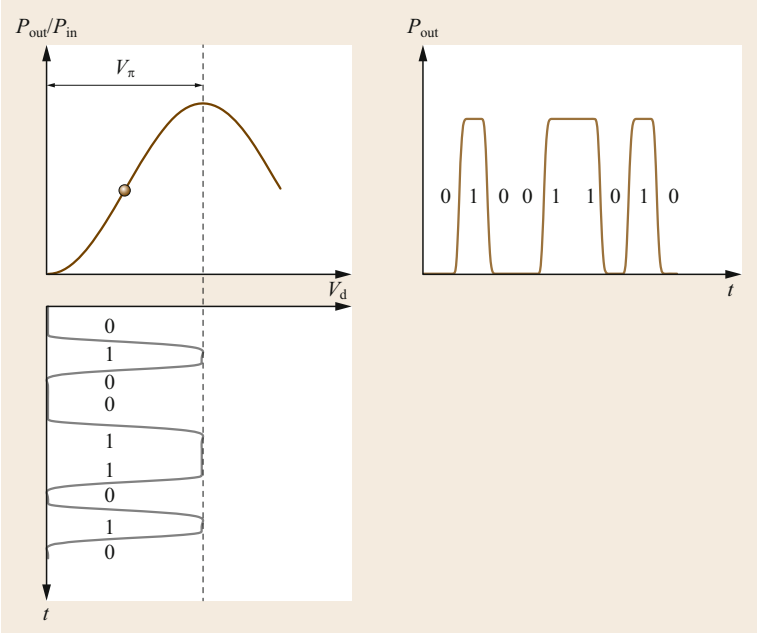


Fig. 4.10 MZM driven by an NRZ drive signal

the MZM needs to be biased at a quadrature point and driven with an input voltage ranging from minimum to maximum transmission, as schematically shown in Fig. 4.10.

RZ Transmitter

RZ transmitters can be implemented either by electronically generating an RZ drive waveform (using the architecture shown in Fig. 4.8) or by carving out pulses from an NRZ signal using an additional modulator, called a pulse carver, as shown in Fig. 4.11. A pulse carver is usually implemented by driving an MZM with a sinusoidal signal. The width of the RZ pulses can be varied by tuning the sinusoidal signal amplitude and frequency. If the sinusoidal signal has a frequency equal to the bit rate and drives the modulator between minimum and maximum transmission, optical pulses with full-width half-maximum of 50% of the bit duration (i.e., with a duty cycle of 50%) are obtained. The pulse width can be reduced by driving the modulator between its transmission minima with a sinusoidal signal whose frequency is equal to half the bit rate. In this case, a duty cycle of 33% can be obtained, as shown in Fig. 4.12.

CSRZ Transmitter

A modification of the RZ line coding, with an increased tolerance to nonlinear propagation effects, was proposed in [4.63]. As for standard OOK formats, the information is encoded on the intensity levels $\{0, 1\}$, but the phase is changed by π every bit, regardless of the data information. Thus, on average, the optical field of half the 1-bits has a positive sign, while the other half has a negative sign, resulting in a zero-mean optical field envelope. As a consequence, the carrier at the optical center frequency vanishes. For this reason, the format is called carrier-suppressed RZ (CSRZ). CSRZ

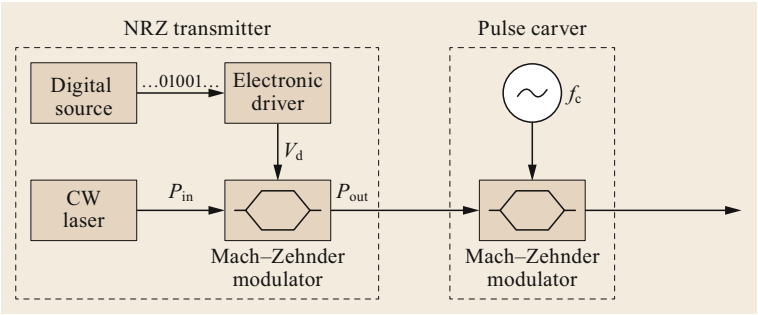


Fig. 4.11 Scheme of an RZ-OOK transmitter

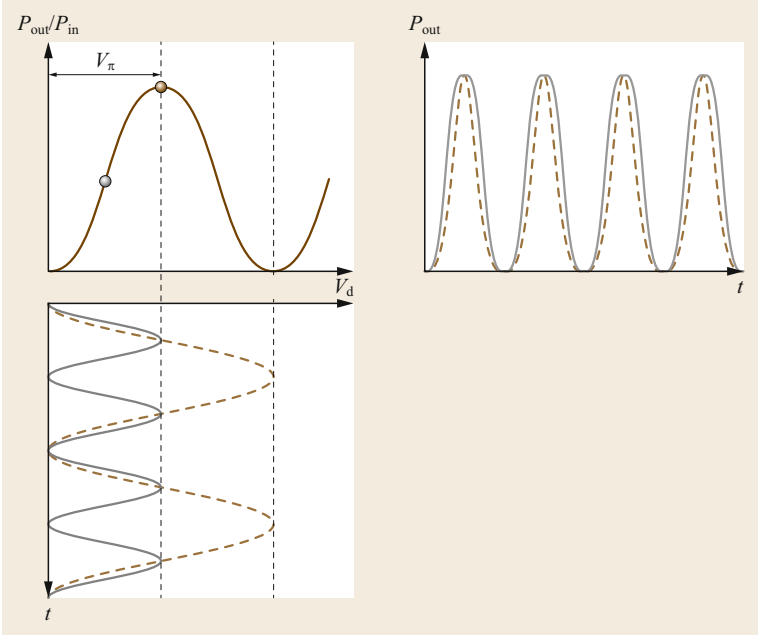


Fig. 4.12 MZM operation as a pulse carver to generate RZ pulses with 33% duty cycle (*dashed lines*) and 50% duty cycle (*solid lines*)

signals can be generated using the same transmitter used for RZ (Fig. 4.11), sinusoidally driving the MZM in the pulse carver at half the data rate between its transmission maxima, as shown in Fig. 4.13. The result is the generation of pulses with 67% duty cycle and alternating phase. Note that the larger duty cycle leads to a lower optical filter penalty (due to either ISI or XTK) with respect to standard RZ pulses.

Chirped Return-To-Zero (CRZ) Transmitter

The chirped-RZ modulation format is obtained by adding a controlled amount of phase modulation to RZ pulses. This can be achieved by inserting a bit-synchronous phase modulator, with an adjustable peak-to-peak level and phase relative to the center of the bit [4.64], after a standard RZ transmitter (just like the one shown in Fig. 4.11). The generated optical phase term is equal to $e^{-jm\cos(2\pi f_0 t)}$, where m is the *phase modulation index* and f_0 is the sinusoidal modulation frequency, which is equal to the data rate. The applied phase modulation reduces the nonlinear propagation distortions [4.65], but it also increases the signal spectrum bandwidth, and consequently the cross-talk between adjacent channels in a WDM system. The optimum amount of phase modulation for CRZ is thus a compromise between tolerance to linear and nonlinear cross-talk. CRZ is typically used in ultralong-haul point-to-point fiber communications, such as submarine systems, with a phase modulation index of approximately 1 rad [4.66].

Alternatively to the three-stage modulation scheme (NRZ modulator, pulse carver, phase modulator), a dual-drive MZM can be used as a chirped pulse carver at various duty cycles by controlling the imbalance of the drive signal amplitudes or their relative phase, as shown in [4.67].

4.2.3 Direct-Detection Optical Receiver

The general scheme of an optically preamplified direct-detection receiver is shown in Fig. 4.14. It is composed of an optical amplifier, followed by an optical filter with impulse response $h_o(t)$, a photodetector, an electrical amplifier, an electrical filter with impulse response $h_e(t)$, and a threshold decision device.

The received optical field at the output of the optical amplifier can be written as

$$y(t) = s_R(t) + n(t)\,, \tag{4.4}$$

where $n(t)$ is the ASE noise introduced by the optical amplifier and $s_R(t)$ is the useful signal component, which is given by

$$s_R(t) = \sum_k a_k q_R(t - kT_s)\,, \tag{4.5}$$

where a_k are the information symbols and $q_R(t)$ is the received optical pulse after propagation in the optical fiber.

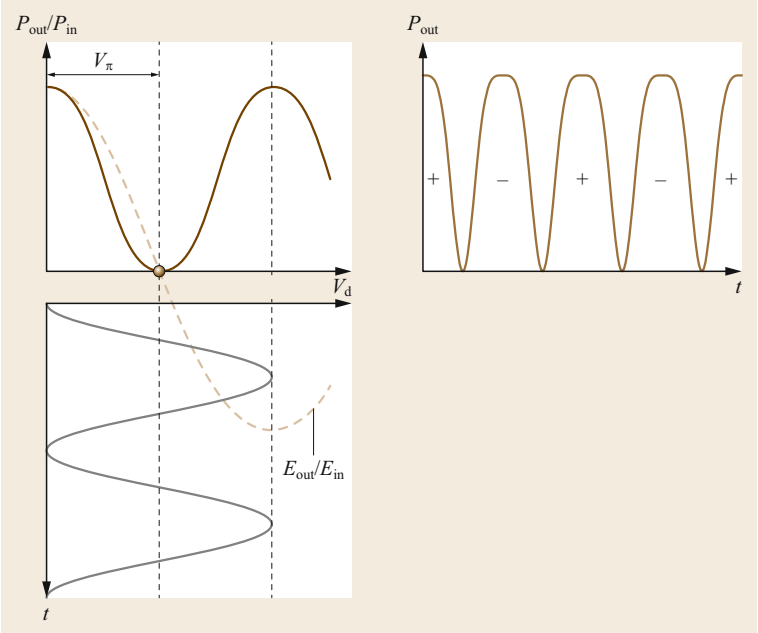


Fig. 4.13 MZM operation as a pulse carver to generate CSRZ pulses with 67% duty cycle

ASE noise is modeled as a white Gaussian noise with spectral density $N_0/2$ [4.57], where

$$N_0 = n_{sp} h\nu (G-1)\,, \tag{4.6}$$

in which n_{sp} is the spontaneous emission factor, h is Plank's constant, ν is the transmission frequency, and G is the amplifier gain. The noise figure F of the optical amplifier, defined as the ratio between the input and output signal-to-noise ratio, is approximately equal to $2n_{sp}$.

The signal at the output of the optical filter can be written as

$$r(t) = s(t) + m(t)\,, \tag{4.7}$$

with

$$\begin{aligned} s(t) &= s_R(t) * h_o(t)\,, \\ m(t) &= n(t) * h_o(t)\,, \end{aligned} \tag{4.8}$$

where the star denotes a convolution and $m(t)$ is a colored Gaussian noise with power spectral density given by

$$G_m(f) = \frac{N_0}{2}|H_o(f)|^2 \tag{4.9}$$

and variance given by

$$\sigma_m^2 = \frac{N_0}{2}\int_{-\infty}^{+\infty}|H_o(f)|^2 \mathrm{d}f = \frac{N_0}{2}\int_{-\infty}^{+\infty}|h_o(t)|^2 \mathrm{d}t\,. \tag{4.10}$$

The detected current is proportional to the incident optical power $P_R(t)$, which is in turn proportional to the squared magnitude of the optical field as shown in (4.7)

$$\begin{aligned} i_R(t) \propto P_R(t) &= |r(t)|^2 = |s(t)+m(t)|^2 \\ &= |s(t)|^2 + |m(t)|^2 + 2\mathrm{Re}\{s(t)m^*(t)\}\,. \end{aligned} \tag{4.11}$$

The first term is the useful signal, the second term is the (noise × noise) beat term, and the third term is the

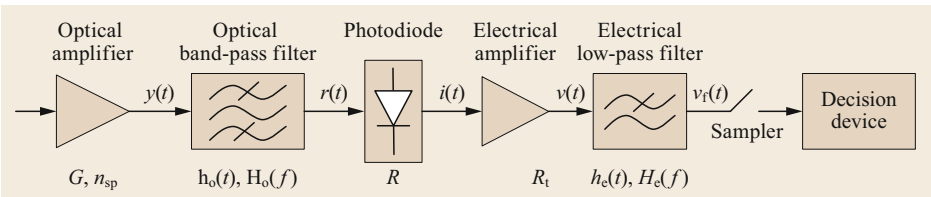


Fig. 4.14 Schematic of an optically preamplified direct-detection receiver

(signal × noise) beat term. The complete photodetection current, including the noise signals generated by the electrical amplifier and the photodetector, can be written as

$$i(t) = i_s(t) + i_{s,ASE}(t) + i_{ASE,ASE}(t) + i_{th}(t) + i_{sh}(t) \,. \tag{4.12}$$

The terms on the right hand side represent the useful signal, the (signal × noise) and (noise × noise) beat terms, the thermal noise and the shot noise respectively [4.57].

The signal after the electrical filter can be written as

$$\begin{aligned} v(t) &= R_t i(t) * h_e(t) \\ &= v_s(t) + v_{s,ASE}(t) + v_{ASE,ASE}(t) \\ &\quad + v_{th}(t) + v_{sh}(t) \,, \end{aligned} \tag{4.13}$$

where R_t accounts for the photodiode and amplifier resistances and $h_e(t)$ is the impulse response of the electrical filter. The signal $v(t)$ is then sampled at the sampling instants $t_k = t_0 + kT$ and compared to a threshold v_{th}. The decision algorithm is then

$$a_k = \begin{cases} 0 & \text{if } v(t_k) < v_{th} \,, \\ 1 & \text{if } v(t_k) > v_{th} \,. \end{cases} \tag{4.14}$$

In general, the system performance is limited by ASE noise, shot noise, and thermal noise. However, in preamplified multispan DWDM systems as addressed in this chapter, shot noise and thermal noise are typically negligible compared to ASE noise. These systems are said to be *ASE limited* and their performance can be characterized by the optical signal-to-noise ratio (OSNR), which is defined as

$$\mathrm{OSNR} = \frac{P_s}{P_N} = \frac{P_s}{2N_0 B_N} \,, \tag{4.15}$$

where P_s is the average power of the optical signal at the input of the receiver optical filter, P_N is the ASE noise power on both polarizations measured over an optical reference bandwidth equal to B_N, and N_0 is the corresponding single-sided ASE noise power spectral density, as shown in (4.6). Often a fixed reference bandwidth B_N of 12.5 GHz is used (corresponding to 0.1-nm wavelength range), which is a standard setting in many optical spectrum analyzers. Alternatively, the symbol rate of the signal R_S can serve as reference bandwidth. In this chapter, we follow the latter approach.

In the 1.55-μm window and with sufficiently high optical amplifier gain, the OSNR (referenced to the symbol rate) can then be rewritten in logarithmic form as

$$\begin{aligned} \mathrm{OSNR_{dB}} &= 58\,\mathrm{dBm} + 10\log\left(\frac{P_s}{1\,\mathrm{mW}}\right) \\ &\quad - 10\log(\mathrm{F}) - 10\log\left(\frac{B_N}{R_S}\right) , \end{aligned} \tag{4.16}$$

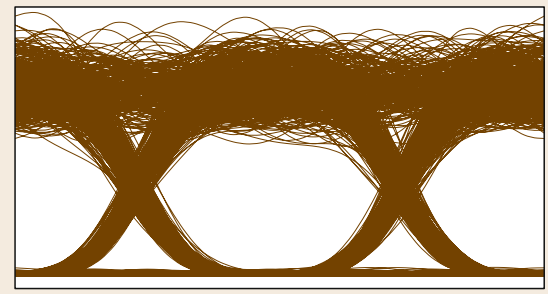


Fig. 4.15 Noisy NRZ-OOK eye diagram after direct detection

where the signal power P_s is given in mW and F is a known amplifier noise figure.

Figure 4.15 shows an example of an eye diagram after direct detection in an NRZ-OOK system limited by ASE noise (where the other noise terms are negligible). Due to the (signal × noise) beat term, the variance of noise is higher on the "1" level and, consequently, the optimum decision threshold v_{th} is closer to the "0" level.

4.2.4 Back-to-Back Performance and Dispersion Tolerance

Ideal Performance of a Binary IMDD System

In ASE-limited systems, for all formats based on intensity modulation and direct detection the optimum receiver is based on the use of an optical filter whose impulse response is matched to the received pulses $q_R(t)$ [4.9]. The optimum achievable performance is independent of the pulse shape, provided that the receiver optical filter is perfectly matched to the electrical field of the input signal. This optimum performance (also known as the *quantum limit* [4.68, 69]) is shown in Fig. 4.16 in terms of bit error rate (BER) versus optical signal-to-noise ratio (OSNR), as defined by (4.16). In Fig. 4.16, the reference noise bandwidth is equal to the symbol rate. This means that the performance curve is valid for any value of the symbol rate (which for binary format is equal to the bit rate).

An analytical derivation of the BER of ASE-noise-limited systems can be found in [4.29]. In the case of a matched optical filter (where $M = 2$ in [4.29])

$$\begin{aligned} \mathrm{BER} &= \frac{1}{2}\Big[\mathrm{e}^{-\varphi}(1+\varphi) + 1 \\ &\quad - Q_2\Big(\sqrt{8\mathrm{OSNR}}, \sqrt{2\varphi}\Big)\Big] \\ \text{with } \varphi &= \frac{v_{th}}{N_0 R_s} \,, \end{aligned} \tag{4.17}$$

where Q_2 is the generalized Marcum Q function of order 2 and φ is the normalized decision threshold that

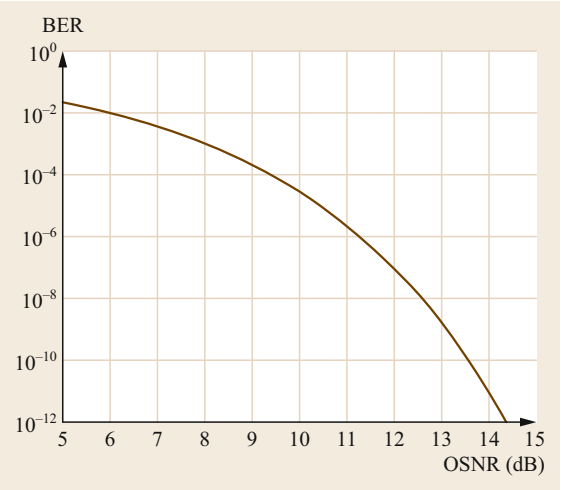


Fig. 4.16 Ideal performance of an IMDD optical system

needs to be optimized in order to minimize the BER. A good approximation of the curve shown in Fig. 4.16 is

$$\mathrm{BER} \cong \frac{1}{2} e^{-0.98\,\mathrm{OSNR}} . \tag{4.18}$$

Optimization of Filter Bandwidths

The ideal performance of IMDD systems is independent of the pulse shape, but only if a perfectly matched optical filter is available, a condition which is difficult to achieve with analog optical filters. This is the reason why a different performance is typically achieved using NRZ and RZ/CSRZ pulses, with the latter showing a better sensitivity in most cases [4.68, 70, 71]. Several research papers have been published on the optimization of the optical and electrical filters in order to achieve system performances as close as is possible to quantum limit [4.69, 72–76]. The final results depend on the filter shapes, but the same general behavior is shown in all papers.

As an example of receiver filter optimization, we report in the following a set of simulation results obtained for NRZ, RZ, and CSRZ transceivers. The transmitter architectures are the ones shown in Figs. 4.8 and 4.11. For NRZ, the bandwidth limitation of the TX is emulated using a 5-pole Bessel low-pass filter with 3 dB bandwidth equal to $0.7R_s$. The FWHM of RZ and CSRZ pulses obtained after the pulse carving operation is 50 and 66%, respectively. At the RX side, the passband optical filter has a 2nd-order super-Gaussian shape with 3 dB bandwidth B_{opt}. An ideal photodetector and a 5-pole Bessel electrical low-pass filter with 3 dB bandwidth B_{el} follow. Since we are considering ASE-noise-limited systems, thermal and shot noise have not been included in the simulations. Noise loading was performed at the receiver input in order to set the OSNR value. The target BER was set to 10^{-3}, which corresponds to the pre-FEC BER value of standard hard-decision EFECs with 7% overhead [4.77]. BER values were evaluated through direct error counting over 2^{18} bit.

The results are shown in Fig. 4.17. The matched filter performance, identical for all systems, is obtained with an OSNR of 8 dB (over a bandwidth equal to R_s). For the NRZ case, the minimum OSNR (with a ≈ 1 dB penalty with respect to the optimum matched filter performance) is achieved for $B_{opt} = 1.2R_s$ and $B_{el} = R_s$. For both the RZ and CSRZ format the penalty with respect to the matched filter performance is lower (≈ 0.5 dB) and the optimum performance is achieved with a larger optical filter bandwidth equal to $B_{opt} = 2R_s$. The reason why the performance of RZ formats is closer to the matched filter limit than NRZ is that there is a greater matching between the RZ pulses and the Super-Gaussian optical filter. The reason why the best performance for NRZ is obtained for a lower B_{opt} value is that NRZ has an intrinsically better bandwidth-efficiency than RZ formats. It's also worth mentioning that RZ/CSRZ contours are less pronounced and therefore low penalty operation can be achieved over a wider (B_{opt},B_{el}) range than for NRZ.

Impact of Finite MZM Extinction Ratio

The extinction ratio ER of a modulator corresponds to the ratio of the power P_1 used to transmit a logic level 1 to the power P_0 used to transmit a logic level 0. Ideally, $P_0 = 0$ and $\mathrm{ER} = \infty$, but in practice $P_0 > 0$ and $\mathrm{ER} < \infty$, because of the presence of spontaneous emission noise and bias errors, as well as the electronic bandwidth limitations in modulator and driver.

A finite value of ER induces a performance penalty, which can be defined as the increase in average power needed to obtain the same bit error rate as an ideal pulse with infinite extinction ratio. In the case of systems limited by additive white Gaussian noise (e.g., IMDD systems limited by thermal noise), the power penalty is given by [4.78]

$$\delta_{\mathrm{ER}} = 10 \log_{10} \left(\frac{\mathrm{ER}+1}{\mathrm{ER}-1} \right) . \tag{4.19}$$

In IMDD systems limited by ASE noise there is an additional penalty due to the increase of ASE noise on the zero level, as shown in [4.79]

$$\delta_{\mathrm{ER}} = 10 \log_{10} \left(\frac{\mathrm{ER}+1}{\mathrm{ER}-1} \right) + 10 \log_{10} \left(\frac{\sqrt{\mathrm{ER}}+1}{\sqrt{\mathrm{ER}}-1} \right) . \tag{4.20}$$

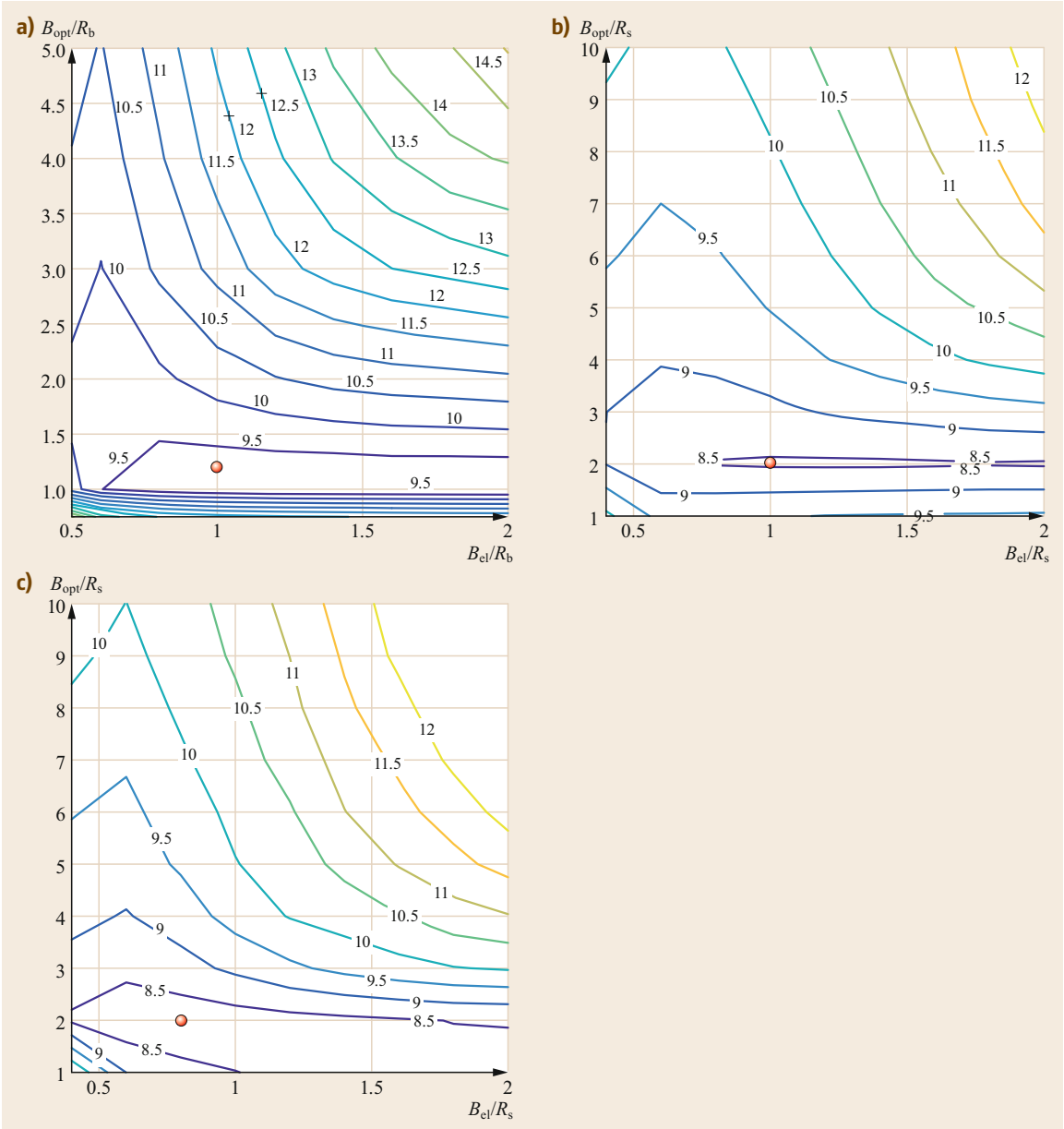


Fig. 4.17a–c Contour plots of the required OSNR to achieve BER $= 10^{-3}$ as a function of the receiver optical and electrical 3-dB bandwidths (B_{opt} and B_{el}) for **(a)** NRZ, **(b,c)** CSRZ pulses. The *red dots* indicate the optimum points (i.e., the points with minimum required OSNR)

Both, (4.19) and (4.20), assume that the eye diagram is perfectly open, i.e., no intersymbol interference (ISI) is present at the sampling instant. Figure 4.18 shows the power penalty as a function of the ER value, for both thermal-noise-limited and ASE-noise-limited receivers. In ASE-noise-limited systems, the power penalty shown in Fig. 4.18 is equivalent to an OSNR penalty. Note that the penalty due to a finite ER might be lower in the presence of ISI and non-perfectly matched receivers, as shown in [4.69, 75]. An approximate estimation of the required OSNR for a certain bit error rate can then be obtained if an effective extinction ratio is calculated from the electrical eye diagram at the receiver and the resulting penalty is added to the reference OSNR at the quantum limit.

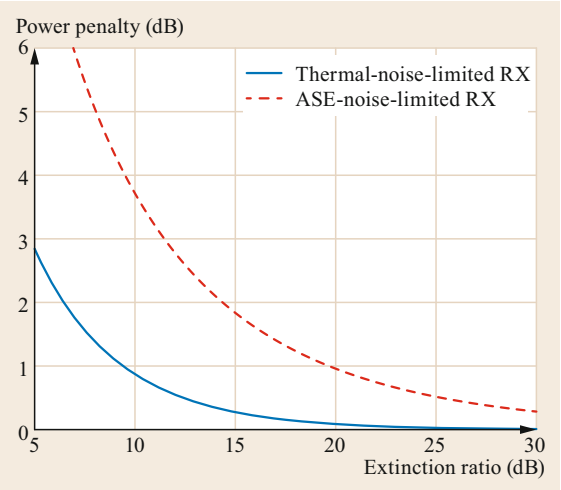


Fig. 4.18 Power penalty due to a finite ER in the absence of ISI

Dispersion Tolerance

Group velocity dispersion (GVD) is the result of the different wavelengths in a light beam arriving at their destination at slightly different times due to the frequency dependence of the refractive index of the fiber. In the literature, the GVD phenomenon is also referred to as chromatic dispersion (CD). Polarization-mode dispersion is a form of modal dispersion where two different polarizations of light in a waveguide travel at different speeds due to random imperfections and asymmetries [4.78]. Both phenomena cause a time spreading of the optical pulses, which in turn induces a performance penalty.

PMD is related to the *differential group delay* (DGD) caused by birefringence in optical fibers. However, birefringence varies along the length of the fiber, with different sections exhibiting a difference birefringence both in terms of levels of refractive index asymmetry and of relative orientation of the slow and fast polarization axis, also known as *principal states of polarization* (PSP). As a result, the instantaneous value of DGD measured at the output of a fiber is a random variable that is well approximated by a Maxwellian distribution [4.80]. The PMD of a fiber is typically described through the parameter δ_{PMD}, which is an estimate of the average value of the statistical distribution of DGD.

The tolerance to GVD and PMD is strongly dependent on the symbol rate of the modulated signal. In particular, the pulse broadening $\Delta\tau_{\mathrm{CD}}$ due to GVD and the average value of DGD between the two principal states of polarizations (normalized to the symbol time $T_{\mathrm{s}} = 1/R_{\mathrm{s}}$) are given by [4.81]

$$\frac{\Delta\tau_{\mathrm{CD}}}{T_{\mathrm{s}}} = 2\pi|\beta_2|kR_{\mathrm{s}}^2 L \times 10^{-6}\,,$$
$$\frac{\langle\Delta\tau_{\mathrm{DGD}}\rangle}{T_{\mathrm{s}}} = \delta_{\mathrm{PMD}}\sqrt{L}R_{\mathrm{s}} \times 10^{-3}\,, \qquad (4.21)$$

where β_2 is the CD coefficient in $\mathrm{ps}^2/\mathrm{km}$, δ_{PMD} is the PMD parameter that can be found in fiber data sheets and is expressed in $\mathrm{ps}/\sqrt{\mathrm{km}}$, R_{s} is the symbol rate in GBd (which for binary formats is identical to the bit rate R_{b}), L is the fiber length in km, and k is a coefficient that depends on the modulation formats, defined as the ratio between the spectral width of the modulated signal and the symbol rate. The coefficient β_2 is related to the dispersion parameter D ($(\mathrm{ps/nm})/\mathrm{km}$) through the relationship $D = -2\pi c\beta_2/\lambda^2$, where c is the speed of light in vacuum and λ is the propagation wavelength.

Equation (4.21) shows that the effect of DGD scales linearly with R_{s}, while the impact of GVD scales quadratically with R_{s}, for example a doubling in the symbol rate reduces the GVD tolerance by approximately a factor of 4. Also, the normalized pulse broadening due to GVD scales linearly with the normalized spectral width of the modulation format, resulting in a higher GVD tolerance (in a linear propagation regime) for NRZ modulation formats. The tolerance to DGD is instead in general higher for RZ modulation formats [4.82].

As an example, Fig. 4.19 shows the performance of NRZ, RZ, and CSRZ modulation formats in the presence of either GVD or a first-order PMD. The simulation setup is the same as described in the first part of this section, with optimized RX filter bandwidth (corresponding to the red dots in Fig. 4.17). Linear fiber propagation was assumed, while first-order PMD was emulated by inserting a deterministic DGD value $\Delta\tau_{\mathrm{DGD}}$ between the two principal states of polarization (PSP) [4.80]. The worst-case of equal power-splitting ratio between the two PSPs was considered [4.83].

In order to make the results symbol-rate independent, the x-axes have been normalized with respect to either R_{s}^2 (in the GVD plot) or R_{s} (in the PMD plot). The value of accumulated dispersion at a given symbol rate ($D\,L$) in ps/nm can be obtained by multiplying the value read on the x-axis of Fig. 4.19a by the coefficient $(2\pi c)/(R_{\mathrm{s}}^2\lambda^2) \times 10^6$, with c expressed in m/s, R_{s} in GBd, and λ in nm. In the 1.55-μm window, this corresponds to a multiplication by a factor $\approx 0.784 \times 10^6/R_{\mathrm{s}}^2$, with R_{s} measured in GBd. As an example, at 10.7 Gb/s the GVD tolerance for a 2-dB penalty is equal to ≈ 550 ps/nm for RZ and to ≈ 750 ps/nm for NRZ and CSRZ. The DGD tolerance for the 2-dB penalty at

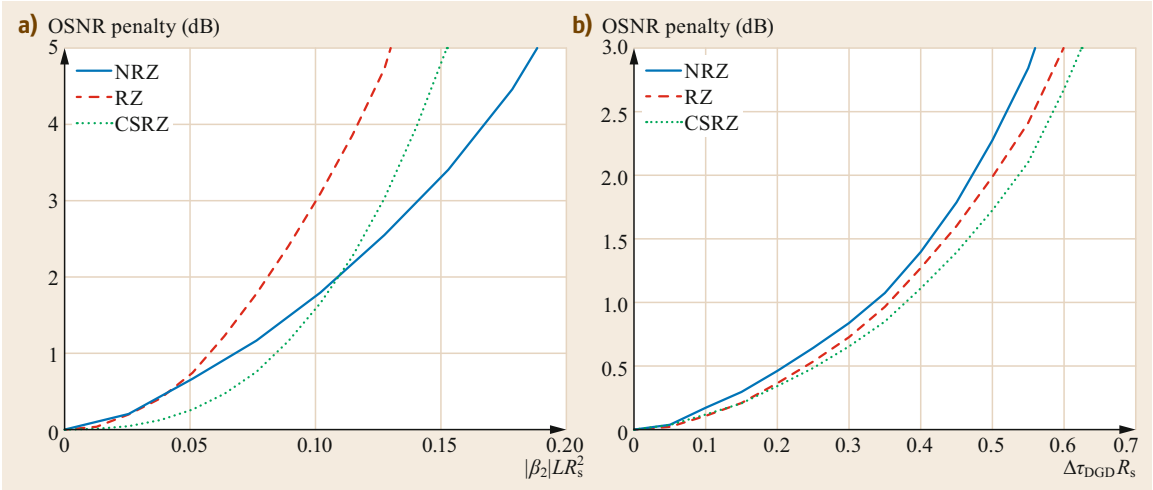


Fig. 4.19a,b GVD and first-order PMD tolerance of NRZ, RZ, and CSRZ in terms of OSNR penalty with respect to back-to-back performance as a function of either **(a)** the normalized accumulated dispersion or **(b)** the normalized DGD in the worst-case power-splitting ratio between the two PSPs. The target BER is 10^{-3} and the value of the reference back-to-back OSNR (over a bandwidth equal to R_s) is 9.0, 8.5, and 8.4 dB for NRZ, RZ, and CSRZ, respectively

10.7 Gb/s is instead equal to 44, 46.5, and 50 ps/nm for NRZ, RZ and CSRZ, respectively. Note that, considering the statistical nature of DGD and allowing a limited outage time, the tolerated PMD (based on the δ_{PMD} parameter) is typically smaller than the tolerable DGD by a factor of ≈ 3 [4.84].

GVD and DGD tolerance, like back-to-back OSNR requirements, can depend in a non-negligible way on pulse and filter shapes used in the system, as well as on other residual distortions [4.35, 85–88]. In general, the impact on a modulation format of various impairments cannot be taken in isolation, but has to be evaluated in the context of the system it is operating in, which might include dispersion maps for nonlinearity mitigation [4.89, 90] and/or advanced digital signal-processing techniques [4.45] that can for instance significantly increase the dispersion tolerance.

4.2.5 Multilevel Intensity Modulation

An M-ary PAM optical signal can be obtained by applying an M-ary electrical signal to an MZM biased at the quadrature point, i.e., half point of its transfer characteristic (Figs. 4.8 and 4.10). The signal bandwidth of an M-ary signal is scaled by a factor $1/\log_2(M)$ compared to a binary signal operating at the same bit rate R_b, where $R_b = R_s \log_2(M)$ and R_s is the symbol rate.

For any optical extinction ratio, it is easy to show that the M levels should be equally spaced (for example $\{0, 1/3, 2/3, 1\}$ for PAM-4) when the noise is stationary, i.e., not signal dependent. This is the case, for example, when receiver thermal noise dominates. Given that the noise in an optically preamplified receiver is strongly signal dependent, it is clear that the lowest probability of error will be achieved when unequal level spacing is used [4.15]. In particular, the photocurrent of each level is approximately Gaussian distributed, with a variance proportional to the intensity. In order to equalize the error probabilities at the M-1 different thresholds, the intensity levels should form a quadratic series, i.e., 0, 1, 4, 9, etc. [4.91]. Under these assumptions, and assuming Gray coding, it can be shown that

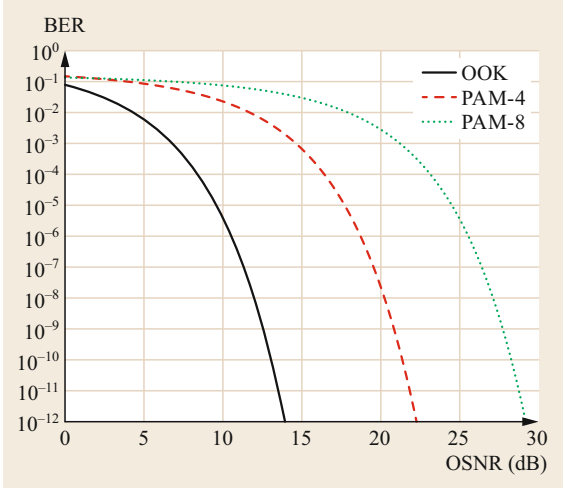


Fig. 4.20 Ideal performance of PAM signals with direct detection in ASE-noise-limited systems (with single-polarization filtering at the RX)

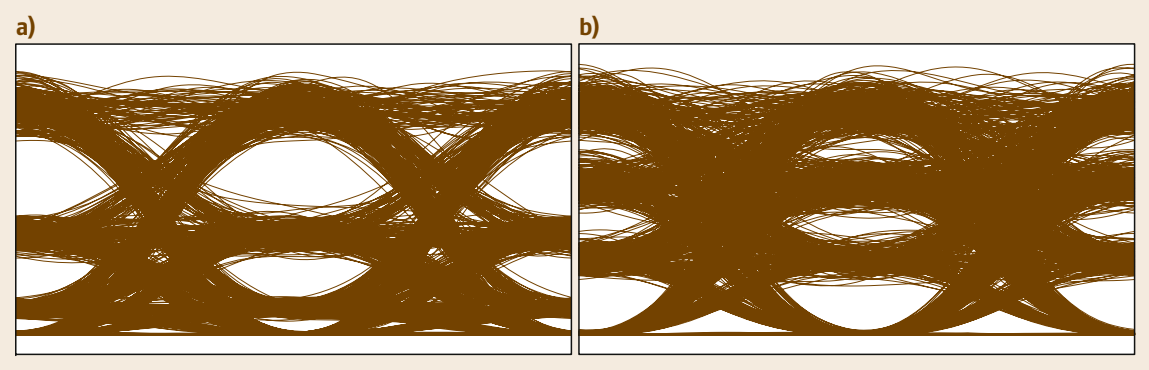


Fig. 4.21a,b Noisy eye diagrams after direct detection of PAM-4 signals with **(a)** quadratic and **(b)** linear level spacing

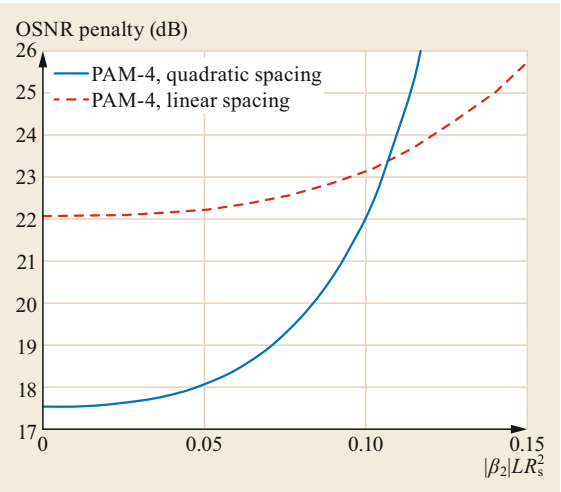


Fig. 4.22 GVD tolerance of PAM-4 in terms of OSNR required to obtain BER $= 10^{-3}$ as a function of the normalized accumulated dispersion

the BER in the optimum matched filter configuration is approximately given by [4.92]

$$\mathrm{BER} \cong \frac{1}{\log_2(M)} \frac{1}{2} \mathrm{erfc}\left(\sqrt{\frac{3\mathrm{OSNR}}{(2M-1)(M-1)}}\right), \tag{4.22}$$

where OSNR is the optical signal-to-noise ratio at the input of the DD receiver, as defined as in (4.15), with $B_\mathrm{N} = R_\mathrm{s}$. Equation (4.22) assumes single polarization filtering at the RX side, i.e., a polarization tracking receiver with a polarization analyzer that blocks the ASE polarized orthogonal to the signal [4.92]. If no polarization filtering is used, the performance is slightly degraded (a fraction of dB of OSNR penalty is typically experienced). Figure 4.20 shows the ideal performance for PAM-2, PAM-4 and PAM-8 constellations, obtained using (4.22). Note that these results hold only at the zero dispersion point. If there is substantial residual dispersion, then a more equidistant level spacing is required. In general, best performance is obtained if the level spacing is optimized. Figure 4.21 shows an example of the back-to-back noisy eye diagrams after the post detection electrical filter with either quadratic or linear level spacing.

Figure 4.22 shows the dispersion tolerance of PAM-4 with either equally spaced or quadratically spaced levels. 5-pole electrical Bessel low-pass filters with 3-dB bandwidth equal to R_s are used both at the TX and at the RX side, and a 2nd-order super-Gaussian passband optical filter with 3-dB bandwidth $2.5R_\mathrm{s}$ is present at the RX input. As expected, while the OSNR sensitivity in back-to-back is significantly better for the quadratic-spacing, the linear-spacing configuration has a higher tolerance to GVD.

4.3 Optical Duobinary with Direct Detection

This section reviews the transmitter and receiver architectures required to generate and detect different types of optical duobinary (ODB) signals. In Sect. 4.3.1, the duobinary modulation concept is introduced. Section 4.3.2 describes the practical implementation of a DB transceiver in an optical system, while Sect. 4.3.3 discusses its back-to-back performance and dispersion tolerance.

4.3.1 Duobinary Line Coding

Figure 4.23 shows the canonical duobinary transmitter structure that can be found in early papers and textbooks [4.9]. It is composed of a precoder, followed by pulse shaping and a duobinary encoder.

The precoder transforms the information bit sequence $a_n \in \{0,1\}$ into a new bit sequence $p_n \in \{0,1\}$ by performing a logical negative exclusive or (XOR) operation between the sequences a_n and p_{n-1}. Then the normalization $b_n = 2p_n - 1$ is applied in order to obtain a bipolar sequence $b_n \in \{-1,+1\}$ [4.17]. Note that the presence of the precoder at the transmitter side allows for easier recovery of the data stream on a bit-by-bit basis and avoids error propagation at the receiver.

The modulated signal after pulse shaping is a standard bipolar ASK signal which can be written as

$$s(t) = \sum_n b_n q_T(t - nT_s)\,, \tag{4.23}$$

where $b_k \in \{-1,+1\}$ are the binary symbols carrying the information, T_s is the symbol time, and $q_T(t)$ is the pulse shape in the time domain. The duobinary coded signal is obtained by adding the data delayed by one symbol period to the present data [4.9, 17], therefore

$$\begin{aligned} s_T(t) &= s(t) + s(t - T_s) \\ &= \sum_n b_n q_T(t - nT_s) + \sum_n b_n q_T(t - nT_s - T_s) \\ &= \sum_n b_n q_{T,DB}(t - nT_s)\,, \end{aligned} \tag{4.24}$$

with $q_{T,DB}(t) = q_T(t) + q_T(t - T_s)$. Note that the signal $s_T(t)$ can be also expressed as

$$s_T(t) = \sum_n 2c_n q_T(t - nT_s)\,, \tag{4.25}$$

with $c_n = b_n + b_{n-1}$. The obtained symbols can thus assume three possible levels ($c_n \in \{-1,0,+1\}$), corresponding to the constellation diagram shown in Fig. 4.24.

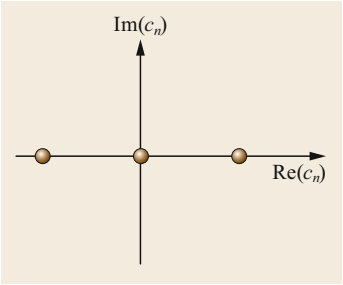


Fig. 4.24 The DB constellation diagram

At the receiver side, the received signal is filtered by a filter with impulse response $h_R(t)$, generating the signal

$$\begin{aligned} s_R(t) &= \sum_n b_n q_{R,DB}(t - nT_s) \\ &= \sum_n 2c_n q_R(t - nT_s)\,, \end{aligned} \tag{4.26}$$

with $q_{R,DB}(t) = q_{T,DB}(t) * h_R(t) = q_R(t) + q_R(t - T_s)$ and $q_R(t) = q(t) * h_R(t)$. The correlated three-level signal is finally demodulated into a binary signal by performing a *modulus* operation which converts -1 and $+1$ to a 1 bit and "0" to a 0 bit. If the precoder shown in Fig. 4.23 is used, $|c_n| = a_n$ and the transmitted sequence is recovered without the need of any further processing. Note that the *modulus* operation is automatically performed by a DD receiver, where the input signal is squared by the photodetector. Figure 4.25 shows an example of the noisy three-level eye diagram before detection and the two-level eye diagram after quadratic detection.

If the pulse $q_R(t)$ satisfies the Nyquist criterion for the absence of ISI, i.e.,

$$q_R(0) \neq 0, q_R(nT_b) = 0, \quad \forall n \neq 0\,,$$

then the duobinary pulse $q_{R,DB}(t)$ has *two* identical nonzero samples T seconds apart [4.9], where

$$\begin{aligned} q_{R,DB}(0) = q_{R,DB}(T_s) \neq 0\,, \\ q_{R,DB}(nT_s) = 0\,, \quad \forall n \neq 0, 1\,. \end{aligned} \tag{4.27}$$

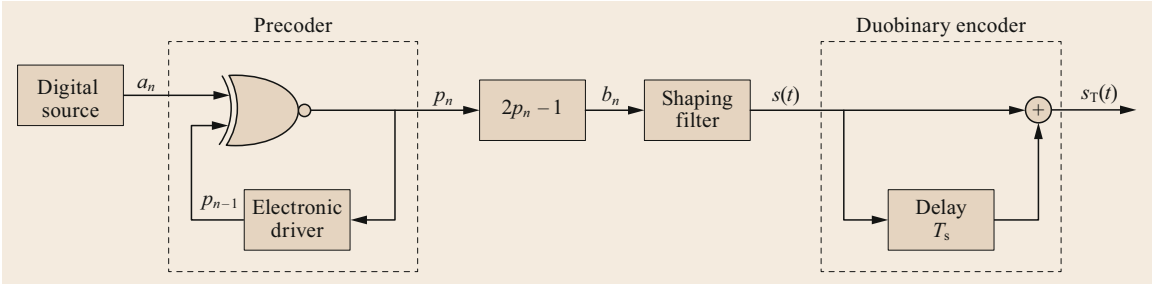


Fig. 4.23 Schematics of a duobinary transmitter

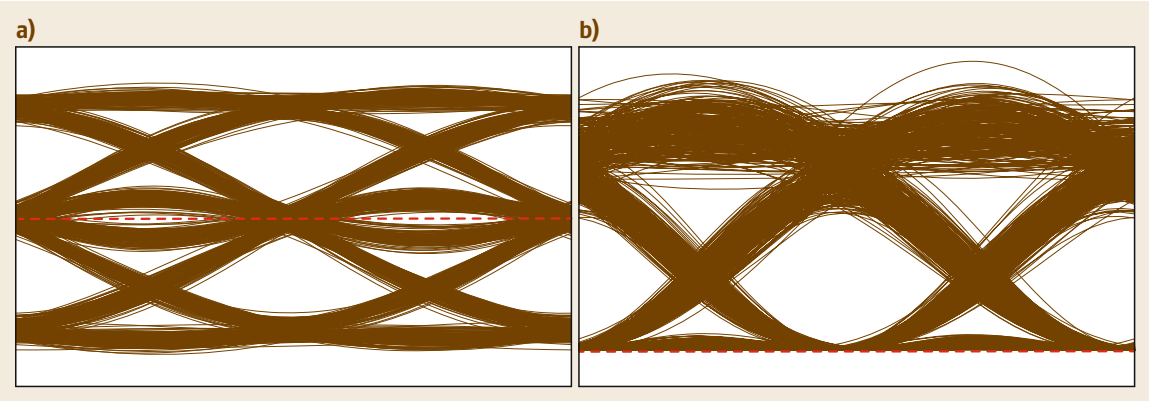


Fig. 4.25a,b The eye diagram of the noisy DB signal before (**a**) and after DD (**b**). The *red dashed lines* indicate the zero power level

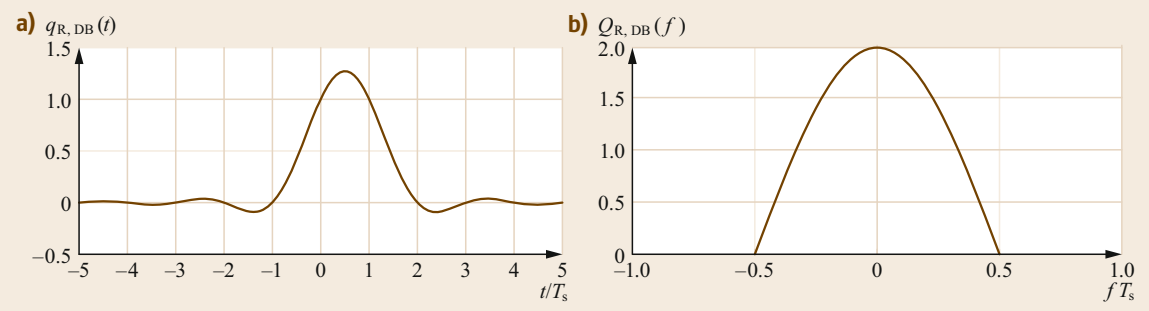


Fig. 4.26 (**a**) The time domain and (**b**) frequency domain characteristics of the minimum bandwidth DB pulse

In practice, DB pulses are defined as pulses that spread over two bit intervals, i.e., pulses that correlate adjacent bits through the introduction of a controlled amount of ISI [4.9].

The DB encoder is equivalent to a filter with transfer function

$$H_{DB}(f) = 1 + e^{-j2\pi f T_s} = 2e^{-j\pi f T_s} \cos(\pi f T_s) \,. \quad (4.28)$$

If an ideal low-pass filter is used for pulse shaping, the minimum bandwidth duobinary signal is obtained [4.9], whose Fourier transform is equal to

$$Q_{R,DB}(f) = \begin{cases} 2e^{-j\pi f T_b} \cos(\pi f T_s) & \text{if } |f| < \frac{1}{2} T_s \\ 0 & \text{if } |f| > \frac{1}{2} T_s \,. \end{cases} \quad (4.29)$$

The time domain and frequency domain characteristics of the minimum bandwidth DB pulse are shown in Fig. 4.26. Note that the spectrum decays to zero smoothly, which means that physically realizable filters can be designed that approximate this spectrum very closely.

In [4.93] a rigorous analysis of the ASE-noise-limited back-to-back sensitivity performance of DB modulation was presented, showing that the quantum limit of DB modulation with DD is 0.91 dB better than that of IMDD (Fig. 4.27). However, as for IMDD, the quantum limit can be reached only using a perfectly matched filter at the receiver (before quadratic detection) [4.94], a condition which is hard to achieve with analog optical filters. In fact, the actual performance of optical systems employing the DB modulation depends on the particular method that is used to perform DB encoding and on the degree of optimization of the system components (for example optical and electrical filters), as shown in the following section.

4.3.2 Optical Duobinary (ODB) Transmitter

The first implementations of optical duobinary (ODB) [4.95] were based on the generation of a three-level electrical signal by tight filtering of a standard electrical binary NRZ signal using a half data-rate raised-cosine filter with roll-off equal to 1. The three-level signal was then applied to a standard MZM to generate a three-level intensity-modulated optical

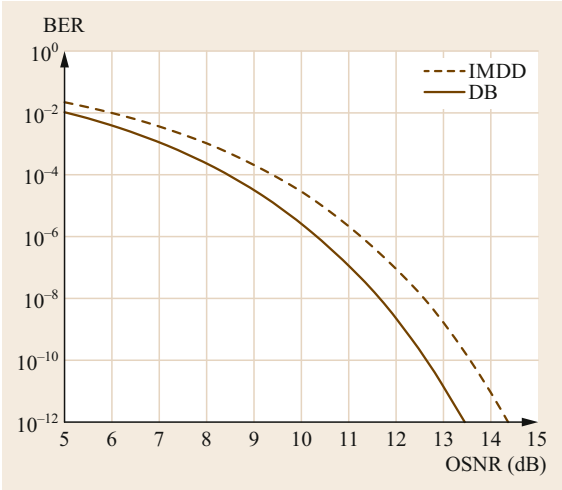


Fig. 4.27 Quantum limit of DB and binary signals with quadratic detection in ASE-noise-limited systems. The quantum limit of binary IMDD is also shown for comparison

signal. A conventional preamplified receiver was used, with an appropriately configured logic circuit to perform a dual threshold detection. The sensitivity penalty of this scheme with respect to binary NRZ was very high (≈ 4 dB at BER $= 10^{-9}$). One of the reasons for this penalty was the use of a pure intensity modulation, allowing transmission and detection of only positive values for the signal levels.

In order to reduce the sensitivity penalty, the use of mixed amplitude and phase modulation was proposed [4.96], which allows the use of a symmetric DB constellation. The transmitter schematic is shown in Fig. 4.28. The ODB signal is obtained by applying a duobinary-encoded three-level electrical signal to a dual-drive MZM driven between its transmission minima (i.e., in a $2V_\pi$ range), as shown in Fig. 4.29. The signal is then decoded using a standard DD receiver. In [4.97] a similar approach was proposed, approximating the delay-and-add operation needed to obtain the DB encoding using a narrow electrical filter. In practice, the effect of this modulation scheme is the generation of an optical signal with two intensity levels (corresponding to the "0" or "1" transmitted bits), but the optical phase of the "1" levels may be equal to either 0 or π, depending on the level ("+1" or "−1") of the three-level driving signal.

These kinds of modulation techniques (based on a binary intensity signal with phase shifts at the higher levels) have subsequently been named phase-shaped binary transmission (PSBT) by *Pennincks* et al. in [4.98], where a similar but simpler approach was proposed, using a conventional low-pass Bessel filter and a single-drive MZM driven between its transmission maxima (Fig. 4.29). The schematic of the PSBT transmitter proposed in [4.21, 98] is shown in Fig. 4.30, where the precoder is followed by a 5-pole Bessel electrical low-pass filter with 3-dB bandwidth $B_{\mathrm{TX,el}} \approx 0.25R_\mathrm{b}$, which is a good approximation of the first arch of an ideal duobinary filter (a cosine-shaped filter). The modulator can be optionally followed by an optical transmitter filter for further pulse shaping [4.23]. Note that Fig. 4.30 shows a transmitter setup with single-ended driver configuration and single ODB filter. In practice, an implementation is sometimes preferred in which both arms of the MZM are driven differentially instead. While such a set-up requires two ODB filters (one for each MZM arm), each driver only needs to drive a swing of V_π instead of $2V_\pi$ which can lead to lower power dissipation and cost.

Figure 4.31 shows an example of the power spectrum of a PSBT signal, compared with a standard

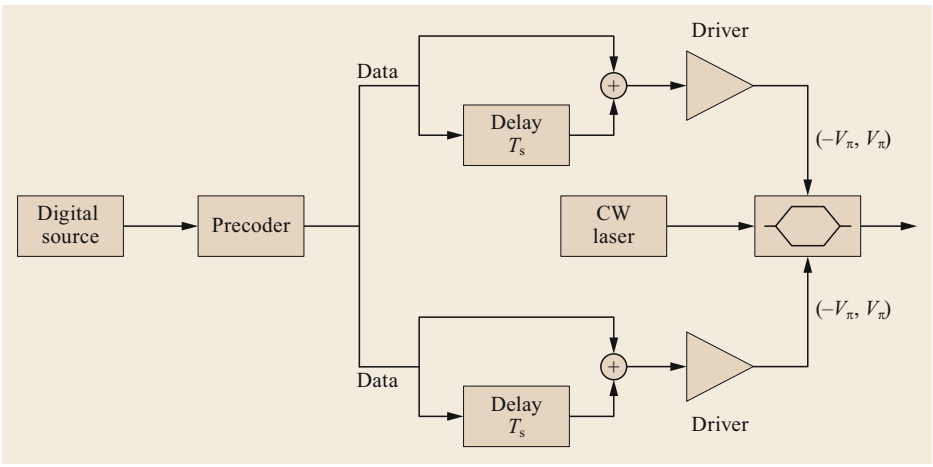


Fig. 4.28 Schematic of the AM/PM ODB transmitter

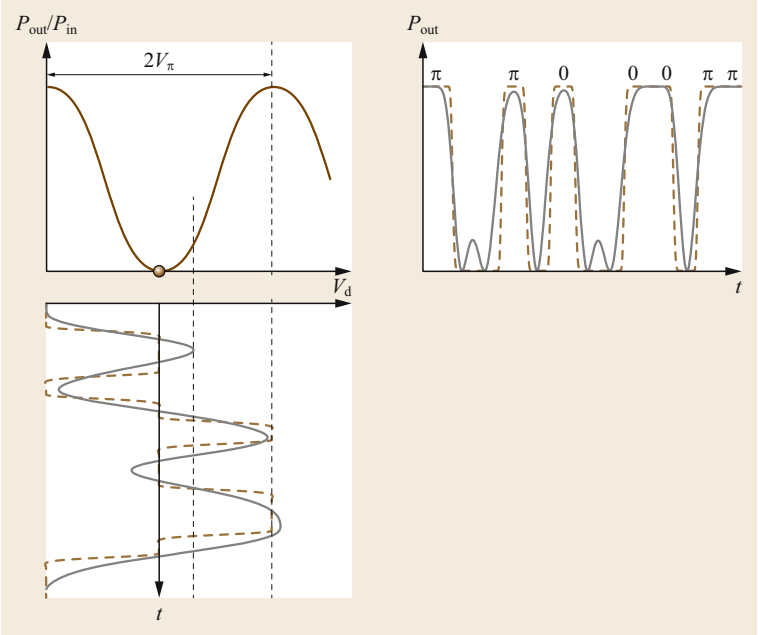


Fig. 4.29 MZM driven by a three-level signal generated as in [4.96] (*dashed lines*) or [4.21] (*solid lines*)

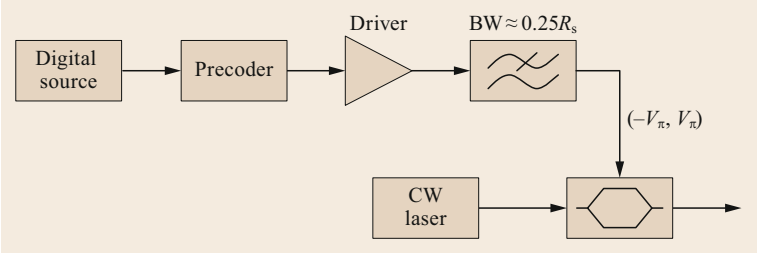


Fig. 4.30 Schematic of the PSBT transmitter

NRZ signal spectrum. Besides its higher resilience to chromatic dispersion [4.18, 19, 98], the narrow optical spectrum of PSBT signals reduces the cross-talk between adjacent channels in ultradense WDM systems [4.17], allowing a higher spectral efficiency than IMDD systems to be achieved.

Among all ODB generation techniques, the PSBT is the one that has become the most popular in optical transmission research because of its simple implementation and good performance. The next subsection is devoted to a performance analysis of ODB transmission implemented through PSBT.

4.3.3 Back-to-Back Performance and Dispersion Tolerance

Several studies have been published on the optimization of electrical and optical filter bandwidths at the transmitter and/or the receiver side [4.22–24, 26, 99], in order to improve the back-to-back performance of PSBT transceivers and approach the theoretical curve shown in Fig. 4.27. All these studies yielded similar results, which are summarized here:

- The optimum electrical filter bandwidth at the transmitter is in the range $0.25-0.35R_s$.
- The use of an optical filter at the transmitter with optimized bandwidth (around $0.75R_s$) can slightly improve the performance.
- The sensitivity can be significantly improved by using a tight optical filter at the receiver (with bandwidth around $0.75R_s$), and the use of the typical bandwidth of MUX/DEMUX in optical systems can significantly degrade the back-to-back performance.
- The optimum electrical bandwidth of the receiver is typically higher than $2R_s$ and the use of the typical bandwidth values around $0.75R_s$ can significantly degrade the performance.

Note, however, that the parameters and results are not all independent. In fact, optical and electrical filter

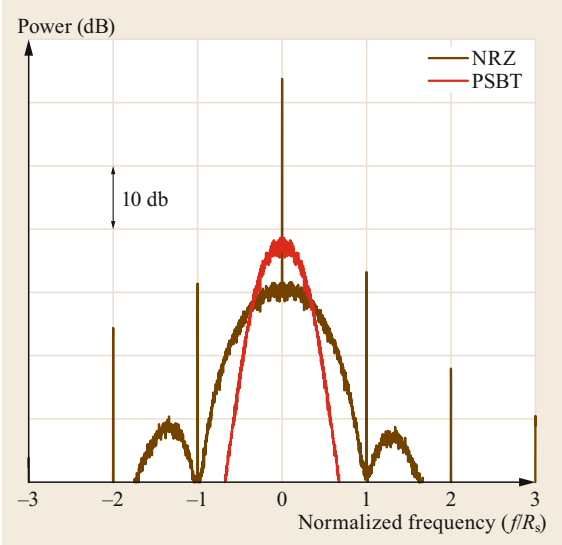


Fig. 4.31 A comparison between the power spectrum of NRZ- and PSBT-modulated signals

bandwidth can be traded-off to some extent. As an example, a narrow optical and wide electrical RX filter can produce similar results to a wider optical filter and a narrower electrical one, as also shown in the following analysis.

Figure 4.32 shows an example of back-to-back performance of PSBT, which compares three different system scenarios at the quantum limit in ASE-noise-limited systems:

- System 1 (an optimized system): 5-pole Bessel electrical low-pass filters with 3-dB bandwidth $0.32R_s$ at the TX and $3.5R_s$ at the RX; second-order super-Gaussian RX optical passband filter with 3-dB bandwidth $0.69R_s$.
- System 2 (a realistic PIN RX, with 7.5-GHz bandwidth at 10 GBd or 30-GHz bandwidth at 40 GBd): 5-pole Bessel electrical low-pass filters with 3-dB bandwidth $0.29R_s$ at the TX and $0.75R_s$ at the RX; second-order super-Gaussian RX optical passband filter with 3-dB bandwidth $0.7R_s$.
- System 3 (standard filter bandwidths at the RX for both electrical and optical filters, for example $B_{RX,el} = 7.5$ GHz and $B_{RX,opt} = 35$ GHz at 10 GBd, with a channel spacing equal to 50 GHz, typical for WDM systems): 5-pole Bessel electrical low-pass filters with 3-dB bandwidth $0.24R_s$ at the TX and $0.75R_s$ at the RX; second-order super-Gaussian RX optical passband filter with 3-dB bandwidth $3.5R_s$.

The optimized PSBT (system 1) has ≈ 0.4 dB penalty with respect to the quantum limit curve both

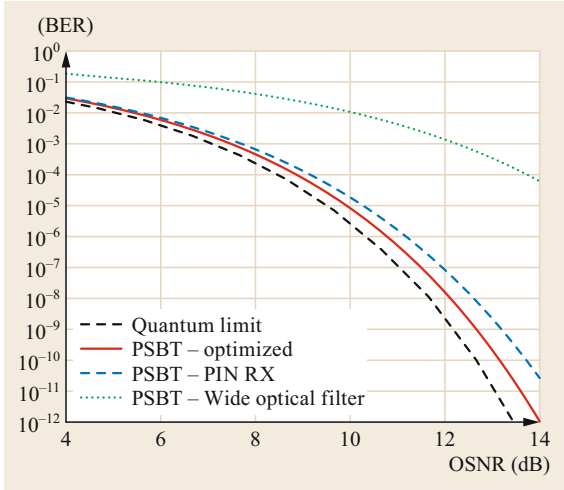


Fig. 4.32 BER versus OSNR characteristics for three configurations of PSBT. The theoretical limit curve is also shown for comparison

at BER $= 10^{-3}$ and BER $= 10^{-9}$. In the more realistic scenario using a standard PIN RX (system 2), the penalty is higher, slightly exceeding 0.6 dB at BER $= 10^{-3}$ and 1 dB at BER $= 10^{-9}$. Note that the use of a wide optical filter at the RX (system 3) yields a much worse performance, i.e., ≈ 5 dB OSNR penalty at a BER of 10^{-3} with respect to the theoretical limit [4.24].

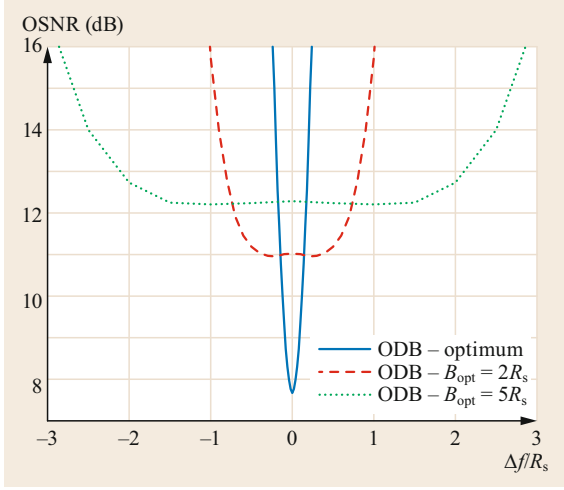


Fig. 4.33 Impact of frequency offset between carrier and optical filter for ODB with optimized RX optical filter bandwidth ($0.69R_s$) and with larger RX filter bandwidths ($2R_s$ and $5R_s$), in terms of required OSNR to obtain a BER equal to 10^{-3}. The electrical filter bandwidths have been optimized in each case

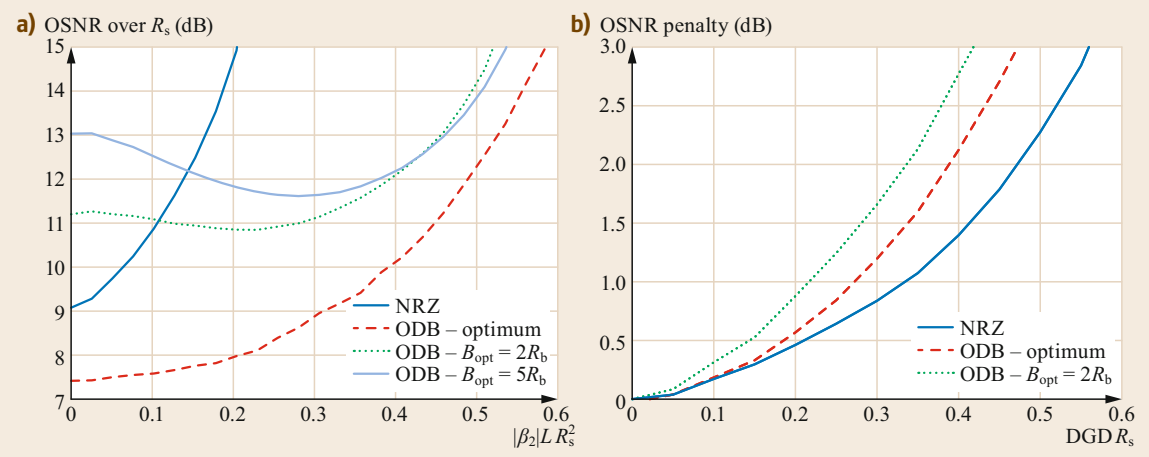


Fig. 4.34 **(a)** GVD tolerance of NRZ and three configurations of ODB in terms of required OSNR (over a bandwidth equal to R_s) to obtain BER $= 10^{-3}$ as a function of the normalized accumulated dispersion. **(b)** First-order PMD tolerance of NRZ and three configurations of ODB in terms of OSNR penalty as a function of the normalized DGD in the worst-case power-splitting ratio between the two PSPs

Impact of RX Optical Filter Detuning

One possible drawback of using narrow-bandwidth optical filtering to tailor the optimum DB pulse shape is the penalty that may be incurred when the carrier frequency drifts away from the filter center frequency [4.23]. Figure 4.33 shows the impact of a frequency detuning Δf between the center frequency of the optical RX filter and the TX laser, in terms of OSNR required to achieve a BER $= 10^{-3}$ as a function of the normalized detuning $\Delta f / R_s$. Because of the narrow bandwidth of the optical filter, the performance of the optimized ODB system is extremely sensitive to the frequency offset. The results shown in Fig. 4.33 demonstrate that the center of the optical filter should be aligned to the carrier frequency within $\pm 0.15 R_s$ in order to maintain the sensitivity advantage over the wide-filter ODB.

Dispersion Tolerance

Because of its narrower spectral width, ODB modulation has a significantly higher tolerance to GVD than binary IMDD. The behavior of ODB in the presence of GVD, though, is strongly dependent on the optical filter bandwidth, as shown in Fig. 4.34a, where the performance of the IMDD-NRZ format is also shown for comparison. The tolerance to PMD is however in general lower for ODB formats [4.35], as shown in Fig. 4.34b. In all cases, the TX and RX electrical filter bandwidths have been optimized in the back-to-back configuration.

4.4 Differential Phase Modulation with Interferometric Detection

This section reviews the transmitter and receiver architectures required to generate and detect different types of optical signals based on differential phase modulation. In Sect. 4.4.1, the concept of differential phase modulation is introduced. Section 4.4.2 describes the optical transmitter structure for the generation of binary and quaternary formats with differential phase modulation, named differential phase-shift keying (DPSK) and differential quadrature phase-shift keying (DQPSK), respectively. Note that DPSK is also referred to as DBPSK in the literature. In Sect. 4.4.3, the interferometric receivers used to extract the phase difference between DPSK and DQPSK symbols are described, while Sect. 4.4.4 shows the back-to-back performance of DPSK/DQPSK systems, focusing on the impact of realistic implementation penalties at the transmitter and receiver sides.

4.4.1 Differential Phase-Shift Keying

Phase-shift keying (PSK) is a modulation technique that encodes the information in the phase of the constellation symbols, keeping the amplitude constant. For instance, in the binary version of PSK (known as BPSK or 2PSK), a "1" may correspond to the phase value 0 and a "0" to the phase value π. In general, for an

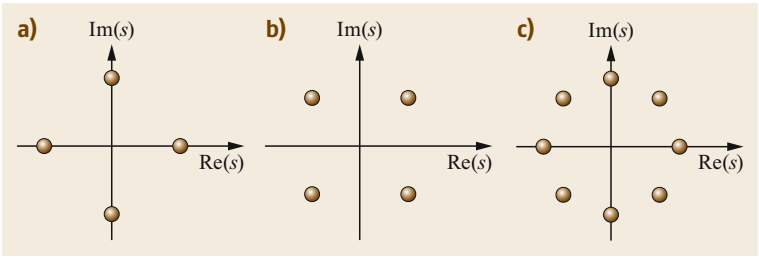


Fig. 4.35a–c Examples of **(a)** 2PSK, **(b)** 4PSK, and **(c)** 8PSK constellations

Table 4.2 The logical operations of the differential precoder for DPSK

Input bit a_k	Phase difference	Output bit p_k
0	π	$\overline{p_{k-1}}$
1	0	p_{k-1}

Table 4.3 The logical operations of the differential precoder for DQPSK [4.101]

Input bits		Phase difference	Output bits	
a_k	b_k		p_k	q_k
0	0	π	$\overline{p_{k-1}}$	$\overline{q_{k-1}}$
0	1	π/2	q_{k-1}	$\overline{p_{k-1}}$
1	0	3π/2	$\overline{q_{k-1}}$	p_{k-1}
1	1	0	p_{k-1}	q_{k-1}

M-level PSK constellation (MPSK), the M information symbols are encoded into M phase values, i.e., $\varphi_n = 2n\pi/M$ with $n = 0, 1, \ldots, M-1$. The complex envelope of PSK-modulated signals can be written as (4.1), with $a_k = e^{j\varphi_k}$ and $\varphi_k \in \{\varphi | \varphi = (2\pi n)/M$, with $n = 0, 1, \ldots, M-1\}$. Figure 4.35 shows examples of PSK constellations with $M = 2$, 4, and 8.

In order to decode a phase-modulated signal, the estimate of the absolute carrier phase is required [4.9]. In practice, however, the carrier phase is extracted from the received signal by performing some nonlinear operation that introduces a phase ambiguity. For instance, the widely used Viterbi&Viterbi phase estimation algorithm [4.100], which raises the signal to the power of M in order to remove the modulation, generates a phase ambiguity equal to $2\pi/M$. Consequently, an absolute estimate of the carrier phase is typically not available for demodulation.

A possible solution to this problem consists in encoding the information in the phase difference between two consecutive symbols. For example, in binary PSK an information bit "1" is transmitted by shifting the phase of the carrier by π with respect to the previous signaling interval, while an information bit "0" is transmitted by applying a zero phase shift relative to the previous signaling interval. In 4PSK the relative phase shifts between successive intervals are 0, π/2, π, and 3π/2, corresponding to the four possible transmitted bit pairs (00, 01, 11, and 10, respectively). Differentially encoded binary ($M = 2$) and quaternary ($M = 4$) modulation formats are often referred to as differential PSK (DPSK) and differential quadrature phase-shift keying (DQPSK).

Tables 4.2 and 4.3 show the logical operations of the differential precoder that is used to encode the information bits on the phase difference of the carrier between two consecutive signal intervals for DPSK and DQPSK, respectively [4.101]. For DPSK, the input bit sequence a_k is encoded onto the output sequence p_k that is used to generate the DSPK waveforms transmitted over the channel. For DQPSK, pairs of input bits (a_k and b_k) are encoded onto two output bit sequences p_k and q_k, corresponding to the in-phase and quadrature components of the DQPSK symbols to be transmitted over the channel.

The optimum detector for PSK signals is based on coherent detection, i.e., the received signal is multiplied by a carrier with the same frequency and phase of the transmitted signal and then goes through a low-pass matched filter [4.9]. Using the optimum demodulator, the bit error probability of both BPSK and QPSK modulations is given by

$$\mathrm{BER} = \frac{1}{2}\mathrm{erfc}\left(\frac{E_\mathrm{b}}{N_0}\right). \tag{4.30}$$

If differential encoding is applied at the transmitter, the same coherent demodulator can be used, followed by a phase comparator that compares the phases of the demodulated signal over two consecutive intervals in order to extract the information. Coherent demodulation of differentially encoded PSK results in a higher BER than standard PSK modulations. In particular, with differentially encoded PSK an error in the demodulated phase of the signal in any given interval will usually result in decoding errors over two consecutive symbols. Consequently, the symbol error rate (SER) of differentially encoded MPSK is approximately twice the SER of MPSK with absolute phase encoding.

As will be shown in Sect. 4.4.3, differentially encoded PSK signals can also be demodulated by a direct-detection optical receiver, composed of a pair of conventional photodetectors, through the use of an

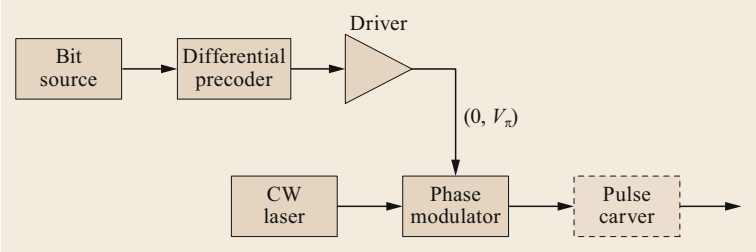


Fig. 4.36 A schematic of a DPSK transmitter using a phase modulator

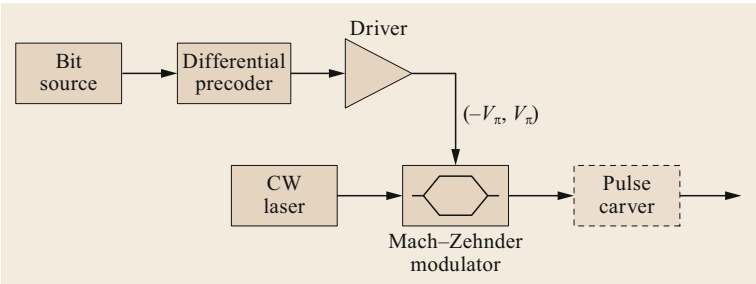


Fig. 4.37 A schematic of a DPSK TX using an MZ amplitude modulator

optical device called the asymmetric MZ interferometer (AMZI) which is able to convert the phase-coded signal into an intensity-coded signal before quadratic detection. The price to pay to avoid the use of standard coherent detection is a performance penalty with respect to the theoretical results determined by (4.30).

4.4.2 DPSK and DQPSK Optical Transmitter

Two possible structures of the DPSK optical transmitter are shown in Figs. 4.36 and 4.37. The first one is based on the use of a phase modulator, driven in the $[0,V_\pi]$ range. The second and most common one employs an MZM, biased at zero power transmission and driven between its transmission maxima. A pulse carver such as the one shown in Fig. 4.11 can be added to generate RZ pulses.

Figure 4.38 shows the schematic of a DQPSK transmitter, which employs an IQ transmitter composed of two MZMs, one for each quadrature of the QPSK modulation. Each MZM is biased at zero power transmission and driven between its transmission maxima. A phase shift of $\pi/2$ is introduced between the two arms before recombining the signals. A pulse carver such as the one shown in Fig. 4.11 can be added to generate RZ pulses.

4.4.3 Interferometric Receiver

The DPSK RX structure is shown in Fig. 4.39. The optical preamplifier is followed by an optical filter and an AMZI with differential delay $T_d = 1/R_s$, where R_s is the symbol rate. Within the AMZI, the input signal is split onto two paths and combined after a path differ-

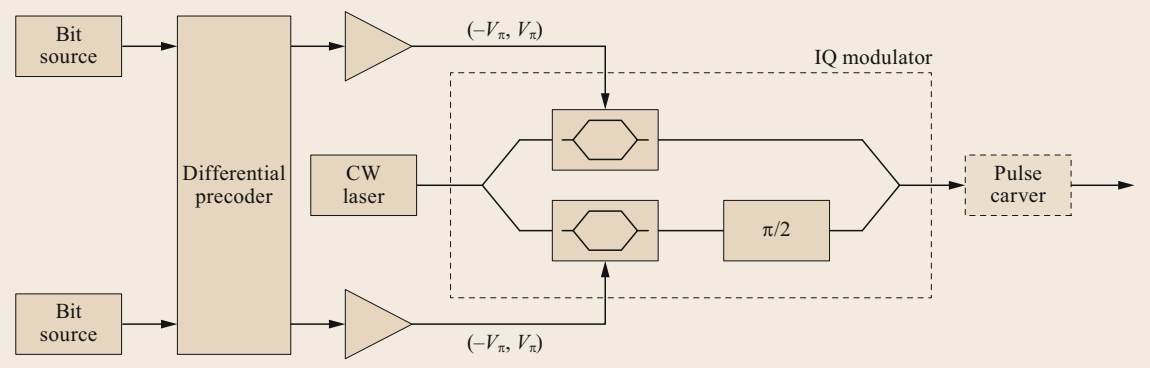


Fig. 4.38 A schematic of a DQPSK TX

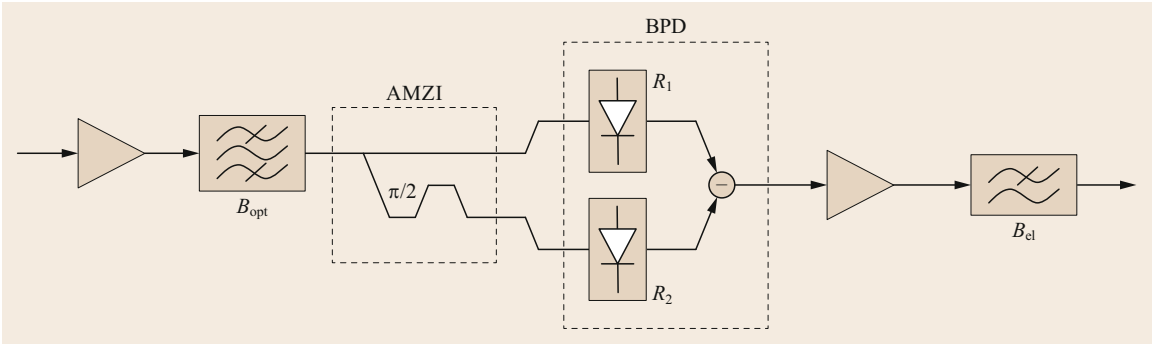


Fig. 4.39 A schematic of a DPSK RX

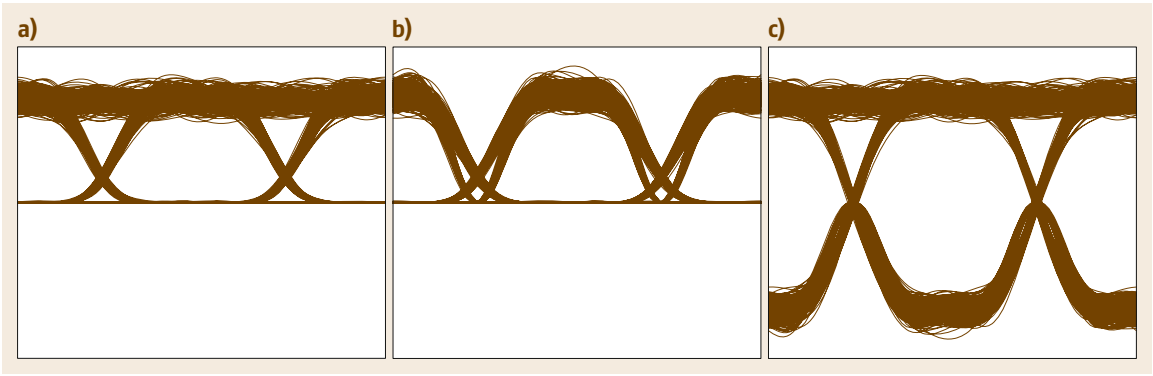


Fig. 4.40a–c Electrical eye diagrams generated at the output of the upper **(a)** and lower arms of the AMZI **(b)** and at the output of the balanced photodetector **(c)**

ence corresponding to one bit delay. If $E(t)$ is the optical field at the input of the AMZI, then the signals at the output of the two arms can be written as

$$E_1(t) = \frac{1}{2}[E(t) + E(t-T_d)e^{j\delta\varphi}], \quad (4.31a)$$

$$E_2(t) = \frac{1}{2}[E(t) - E(t-T_d)e^{j\delta\varphi}], \quad (4.31b)$$

where $\delta\varphi$ is a quantity that should ideally be equal to 0 and that can be controlled, typically by thermal adjustments. A balanced photodetector (BPD) follows, which consists of two photodetectors (one for each output branch of the AMZI), with responsivities R_1 and R_2, connected so as to subtract their currents from each other, and a postdetection electric filter. Ideally, for optimal reception, it should be the case that $R_1 = R_2$ and $\delta\varphi = 0$.

In practice, the AMZI can be viewed as a filter that converts phase modulation into amplitude modulation. The amplitude modulation is then detected by the photodiodes. When two subsequent bits have the same phase then $E_1(t) = E(t)$, $E_2(t) = 0$, while, if they have opposite phase then $E_1(t) = 0$, $E_2(t) = E(t)$. In practice, the AMZI performs a *differential demodulation* of the received signal.

The two ports form slightly different IM signals, which give rise to different *eyes*, as shown in Fig. 4.40, because of the different transfer functions of the two arms of the AMZI (in the absence of phase modulation, the upper and lower branches experience a constructive or destructive interference, respectively). The BPD sums the eyes reinforcing each other. In principle, DPSK can be detected using only one photodiode, on either Out1 or Out2. However, using detection on only one port, the 2.7-dB sensitivity advantage over IMDD is lost [4.102].

The AMZI can also be viewed as a 3-dB splitter followed by a pair of filters whose transfer functions are given by

$$|H_p(f)|^2 = \cos^2\left[\pi(f-f_0)T_d + \frac{\delta\varphi}{2}\right], \quad (4.32a)$$

$$|H_m(f)|^2 = \sin^2\left[\pi(f-f_0)T_d + \frac{\delta\varphi}{2}\right]. \quad (4.32b)$$

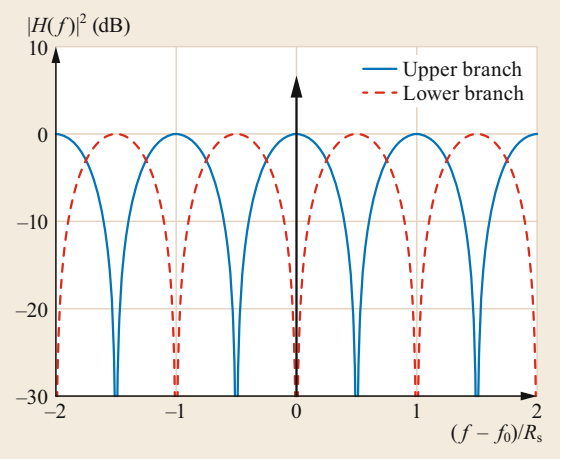


Fig. 4.41 Ideal AMZI transfer functions (given by (4.32), with $\delta\varphi = 0$). The *thick black line* corresponds to the center frequency of the signal at the input of the AMZI

Assuming that DPSK is transmitted at an optical frequency f_0, the optimum demodulation requires that $|H_p(f_0)| = 1$ and $|H_m(f_0)| = 0$. The phase error $\delta\varphi$ can thus be seen as an incorrect tuning of the AMZ filter, by a quantity $\Delta f = R_s\delta\varphi/(2\pi)$. This corresponds to a shift by Δf of the optical frequency of the AMZ filter transfer function. Figure 4.41 shows the AMZI transfer functions of (4.32) when $\delta\varphi = 0$.

DQPSK Receiver

The DQPSK RX structure is shown in Fig. 4.42. It consists of an optical filter, a pair of AMZIs with a differential delay ideally equal to the time duration of a transmitted symbol, each followed by a BPD consisting of two photodetectors (one for each output branch of the AMZIs), connected so as to subtract their currents from each other, and then followed by a postdetection electric filter. The differential optical phase between the interferometer arms should be set to $\pi/4$ and $-\pi/4$ for the upper and lower branches, respectively.

If $E(t)$ is the optical field at the output of the RX optical filter, the signals at the output of the upper AMZI can be written as

$$E_1(t) = \frac{1}{2}[E(t) + E(t - T_d)e^{j\frac{\pi}{4}}e^{j\delta\varphi}], \tag{4.33a}$$

$$E_2(t) = \frac{1}{2}[E(t) - E(t - T_d)e^{j\frac{\pi}{4}}e^{j\delta\varphi}], \tag{4.33b}$$

where $\delta\varphi$ is a quantity that can be controlled, typically by thermal adjustments of the AMZ; $\delta\varphi$ can be transformed into a frequency-detuning parameter Δf of the AMZI transfer function through the relation $\Delta f = R_s\delta\varphi/(2\pi)$. The value $\Delta f = 0$ corresponds to perfect tuning of the AMZI.

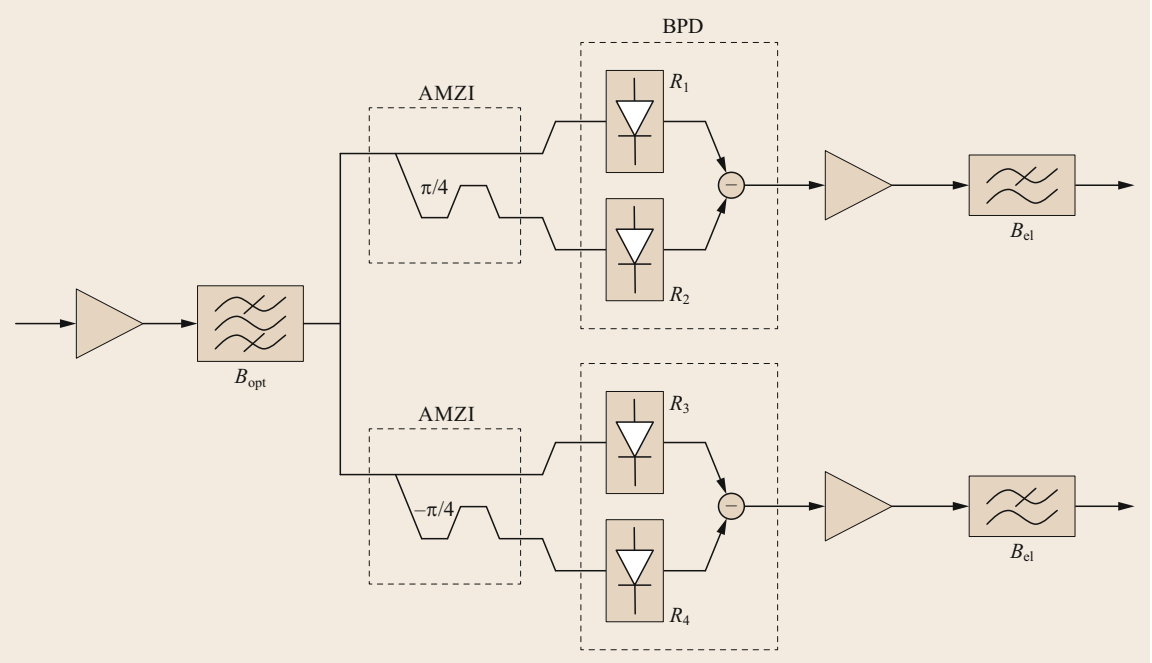


Fig. 4.42 A schematic of a DQPSK RX

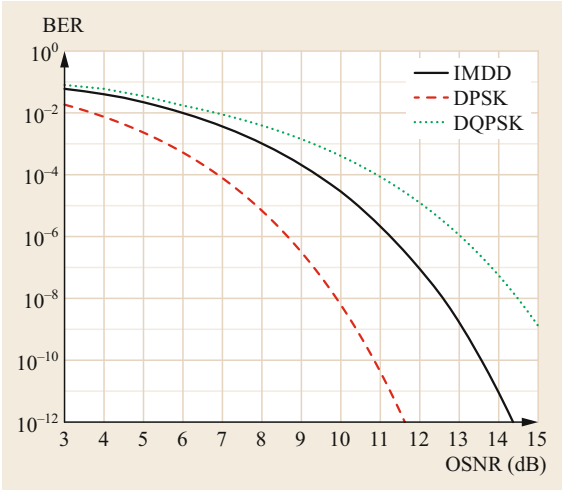


Fig. 4.43 Ideal performance of DPSK and DQPSK systems with interferometric detection. The performance of IMDD is also shown for comparison

4.4.4 Back-to-Back Performance and Dispersion Tolerance

The best performance is achieved when the receiver contains an optical filter matched to the received optical signal, ideal photodetectors, and no postdetection electrical filters. In this ideal case, and in the absence of any TX or RX impairment, the BER achievable with the DPSK modulation is equal to [4.29, 103]

$$\mathrm{BER} = \frac{1}{2}\mathrm{e}^{-2\mathrm{OSNR}}\left(1+\frac{1}{2}\mathrm{OSNR}\right), \qquad (4.34)$$

where OSNR is defined by (4.15) over a bandwidth equal to the symbol-rate R_s. The BER versus OSNR curve for DPSK is shown in Fig. 4.43 as a dashed red line. Comparing it to the optimum matched filter performance of OOK-IMDD (solid black line), an OSNR gain of approximately 3 dB is achieved by DPSK. The performance curve of DQPSK is shown in Fig. 4.43 as a dotted green line. It has been estimated using the semianalytical technique described in [4.104], since a closed-form formula for performance evaluation is not available for DQPSK. The reference bandwidth for the OSNR is equal to R_s, thus the reported curves give a direct comparison of systems operating at the same symbol rate. DQPSK allows transmitting information at a double rate with respect to the binary formats, at the expense of an OSNR penalty of ≈ 1 dB with respect to IMDD and ≈ 3.5 dB with respect to DPSK, at a reference BER of 10^{-3}.

The sensitivity gain of DPSK and DQPSK with respect to IMDD depends on the practical implementation of the transceiver (for example on the electrical and optical filter bandwidths). In [4.34, 102, 104] it was shown that, like for OOK formats, a careful optimization of the filter bandwidths can yield a performance very close to the theoretical matched filter limits, provided that a narrowband optical filter is available at the receiver. If the optical filter bandwidth is wide, the performance penalty with respect to the ideal performance can be kept below 1 dB by optimizing the TX and RX electrical bandwidths [4.104].

Table 4.4 shows an example of optimized 3-dB filter bandwidths when a second-order super-Gaussian optical passband filter and a 5-pole Bessel electrical low-pass filter are used at the RX. Six different cases have been considered, i.e., IMDD, DPSK, and DQPSK all with either NRZ or RZ modulation. In the case of NRZ modulations, the transmitted pulses are shaped by a 5-pole Bessel low-pass filter with 3-dB bandwidths equal to $0.75R_s$ for IMDD, $0.9R_s$ for DPSK, and $0.45R_s$ for DQPSK. The OSNR (over a bandwidth equal to R_s) required to achieve BER $= 10^{-3}$ is also shown, together with the penalty with respect to the matched filter case.

Interferometric Receiver Implementation Impairments

The receiver architecture used for both DPSK and DQPSK is more complex than a standard DD receiver, requiring the additional use of an AMZI and a pair of BPDs for each quadrature. The performance of the system can thus be affected by imperfections in the AMZI or in the BPD:

Table 4.4 Optimized RX filter bandwidths in different configurations and corresponding system performance and OSNR penalty with respect to the ideal matched filter receiver

Case	B_{opt}/R_s	$B_{el,RX}/R_s$	OSNR@BER $= 10^{-3}$ (dB)	OSNR penalty (dB)
IMDD-NRZ	1.2	0.8	9.0	1.0
IMDD-RZ	2.0	1.0	8.5	0.5
DPSK-NRZ	1.2	1.6	5.9	0.3
DPSK-RZ	2.0	1.2	5.7	0.1
DQPSK-NRZ	1.2	1.6	9.7	0.4
DQPSK-RZ	2.0	1.2	9.5	0.2

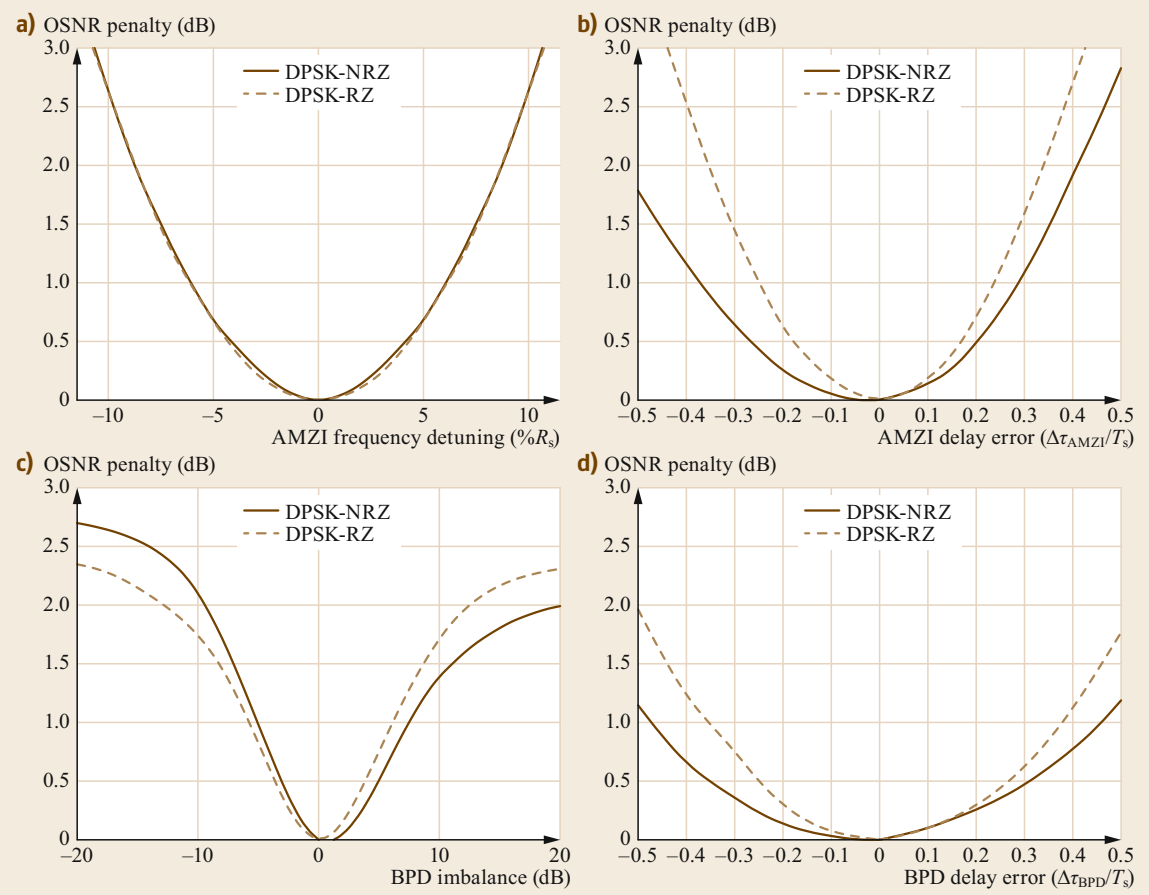


Fig. 4.44a–d Impact of interferometric RX implementation impairments on DPSK-NRZ and DPSK-RZ in terms of OSNR penalty. The target BER is 10^{-3}, and the values of reference OSNR (over a bandwidth equal to R_s) in the absence of impairments are shown in Table 4.4

- AMZI frequency-detuning Δf with respect to the transmit laser, due to a nonzero phase mismatch $\delta\varphi$ between the two branches of the AMZI ((4.32) and (4.33)): $\Delta f/R_s = \delta\varphi/(2\pi)$.
- AMZI delay error $\Delta\tau_{AMZI}$, due to a mismatch between the delay T_d between the two arms of the AMZI and the symbol time T_s: $\Delta\tau_{AMZI} = T_d - T_s$.
- BPD delay imbalance $\Delta\tau_{BPD}$, due to a time mismatch in the propagation paths in the two arms of the BPD.
- BPD gain imbalance K, due to a mismatch between the responsivities R_1 and R_2 of the photodiodes in the two arms of the BPD defined as $K = R_1/R_2$.

In the following, the impact of such impairments on the performance is estimated for both DPSK and DQPSK, in terms of OSNR penalty with respect to the ideal case with no imperfections at a reference BER equal to 10^{-3} (which assumes the use of a hard-decision FEC scheme). The filter parameters used in simulations are those shown in Table 4.4. The results, shown in Fig. 4.44 for DPSK and Fig. 4.45 for DQPSK, indicate that the tolerance to AMZI and BPD imperfections is in general higher for DPSK than for DQPSK, when operating at the same symbol rate (note that all time and frequency errors in the x-axes of Figs. 4.44 and 4.45 are normalized with respect to symbol time/symbol rate). RZ-DPSK is less tolerant that NRZ-DPSK to delay errors (both in the AMZI and in the BPD), whilst the impact is almost the same for RZ-DQPSK and NRZ-DQPSK. The most critical parameter is the frequency detuning between the AMZI and the transmit laser: an offset of only 6% of the symbol rate induces a 1-dB penalty on DPSK and more than 3-dB penalty on DQPSK. Note that this frequency offset could fully be compensated for by proper tuning of the path length difference within the AMZI on an optical wavelength scale, by inserting a phase shifter in

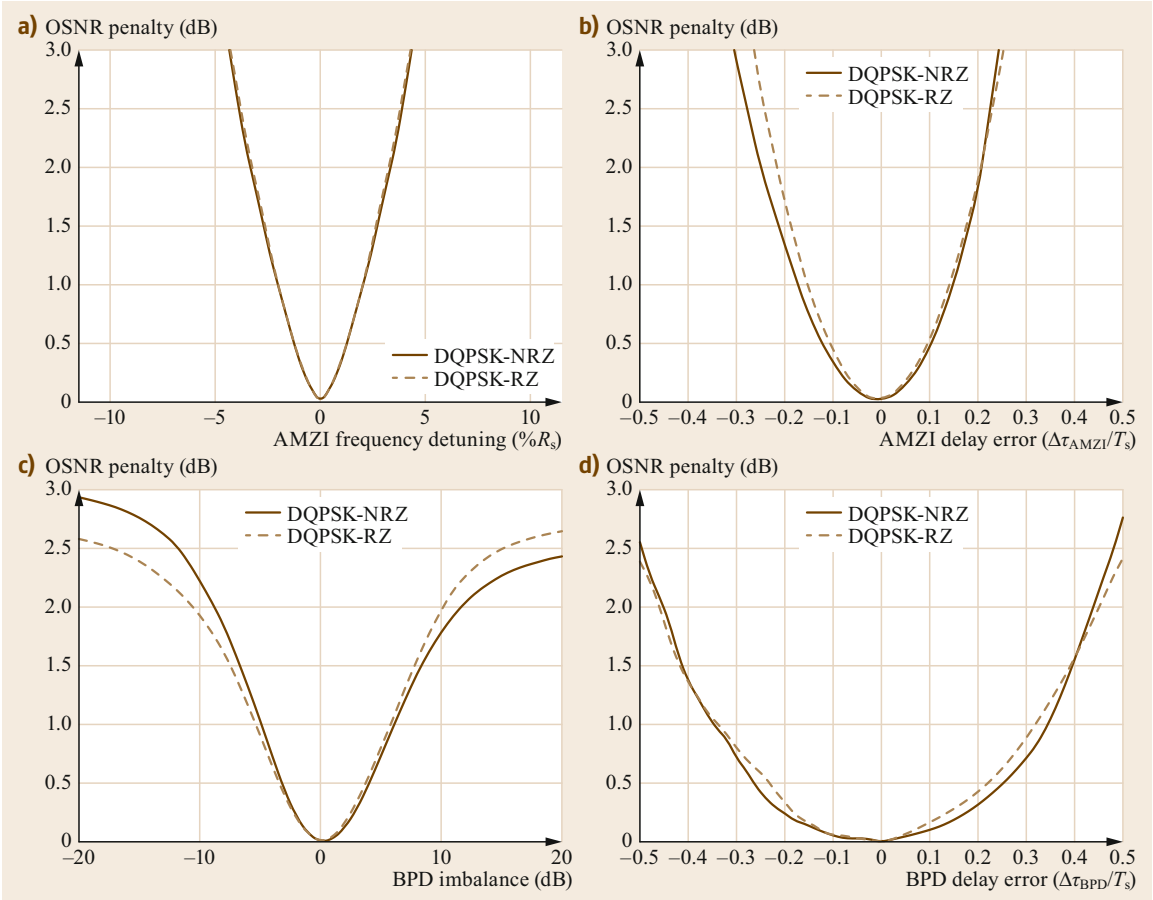


Fig. 4.45a–d Impact of interferometric RX implementation impairments on DQPSK-NRZ and DQPSK-RZ in terms of OSNR penalty. The target BER is 10^{-3}, and the values of reference OSNR (over a bandwidth equal to R_s) in the absence of impairments are shown in Table 4.4

one of the AMZI paths [4.105]. However, the results shown in Figs. 4.44a and 4.45a may set stringent requirements to the wavelength stability of the transmit laser. Similar analyses can be found in [4.104–108], where much lower values of reference BER (10^{-9} or 10^{-10}) were used, which imply a slightly lower tolerance to the imperfections than the one observed at BER $= 10^{-3}$.

Dispersion Tolerance

Figure 4.46 shows the performance of DPSK and DQPSK modulation formats (with either RZ or NRZ line coding) in the presence of either GVD or a first-order PMD. The simulation setup is the same as described in Sect. 4.2.4, with optimized RX filter bandwidth (as shown in Table 4.4). Linear fiber propagation was assumed, while first-order PMD was emulated by inserting a deterministic DGD value $\Delta\tau_{DGD}$ between the two principal states of polarization. The worst-case of equal power-splitting ratio between the two PSPs was considered. The curve for IMDD-NRZ is also shown for comparison.

For the system configurations considered in Fig. 4.46, the highest GVD tolerance is achieved by DPSK-NRZ, while the other formats have a much lower GVD tolerance. However, it must be noted that the comparison is performed at the same symbol rate, i.e., DQPSK has double bit rate with respect to DPSK and IMDD. The tolerance to GVD for DQPSK operating at the same bit rate would thus be increased by a factor of 4, achieving a higher tolerance than the binary formats, thanks to its narrower spectral width. The DGD tolerance is instead similar for all formats (at the same symbol rate).

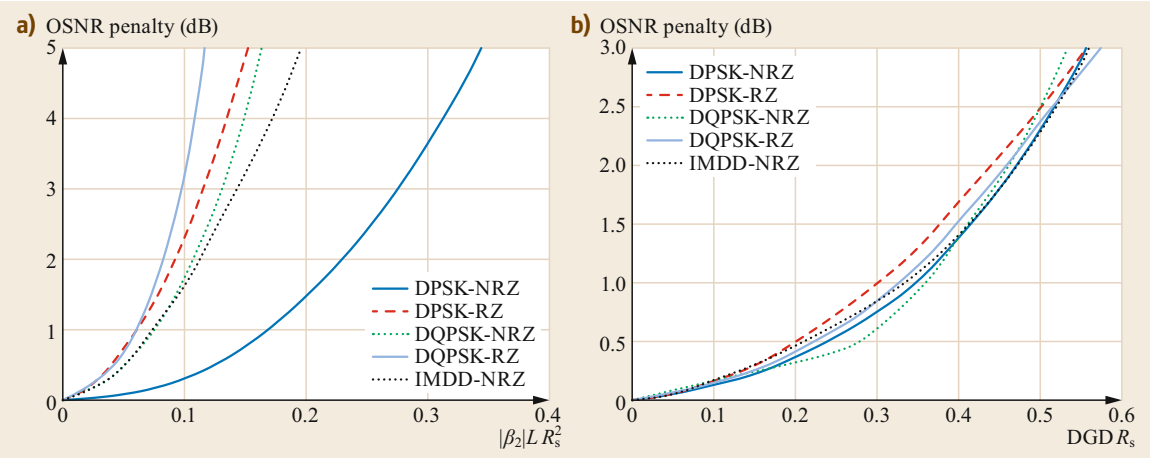


Fig. 4.46a,b GVD and first-order PMD tolerance of DPSK and DQPSK in terms of OSNR penalty with respect to back-to-back performance as a function of either **(a)** the normalized accumulated dispersion or **(b)** the normalized DGD in the worst-case power-splitting ratio between the two PSPs. The performance of IMDD-NRZ is also shown for comparison. The target BER is 10^{-3} and the values of reference back-to-back OSNR (over a bandwidth equal to R_s) are shown in Table 4.4

4.5 IQ Modulation with Coherent Detection

This section reviews the transmitter and receiver architectures required to generate and detect different types of high-order modulation formats employing coherent detection. Section 4.5.1 reviews the architecture of an IQ transmitter, which can be used to generate arbitrary two-dimensional (2D) constellations or four-dimensional (4D) constellations if polarization multiplexing is employed. The architecture of a coherent-detection optical receiver is described in Sect. 4.5.2, while Sect. 4.5.3 introduces the Nyquist-WDM multiplexing technique, highlighting the compromise between spectral efficiency and performance. Finally, Sect. 4.5.4 focuses on flexible transponders, designed to dynamically adapt the transmission speed and modulation format to the channel conditions.

4.5.1 IQ Transmitter

Combining amplitude modulation and phase modulation, the class of modulation formats known as QAM (quadrature amplitude modulation) is obtained. The complex envelope of QAM modulated signals can be written as [4.9]

$$s_T(t) = \sum_k (p_k + \mathrm{j}q_k) g_T(t - kT_s)\,, \qquad (4.35)$$

where p_k and q_k are the coordinates of the constellation points on the in-phase and quadrature axes, respectively, T_s is the symbol time, and $g_T(t)$ is the optical pulse shape in the time domain. For each symbol slot, M alternative waveforms can be transmitted, depending on the value of the information bits; M is called the cardinality of the constellation. The maximum number of bits n_b that can be carried by each constellation symbol is equal to $\log_2(M)$. The transmission speed, i.e., the bit rate, is given by $R_b = n_b/T_s$. Figure 4.47 shows a few examples of QAM constellations (for $M = 4, 16, 32, 64$).

Any optical QAM signal could be in principle generated by using a single dual-drive MZM [4.109]. However, the required number of levels of the electrical driving signals would be quite high (for example for a 16-ary driving signals are needed for 16QAM), trading the simplicity of the optical part for a big electrical effort. Another method to generate arbitrary QAM constellations is based on the use of two consecutive optical modulators: an MZM for amplitude modulation and a phase modulator to set the phase. Also in this case, the simplicity of the optical TX part implies the use of a complex electrical level generator which cannot be easily implemented for high data rates [4.110].

The most commonly used TX configuration to generate QAM signals is based on an IQ modulator, like the one used in the generation of DQPSK signals (Fig. 4.38). This configuration exploits the fact that each constellation point can be projected on the in-phase and

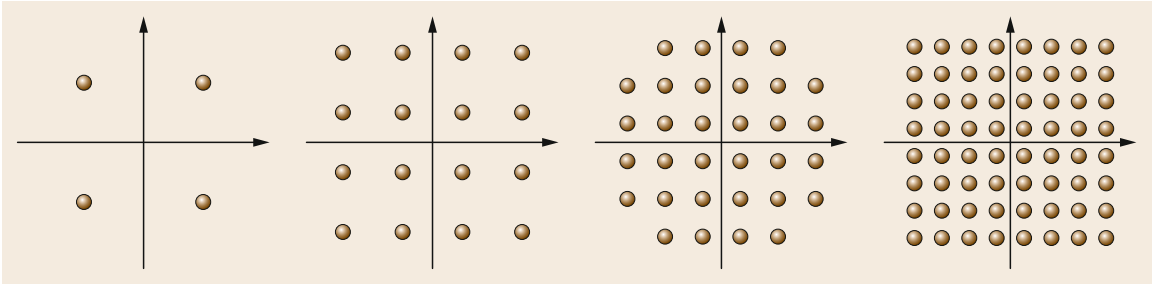


Fig. 4.47 High-order square QAM modulation format constellations (for $M = 4, 16, 32, 64$)

quadrature axes, in order to reduce the number of required levels for the driving signal (which is equal to $\log_2(M)$ for M-QAM). The scheme of the conventional IQ transmitter is shown in Fig. 4.48. Two multilevel electrical driving signals are generated (one for the in-phase and the other for the quadrature component of the QAM signal), each of them driving one of the two MZMs in the IQ modulator. In general, an adaptive and precise bias control algorithm for the IQ modulator is essential to minimize transmitter impairments and to maintain long-term stable operation [4.111–115].

Combining two IQ transmitters using a polarization beam splitter/combiner, polarization-multiplexed QAM (PM-QAM) signals can be generated, as shown in Fig. 4.49. The light from a single laser signal is split and sent into four separate Mach–Zehnder modulators. The upper and lower portions of this super-MZM structure each generate an M-QAM signal. The signals are then sent into a polarization beam combiner (PBC), after applying a polarization rotation to the lower branch, so that the signal from the upper half of the circuit becomes X-polarized, while the signal from the lower half of the circuit becomes Y-polarized.

Generation of the Multilevel Driving Signals

Multilevel driving signals with a desired pulse shape can be efficiently generated using a digital signal processing (DSP) block, followed by a digital-to-analog converter (DAC) which generates the driving electrical signal. Any DAC device is characterized by two main parameters, which set a limit to its performance. These are:

- The sampling speed f_{DAC}, which limits the achievable symbol rate $R_{\mathrm{s}} = f_{\mathrm{DAC}}/N_{\mathrm{SpS}}$, where N_{SpS} is the number of samples per symbol (SpS), also indicated as the *oversampling factor*.
- The number of resolution bits N_{DAC}, which limits the cardinality of the modulation format. In fact, the higher the order of the modulation format, the higher is the required value of N_{DAC}.

Typically, if f_{DAC} increases, N_{DAC} decreases. The achievable symbol rate can clearly be increased by decreasing the oversampling factor. In doing so, penalties could be incurred due to interference produced by spectral replica of the useful spectrum in the DAC process. In [4.116], a 1.5 SpS DAC-supported Nyquist-WDM PM-16QAM experiment was reported, using a DAC with $f_{\mathrm{DAC}} = 23.4$ GHz and thus achieving a symbol rate $R_{\mathrm{s}} = 15.6$ GBd. In [4.117], 1.33 SpS were employed in a 100-km PM-64QAM single-channel transmission at 252 Gb/s. In [4.118], an oversampling factor as low as 1.15 was used, limiting the penalty due to spectrum replica thanks to the use of antialiasing electrical

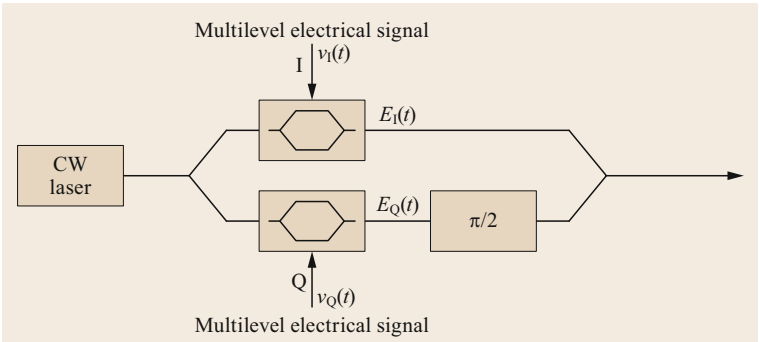


Fig. 4.48 A single polarization IQ transmitter

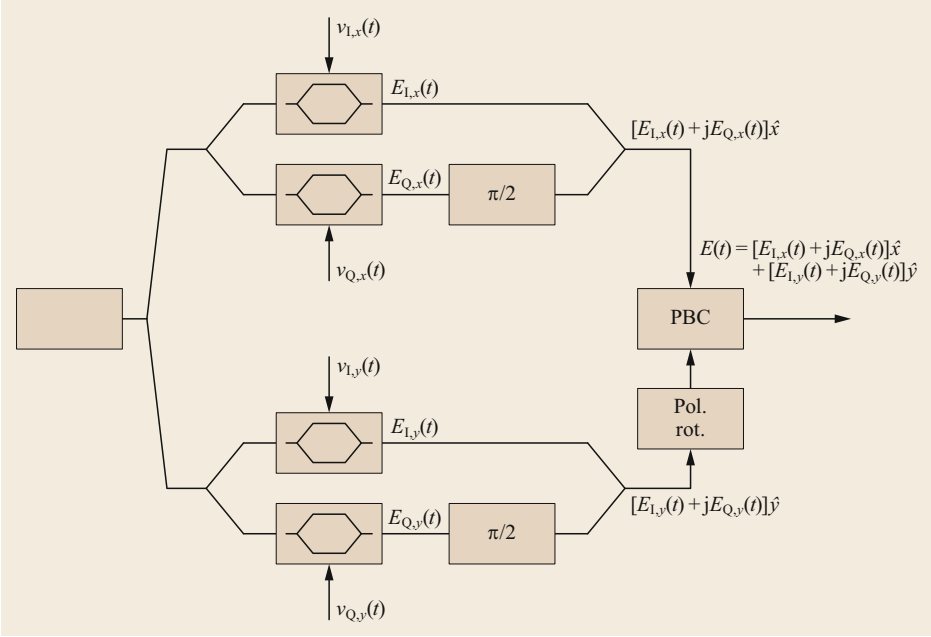


Fig. 4.49 A dual polarization IQ transmitter (PBC: polarization beam combiner)

filters. The distortions on the useful signal introduced by the bandwidth limitations of the TX can be compensated for in the digital domain using a properly designed pre-equalizer in the DSP block before the DAC [4.119].

4.5.2 Coherent Optical Receiver

The basic structure of the most commonly used coherent optical receiver is shown in Fig. 4.50. It is a phase-diversity RX [4.39, 120] which is able to retrieve the in-phase and quadrature components of the complex amplitude of an optical signal without locking the frequency and phase of the local oscillator (LO). The only requirement is that the frequency offset between the TX laser and the LO is limited to a fraction of the symbol rate. This type of coherent detection without active optical phase lock loop is also known as *intradyne* coherent detection.

In the schematic shown in Fig. 4.50, the optical data signal is mixed with a signal generated by an LO through a six-port optical device called a 90° *hybrid* [4.121], which consists of a two-by-two optical coupler with a 90° phase delay function implemented in one arm of the coupler. The input/output characteristic of a 90° hybrid is given by

$$\begin{pmatrix} E_1(t) \\ E_2(t) \\ E_3(t) \\ E_4(t) \end{pmatrix} \begin{pmatrix} +1 & +1 \\ +1 & -1 \\ +1 & -\mathrm{j} \\ +1 & +\mathrm{j} \end{pmatrix} \begin{pmatrix} E_S(t) \\ E_{LO}(t) \end{pmatrix} , \tag{4.36}$$

with

$$E_S(t) = \sqrt{P_S(t)}\,\mathrm{e}^{\mathrm{j}[2\pi f_S t+\varphi_D(t)]} , \tag{4.37a}$$

$$E_{LO}(t) = \sqrt{P_{LO}}\,\mathrm{e}^{\mathrm{j}[2\pi f_{LO} t+\varphi_{LO}(t)]} . \tag{4.37b}$$

The output signals are then converted from the optical domain into the electronic domain with a pair of balanced photodetectors, whose output currents are given by

$$I_I(t) = R\left(|E_1(t)|^2 - |E_2(t)|^2\right) , \tag{4.38a}$$

$$I_Q(t) = R\left(|E_3(t)|^2 - |E_4(t)|^2\right) , \tag{4.38b}$$

where R is the responsivity of the photodiode (which is ideally the same for all four photodetectors). Substituting (4.36) and (4.37) into (4.38), gives

$$\begin{aligned} I_I(t) &= 4R\,\mathrm{Re}(E_S(t)E_{LO}^*(t)) \\ &= 4R\sqrt{P_S(t)P_{LO}^*}\cos\left[2\pi\delta f t+\varphi_D(t)-\varphi_{LO}(t)\right] , \end{aligned} \tag{4.39}$$

$$\begin{aligned} I_Q(t) &= 4R\,\mathrm{Im}(E_S(t)E_{LO}^*(t)) \\ &= 4R\sqrt{P_S(t)P_{LO}^*}\sin\left[2\pi\delta f t+\varphi_D(t)-\varphi_{LO}(t)\right] , \end{aligned} \tag{4.40}$$

where $\delta f = f_S - f_{LO}$ is the frequency offset between the TX laser and the LO. Combining the two photocurrents

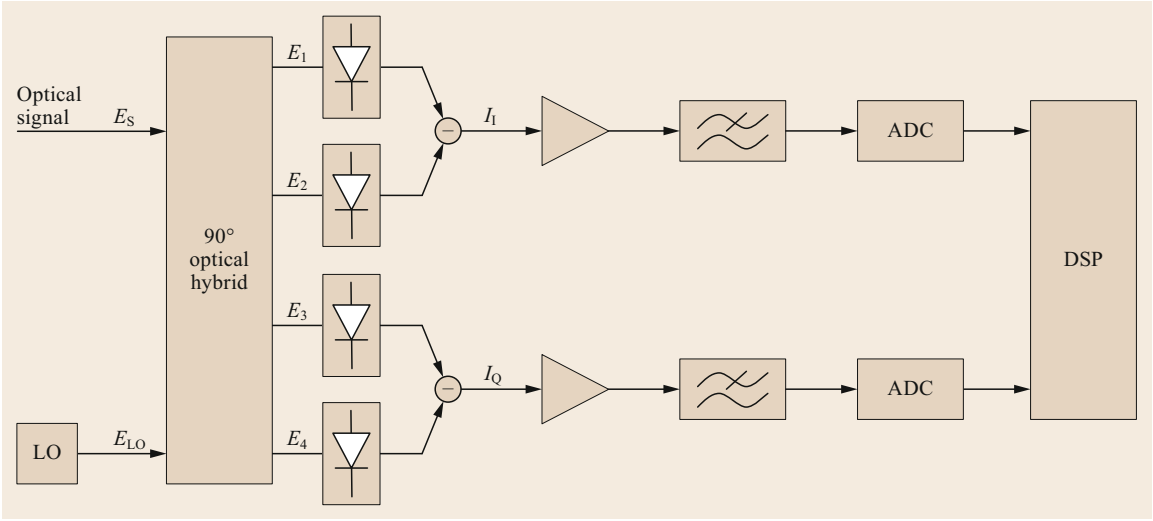


Fig. 4.50 A coherent intradyne receiver based on phase-diversity detection (ADC: analog-to-digital converter, LO: local oscillator; DSP: digital signal processing)

into a complex signal format gives

$$I_{IQ}(t) = I_I(t) + jI_Q(t) = 4R\sqrt{P_S(t)P_{LO}^*}\,e^{j[2\pi\delta ft+\varphi_D(t)-\varphi_{LO}(t)]}\,, \quad (4.41)$$

which has the same phase as the data signal $E_s(t)$ except for the frequency offset and the additional phase noise introduced by the LO. Both, phase noise and frequency offset, can be compensated for using properly designed DSP algorithms that can be applied to the received samples after the analog-to-digital conversion process, together with the compensation of the propagation effects. Note that, if a single photodetector was used instead of a balanced one, it would be necessary to employ a local oscillator laser with 20–25 dB more power than the incoming signal in order for the resulting coherent signal term to dominate [4.121]. By using a balanced photodetector it is possible to employ a much less powerful local oscillator.

The electrical low-pass filters (LPFs) shown in Fig. 4.50 represent the cascade of all band-limiting components in the RX. An additional antialiasing electrical filter may be needed before the ADCs in order to reduce the bandwidth of the input signal, thus relaxing the requirements for the ADC sampling frequency f_{ADC}, which, in order to avoid a performance degradation due to *aliasing* [4.9], has to be higher than twice the bandwidth of the input signal. The distortions introduced on the useful signal by the bandwidth limitations of the RX can be compensated for in the digital domain by the adaptive or static equalizers present in the DSP chain (refer to Chap. 6 for further details).

A polarization controller is needed at the input of the receiver in Fig. 4.50 in order to align the polarization states of the LO and the data signal. However, in practical systems, the polarization of the incoming signal is unlikely to remain aligned to the SOP of the LO because of random changes in the birefringence of transmission fibers [4.122]. To deal with this problem, polarization-diversity detection has been introduced into the coherent receiver. The schematic of the dual-polarization coherent receiver, based on polarization diversity, is shown in Fig. 4.51. The incoming optical signal passes through a polarization beam splitter (PBS), which divides the signal into two orthogonal polarization signal components. Each component is combined with an LO through an independent 90° optical hybrid circuit. The eight signals are then converted from the optical domain into the electronic domain using four balanced photodetectors. In the case of the polarization-multiplexed system, the polarization tributaries can be demultiplexed from the four outputs of the receiver using DSP algorithms [4.45].

Customarily, the performance of optical coherent systems is estimated by means of the OSNR, defined by (4.15). To find the BER, the OSNR is inserted into a suitable formula, which depends on the transmission format. For instance, the ideal BER performance of square PM-QAM formats with M constellation points is given by [4.9]

$$\mathrm{BER}_{\mathrm{PM-MQAM}} = \frac{2(1-1/\sqrt{M})}{\log_2 M} \times \mathrm{erfc}\left(\sqrt{\frac{3\mathrm{OSNR}}{2\,(M-1)}}\right), \quad (4.42)$$

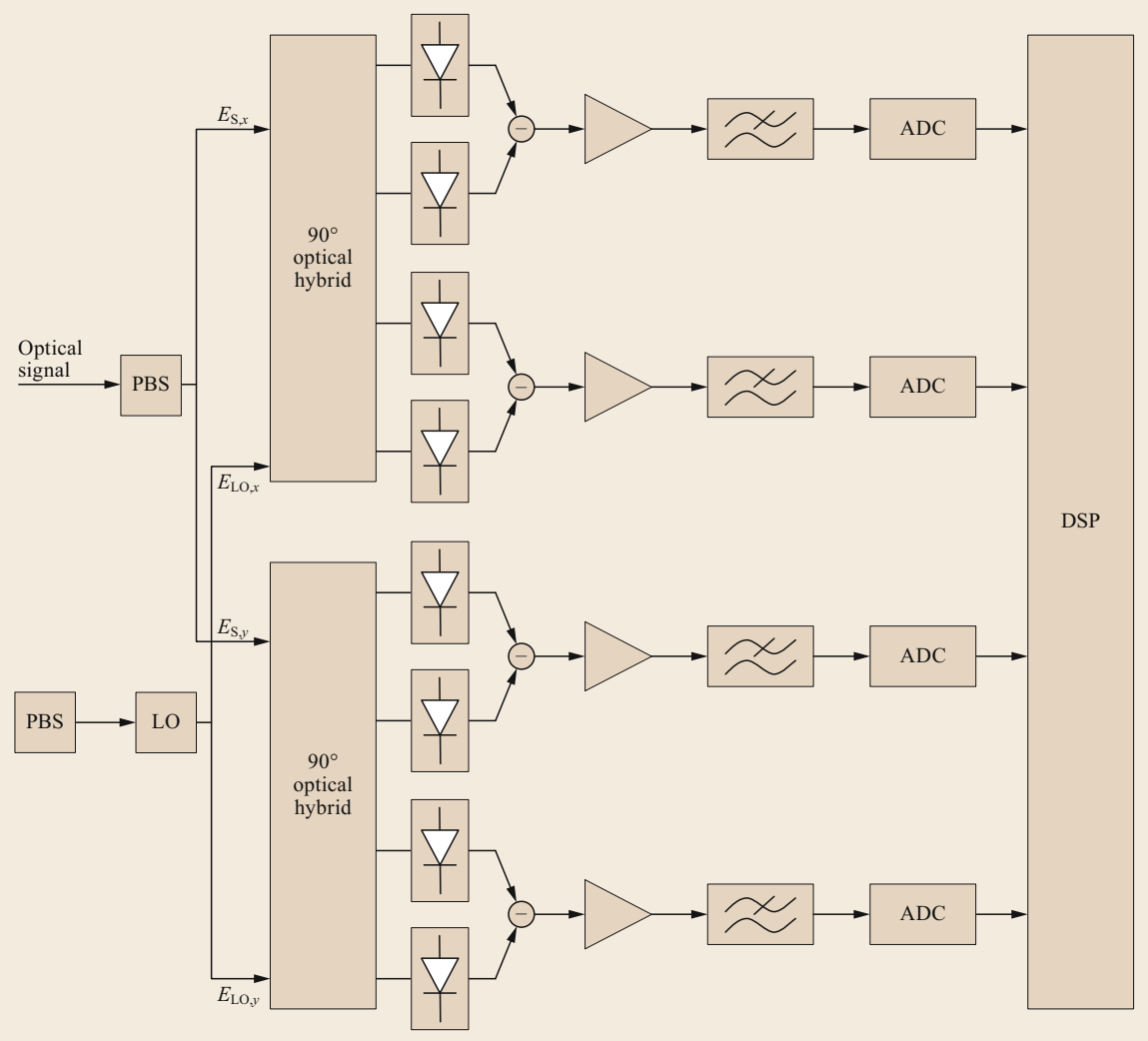


Fig. 4.51 A dual-polarization coherent intradyne receiver based on phase- and polarization-diversity detection (ADC: analog-to-digital converter, LO: local oscillator; DSP: digital signal processing; PBS: polarization beam splitter)

where OSNR is measured over a bandwidth equal to the symbol rate. Equation (4.42), as well as textbook formulas addressing other formats, assume that the RX operates by filtering the incoming signal through a baseband transfer function $H_{RX}(f)$ that is matched to the transmitted signal baseband pulse. It is also assumed that the ISI be absent. Otherwise, they are no longer valid, in the sense that there is a penalty with respect to what they predict. Note that in modern coherent systems, the DSP adaptive equalizer tends to make $H_{RX}(f)$ converge to a matched shape, so that this condition is typically well satisfied.

Under the mentioned assumptions of a matched $H_{RX}(f)$ and no ISI, the OSNR in (4.15) also corresponds exactly to the signal-to-noise ratio SNR that can be directly measured on the RX electrical signal constellation, at the input of the decision stage. In addition, such SNR also coincides with the communications theory widely used parameter *energy-per-symbol versus noise spectral density* [4.9], that is SNR $= E_s/N_0$.

4.5.3 Nyquist-WDM

Today, transmission at 100 Gb/s, enabled by the use of quadrature phase-shift keying (QPSK) at 32 GBd with polarization multiplexing (PM), has reached maturity and widespread deployment, and 200-Gb/s PM-16QAM coherent transceivers are already commer-

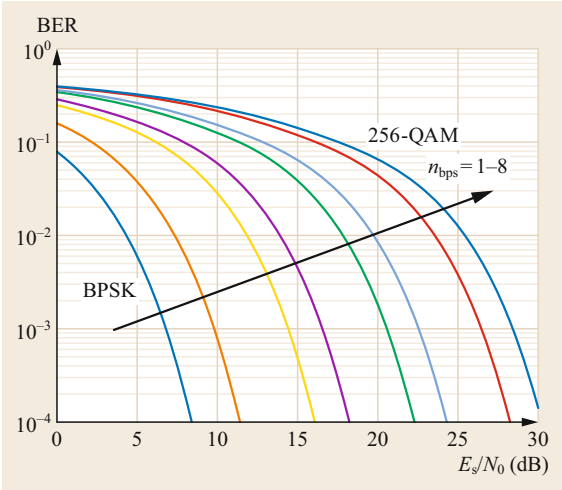


Fig. 4.52 Ideal back-to-back performance of high-order QAM formats with coherent detection

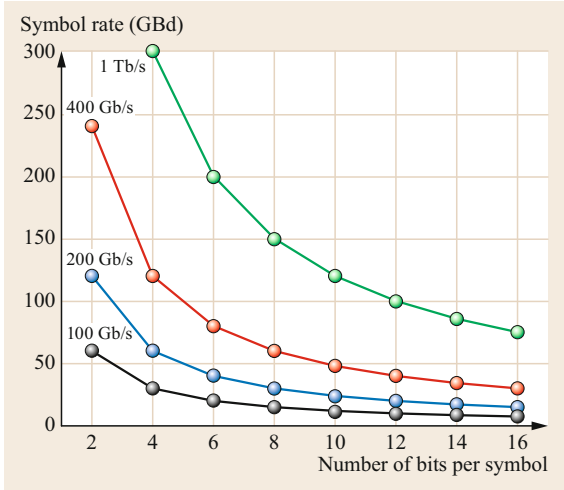


Fig. 4.53 Symbol rate versus number of bits per symbol needed to achieve a target net transmission speed of 100, 200, 400 Gb/s, or 1 Tb/s (assuming that an FEC with 20% overhead is used) (after [4.127])

cially available. Industry efforts are currently devoted to the ASIC development for transceivers for 400 Gb/s and soon they will move to 1 Tb/s. Since the transmission speed R_b is given by the product of the symbol rate R_s and the number of bits per symbol n_{bps} ($R_b = R_s n_{bps}$), higher values of R_b can be obtained by increasing either R_s or n_{bps} (i.e., the order of the modulation format). Note that the first solution would have a strong impact on the hardware complexity, whilst the second one would mainly impact the DSP complexity. As an example, a raw transmission rate of 480 Gb/s can be achieved either with 120-GBd PM-QPSK ($n_{bps} = 4$) or 30-GBd PM-256QAM ($n_{bps} = 16$). Figure 4.52 shows the theoretical back-to-back performance of high-order modulation formats, which highlights the fact that an increase in the n_{bps} introduces a higher requirement in terms of SNR, which in turn results in a significantly reduced achievable transmission distance [4.123].

Figure 4.53 shows the values of symbol rate and number of bits per symbol needed to achieve a target net transmission speed of 100, 200, 400 Gb/s, or 1 Tb/s, assuming that an FEC with 20% overhead is used. The curves in Fig. 4.53 highlight the fact that serial interface rates well beyond those provided by state-of-the-art components (such as electro-optic (E/O) converters, O/E converters, DACs, and ADCs) are needed to achieve transmission speeds beyond 200 Gb/s. This electronic bottleneck can be efficiently circumvented through the use of superchannels, which exploit optical parallelism to provide high per-channel data rates and better spectral utilization [4.124]. A superchannel is a collection of optical signals that are modulated and multiplexed together with high spectral efficiency (SE) at a common originating site, transmitted and routed over a common optical link as a single entity and received at a common destination site. Typically, the superchannels use spectrally efficient advanced modulation formats in combination with advanced multiplexing schemes such as orthogonal frequency-division multiplexing (OFDM) [4.125] or Nyquist-WDM [4.123]. In [4.126] it was shown that both techniques have potentially the same performance, but Nyquist-WDM is much more robust to practical implementation RX constraints, such as limited analog bandwidth and limited ADC speed. In the following, we focus on the use of the Nyquist-WDM technique, briefly describing its generation and detection processes.

Generation of a Nyquist-WDM Signal

Nyquist-WDM signal generation and transmission have been demonstrated using either optical filters or DSP-based electrical filters [4.128–130]. Compared to the optical filter solution, the DSP-based approach is more flexible as the same hardware can be used to generate different filter shapes and modulation formats [4.131, 132]. In practice, the system cost is limited by the complexity, speed, and power consumption. Therefore, it is necessary to optimize the digital Nyquist design for the required signal in order to minimize complexity and speed [4.133–135].

In order to satisfy the Nyquist criterion for the absence of ISI among adjacent pulses [4.9], a raised-cosine (RC) spectral shape is typically used. The RC

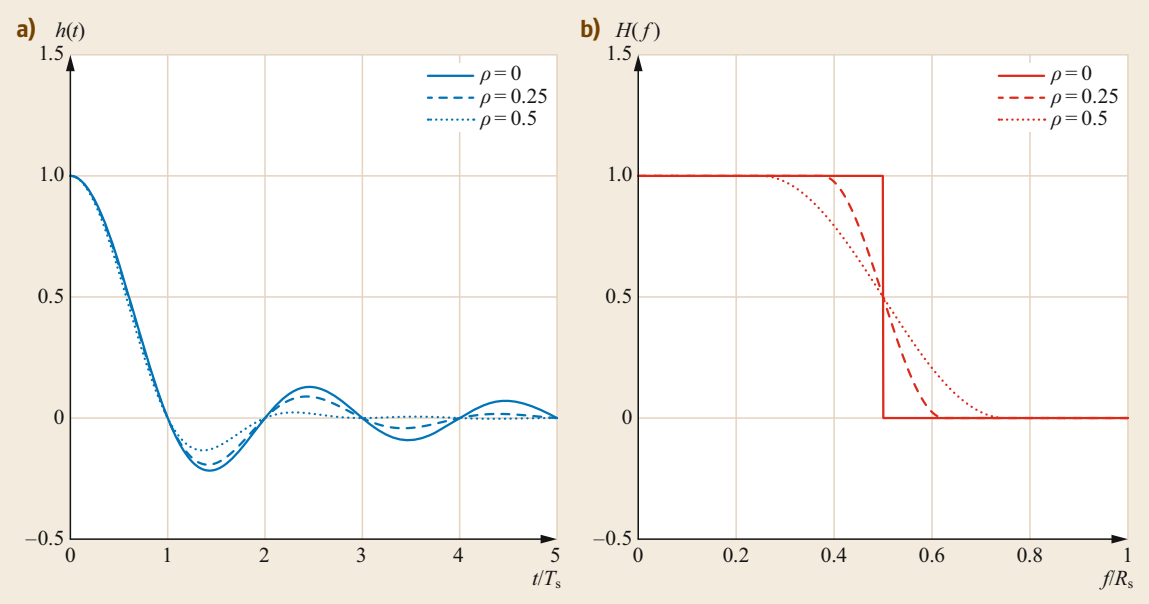


Fig. 4.54a,b RC pulses in the time domain **(a)** and frequency domain **(b)**

transfer function is defined as follows

$$H_{\mathrm{RC}}(f) = \begin{cases} T_\mathrm{s}, & \text{if } |f| \le \frac{1-\rho}{2T_\mathrm{s}} \\ \frac{T_\mathrm{s}}{2}\left\{1-\cos\left[\frac{\pi T_\mathrm{s}}{\rho}\left(f-\frac{1+\rho}{2T_\mathrm{s}}\right)\right]\right\}, & \text{if } \frac{1-\rho}{2T_\mathrm{s}} \le |f| \le \frac{1+\rho}{2T_\mathrm{s}} \\ 0, & \text{if } |f| \ge \frac{1+\rho}{2T_\mathrm{s}} \end{cases}, \tag{4.43}$$

where T_s is the symbol time (reciprocal of the symbol rate) and ρ is the *roll-off* factor, that can take values in the range $[0, 1]$. The impulse response corresponding to an RC spectrum is

$$h_{\mathrm{RC}}(t) = \frac{\sin(\pi t/T_\mathrm{s})}{\pi t/T_\mathrm{s}} \frac{\cos(\rho\pi t/T_\mathrm{s})}{1-(2\pi t/T_\mathrm{s})^2}. \tag{4.44}$$

In the case in which the channel has a flat frequency response, the optimum receiver filter is matched to the transmit one and both assume a square-root raised cosine (SRRC) shape: $H(f) = \sqrt{H_{\mathrm{RC}}(f)}$. Figure 4.54 shows examples of RC pulses in the time and frequency domain, for different values of ρ.

In Nyquist-WDM transmission with RC or SRRC spectra and frequency spacing equal to Δf, ideally no cross-talk is present between adjacent WDM channels if $\Delta f > (1+\rho)R_\mathrm{s}$. As an example, if $\rho = 0.1$, the penalty due to linear cross-talk is negligible if the channel spacing is larger than $1.1R_\mathrm{s}$, i.e., if the normalized spacing $\Delta f/R_\mathrm{s}$ exceeds 1.1. For lower values of channel spacing, a cross-talk penalty is present, as shown in Fig. 4.55 for three different modulation formats (QPSK, 16QAM, and 64QAM). If $\Delta f/R_\mathrm{s} > 1.1$, the performance of all three formats converges to the theoretical value, whilst for lower spacing values an SNR penalty is present, which is higher for higher cardinality modulation formats.

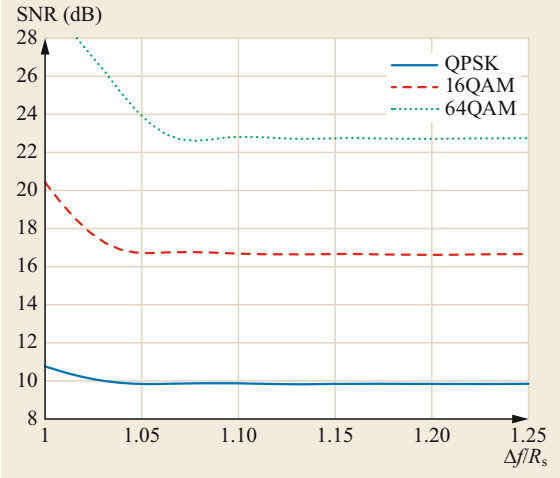


Fig. 4.55 The SNR needed to achieve a target bit error rate of 10^{-3} as a function of the normalized frequency spacing $\Delta f/R_\mathrm{s}$. SRRC filters with roll-off equal to 0.1 when used both at the TX and RX sides [4.137]

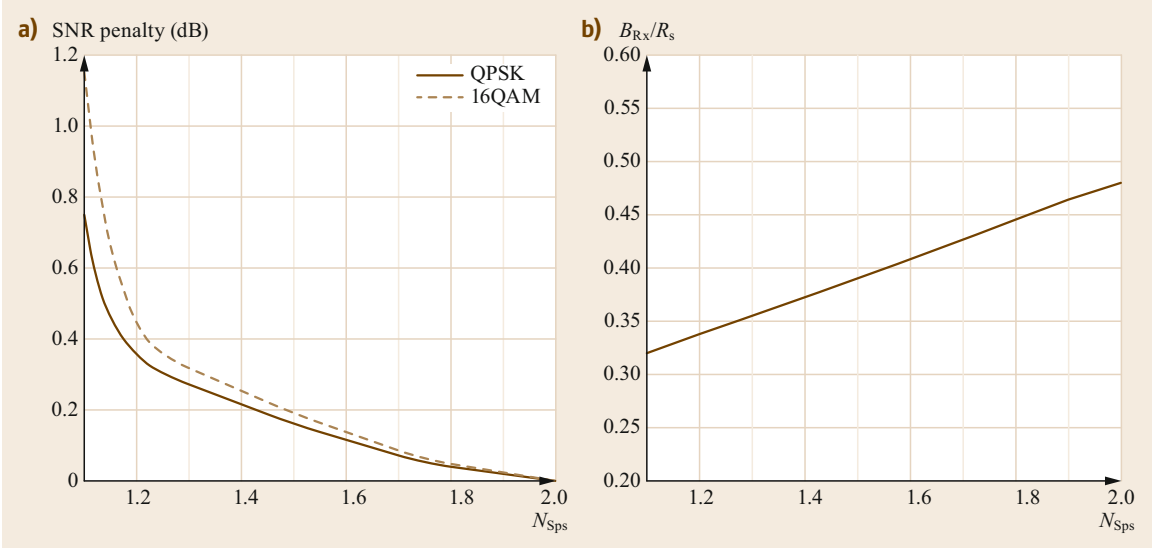


Fig. 4.56a,b The SNR penalty **(a)** and optimum RX bandwidth **(b)** as a function of the number of samples per symbol at the output of the ADC. The reference BER is 10^{-3}

Typically, Nyquist pulse shaping is performed in the digital domain using FIR filters. The required length of the digital filters that generate SRRC or RC pulses depends on the roll-off factor: For low values of ρ, the length of the pulse in time domain increases (as shown in Fig. 4.54a), thus a larger number of FIR filter taps is needed to properly approximate it. Complexity and performance of digital pulse shaping was investigated in [4.138], showing that the use of FIR filters with 17 taps allows for a reduction in channel spacing to 1.1 times the symbol rate with a 1-dB penalty. If the channel spacing is tighter, a higher number of FIR filter taps are needed, as shown in [4.139], where a 32-tap FIR filter was used to shape 14-GBd PM-16QAM Nyquist-like pulses with frequency spacing equal to $1.06R_s$ without substantial cross-talk penalty.

The achievable symbol rate and cardinality of the modulation format are limited by the DAC characteristics, as discussed in Sect. 4.5.1.

Detection of a Nyquist-WDM Signal

Nyquist-WDM signals can be efficiently detected using the standard coherent optical receiver described in Sect. 4.5.2. The single WDM channel is selected by tuning the local oscillator (LO) to its center frequency and the sharp filter shapes needed to cut out the adjacent channels are efficiently implemented in the digital domain. This enables the detection of single WDM channels without the need of any tight optical filtering at the RX. However, the use of antialiasing electrical filters before the ADCs might be beneficial, since they would reduce the signal bandwidth and thus relax the requirements for the ADC sampling frequency f_{ADC} (which, in order to satisfy the Nyquist sampling theorem, needs to be larger than twice the absolute bandwidth of the input signal [4.9]). The ratio between the ADC sampling frequency f_{ADC} and the symbol rate R_s determines the number of samples per symbol N_{SpS} at the output of the ADC. Typically, $f_{ADC} = 2R_s$, which corresponds to $N_{SpS} = 2$, but lower values can be used without incurring any substantial penalties [4.126, 137, 140].

Figure 4.56a shows an example of back-to-back SNR penalty as a function of the number of samples per symbol N_{SpS} at the output of the ADC. The reference BER was 10^{-3} and the penalty was evaluated with respect to the two samples per symbol case. The results were obtained through numerical simulations [4.137], where Nyquist-WDM signals were generated with an SRRC shape and $\rho = 0.1$. A fifth-order electrical Bessel low-pass filter with 3-dB bandwidth B_{RX} was used at the RX. The value of B_{RX} was optimized for each value of N_{SpS}, obtaining the optimum values as shown in Fig. 4.56b (normalized with respect to the symbol rate). Since a low value of N_{SpS} corresponds to a low value of f_{ADC}/R_s, the RX bandwidth has to be decreased accordingly in order to avoid aliasing. This, however, causes a distortion on the useful signal, which in turn induces an SNR penalty, which is greater for higher order modulation formats, as shown in Fig. 4.56a.

In [4.133] it was shown that the complexity of the blind equalizer can be significantly reduced by incorporating the matched SRRC filter in the bulk CD equalizer. It was also pointed out that Nyquist-WDM systems with matched filtering can be extremely sensi-

tive to the frequency offset between the transmitter laser and the local oscillator, and that the induced penalty is higher for low values of the SRRC roll-off factor.

Another critical point of Nyquist-WDM signals that has to be taken into account is related to the difficulty in performing clock recovery, i.e., in extracting the timing information from the incoming signal so that the frequency and phase of the local clock is adjusted, and the transmitted symbols may be correctly recovered. In fact, modern optical transceivers have demanding jitter tolerance and generation requirements [4.141], and must operate under high noise-loading conditions with signals that are distorted by GVD, PMD, optical filtering, and nonlinear phase noise. Whilst Nyquist pulse shaping does not result in ISI in the center of the eye, significant horizontal eye closure is introduced, presenting further difficulties for clock recovery and requiring the use of advanced DSP algorithms to minimize the penalty, as discussed in [4.142].

4.5.4 Flexible Transceivers

The key design parameters to optimally exploit all available resources in long-haul and high-capacity optical transmission systems with coherent detection are the spectral efficiency (SE) and the maximum transmission reach. Nonbinary modulation with coherent detection maximizes SE and improves tolerance to transmission impairments, while enabling effective, low-complexity electrical compensation of these impairments [4.143]. SE can be increased by minimizing the frequency spacing between WDM channels (e.g., using the *Nyquist-WDM* technique), using FEC codes with low overhead or increasing the modulation cardinality. In turn, the maximum transmission reach can be increased by using energy-efficient modulation formats (e.g., low-cardinality QAM) which reduce the SNR requirements at the receiver, using FEC codes with high overhead to ensure reliable transmission or applying proper techniques to mitigate the nonlinear propagation effects in the optical fiber. In practice, a compromise has to be made between SE and achievable reach. For example, a transceiver that operates on a short network segment with high SNR can achieve a high SE to maximize the net data rate. Similarly, a transceiver operating on a long network segment with low SNR should use either a lower order modulation format or an FEC code with high overhead to ensure reliable transmission, at the expense of a loss in SE.

Since the huge amount of traffic foreseen for the near future cannot be efficiently supported by static increases in network capacity, optical transport networks are currently migrating from a static configuration with little flexibility to the concept of software-defined optical networks [4.144]. In the scenario of next-generation optical networks, a key role will be played by flexible and scalable optical transceivers, able to dynamically adapt the modulation format and the transmission rate to the network conditions [4.145].

For flexibility, today's coherent optical transceivers typically use a set of different modulation formats based on standard QAM constellations, generated through the combination of two PAM formats in the in-phase and quadrature components. For these standard QAM formats, a gap to Shannon capacity remains, which can be expressed in terms of a penalty in SNR. Also, the standard QAM formats only offer a coarse granularity in spectral efficiency and hence also a coarse granularity in the achievable transmission reach.

To overcome the lack in granularity, several techniques are currently under study, which take advantage of the high flexibility provided by commercially available high-speed digital-to-analog converters (DACs), allowing the transmitted signal in the digital/electrical domain to be fully designed. This means that, using the same hardware, different modulation formats can be generated and different transmission rates can be achieved. Examples of techniques used to increase the flexibility of optical transceivers are: subcarrier multiplexing (SCM), implemented as either Nyquist frequency-division multiplexing (Nyquist-FDM) [4.146] or orthogonal frequency-division multiplexing (OFDM) [4.147], time-domain hybrid formats (TDHF) [4.148], and probabilistic shaping (PS) [4.149]. The great advantage of all these methods is that the flexibility is achieved without requiring any additional hardware, thus allowing currently available single-carrier transceivers to be reused by simply reprogramming the transmitted signal. The general schematics of the TX/RX pair that can be used to generate SCM, TDHF, or PS signals, as well as standard QAM signals, is shown in Fig. 4.57: the same hardware is used in all cases, the difference is in the DSP algorithms implemented at the TX and RX side. In the following, the above-mentioned techniques are briefly reviewed.

Orthogonal Frequency-Division Multiplexing (OFDM)

Orthogonal frequency-division multiplexing (OFDM) has been widely researched in optical communications due to its high spectral efficiency, resistance to chromatic dispersion, simplicity of one-tap equalization, and flexibility in modulation format [4.147, 150, 151]. This technique is based on the use of electrical subcarriers, each carrying symbols of a QAM constellation, with a sinc-function spectrum to achieve subchannel orthogonality.

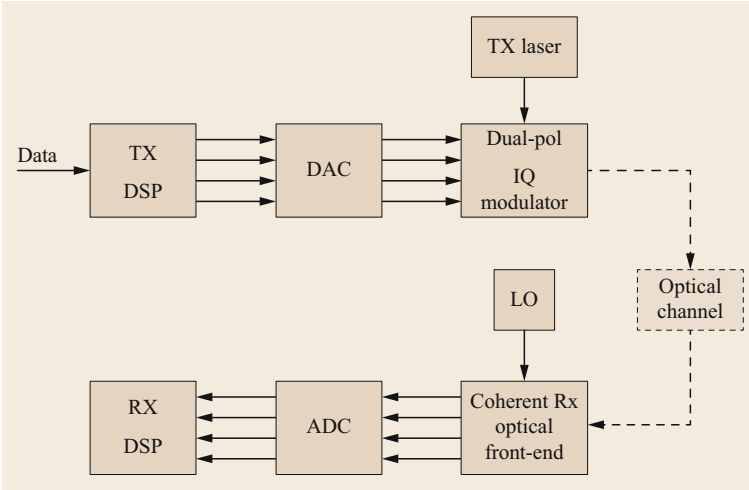


Fig. 4.57 Block diagram of a DSP-based optical TX/RX pair that can be used to generate/detect signals employing SCM, TDHF, or PS techniques, as well as standard single-carrier QAM modulation formats (DAC: digital-to-analog converter; ADC: analog-to-digital converter; DSP: digital signal processing; LO: local oscillator)

Recently, the real-valued OFDM signal (i.e., discrete multitone, DMT) has been the subject of intense investigations for intensity-modulation and direct-detection (IM/DD) optical systems, as an alternative to multilevel PAM formats for the optical interconnects such as passive optical networks, data centers, and system-on-chip interconnects [4.152–154]. However, OFDM signals show relatively high peak-to-average power ratio (PAPR) and high computational complexity, which hinder its practical application in the cost-sensitive and power-sensitive optical interconnects [4.155].

Nyquist-FDM

Like OFDM, this technique consists in electrically decomposing a high symbol-rate signal into a given number of electrical subcarriers, each of which operates at a lower symbol rate [4.156–158]. In Nyquist-FDM signals the cross-talk between adjacent subcarriers is avoided by using Nyquist shaping on the electrical subcarriers. Depending on the optical system being tested, several degrees of freedom can be optimized, such as the number of subcarrier components, the intersubcarrier spacing and the associated modulation format and pulse shaping. The granularity of the subcarriers allows control of the subcarrier symbol duration by keeping a constant overall bandwidth. This ability of optimizing the symbol duration allows the signal for nonlinearity tolerance to be adapted and it can therefore be used to extend the reach of optical transport systems [4.146, 159]. Furthermore, Nyquist-FDM allows the flexible adaption of the bandwidth [4.160] and modulation format [4.161] of each individual subcarrier. This allows for a flexible design of the spectral efficiency and therefore a finer granularity in transmission reach. A disadvantage of the Nyquist-FDM scheme is its larger sensitivity to transmitter impairments, like the IQ-skew which needs a more precise calibration or additional calibration algorithms [4.136, 162].

Time-Domain Hybrid Formats (TDHF)

The TDHF technique consists in transmitting different modulation formats interleaved in the time domain. By changing the ratio of symbols carrying the different modulation formats, the spectral efficiency can be adjusted in fine steps [4.163]. In this way, a high degree of flexibility is achieved, since the transmission rate and, consequently, the SNR performance can be tuned by simply modifying the pattern of the QAM formats. However, even if the composition of hybrid formats is optimized, the SNR performance is still far from that of the optimum Gaussian constellation [4.164–166].

Constellation Shaping (PS)

To overcome the gap to Shannon capacity of QAM constellations, modulation formats that have a Gaussian-like shape can be used [4.167–169]. Two main approaches have been proposed to achieve a Gaussian distribution, i.e., *geometrical shaping* (GS) and *probabilistic shaping* (PS). The first approach (GS) consists in designing nonsquare constellations with a Gaussian-like distribution of points and equal transmission probability for all symbols. The second approach (PS) is based on the use of a standard QAM modulation with a nonuniform distribution of the occurrence probability of the symbols of a given constellation. PS can be implemented either using modified FEC codes [4.169, 170] or using a distribution matcher [4.149]. The main advantages of PS over GS in terms of flexibility are that the net data rate can be adjusted with arbitrary granular-

ity and that the same DSP unit can be used to equalize the signal at different spectral efficiencies. In addition, a sensitivity gain on the order of 1 dB can be achieved with respect to standard QAM constellations [4.149, 171, 172], whilst a more limited gain can be obtained with GS [4.173].

4.6 Transponder Architectures

In the previous subchapters, the concept of a transponder has been introduced and different line interface variants have been presented. In what follows, the transponder concept will be generalized in Sect. 4.6.1. An explanation of which transponder types exist and what their hardware architectures look like will be given in Sect. 4.6.2. In Sect. 4.6.3, the relevant interface standards will be listed and the pluggable optical transceiver modules used in modern transponder implementations will be discussed.

Transponders are realized on printed circuit boards (PCBs) and are therefore often called *transponder cards*. They fit into monolithic or modular WDM transport platforms (Fig. 4.58) and may comprise multiple ports to improve equipment density and to provide additional functionality. Monolithic WDM equipment commonly uses a 1 HU (height unit) chassis which is sometimes referred to as a *pizza box* due to its apparent similarity (1 HU equals 1.75 in or 44.45 mm). Multiples of these shelves can be combined in a rack and stack fashion if more capacity is needed at a particular location. A modular chassis provides multiple slots (4–64 being typical amounts) into which transponders and other cards (for instance amplifier, filter, and ROADM cards) can be inserted. Placed into a 19 (or 21)-in rack with 300- or 600-mm depth, the height of a particular node can vary between 1 and 45 HU depending on the slot configuration and the amount of optical cards required. Modular shelves are typically used for large node configurations. Common equipment functions such as power supplies, fans, management and control, alarm contacts, and system timing can be shared between all cards in the same shelf. An optional high-speed backplane may provide data plane interconnections between cards. This allows the pairing of card slots and the splitting of transponder functions across multiple cards, for example into cards containing only network ports (*line cards*) and cards containing only client ports (*tributary cards*). An electronic switch fabric, typically distributed across two or more cards for redundancy reasons, can turn a transponder shelf into an electronic cross-connect, which can switch at the wavelength, OTN, and/or SDH/SONET traffic levels. Adding packet processing functions converts such equipment into a packet optical transport platform (POTP). Traffic processing at layer 2 or 3 only

Fig. 4.58 WDM transport equipment comprising different shelf variants and different transponder cards (courtesy of ADVA Optical Networking)

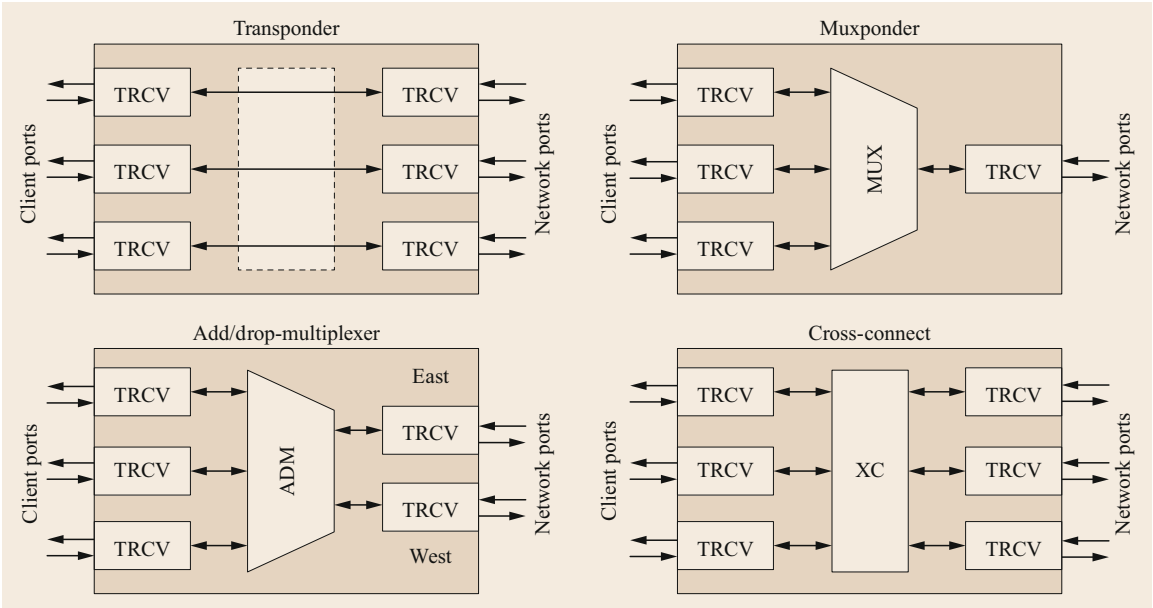


Fig. 4.59 Basic card types and operational modes (TRCV: transceiver, MUX: multiplexer, ADM: add/drop multiplexer, XC: cross-connect)

effectively turns the platform into a packet switch or router with integrated colored interfaces.

4.6.1 Transponder Types

Generalizing the transponder definition from Sect. 4.1, four basic card types can be distinguished: a transponder, a muxponder, an add-drop multiplexer, and a cross-connect. Line cards can be made configurable to support more than one of these functionalities. Figure 4.59 illustrates the basic card types.

A *transponder* is characterized by a 1-to-1 relationship between client and network ports. There are two basic modes of operation: transparent and mapped, both will be described below.

In *transparent* mode, the transponder performs a 2R (reamplifying and reshaping) or 3R (with additional retiming) regeneration between client and network ports without any intermediate protocol processing. Such a configuration can be attractive because of its low complexity and latency, but only allows limited signal monitoring and conditioning. If the same optical interface types are employed on the network and client side, the transponder acts as a simple regenerator. If different optical interface types are used on both sides, the transponder performs a media conversion and can serve as a demarcation point between different network or equipment domains. A switching function between client and network ports – if available – can be used for signal restoration: in the case of failures on the network side, client ports can be switched to an alternative network port. Depending on the switching capabilities and the number of supported ports then $1+1$, $1:1$ or $1:N$ protection schemes can be supported. If client and line interfaces are realized with pluggable transceivers, only those transceivers need to be equipped which are required for the desired functionality.

In *mapped* mode, the transponder then also performs a protocol conversion between the client and line signals. Multiple services such as local area network (LAN), storage area network (SAN), high-performance computing (HPC), video distribution, passive optical networking, synchronous digital hierarchy (SDH)/synchronous optical network (SONET), optical transport network (OTN) and generic constant bit rate (CBR) or packetized signals on the client side can be mapped into a transport signal on the line side. Please see Chap. 13 for a summary of the most popularly used client signals. While SDH and SONET have been used since the late 1980s, OTN was introduced in 2001 and has since then established itself as the common transport protocol on the network side. Adding a protocol mapper to a transponder allows the monitoring of client and line signals which eases operations, administration, and maintenance (OAM) of the signal and offers increased performance by optional forward error correction (FEC). It can also provide fault and performance management capabilities as well.

A *muxponder* is a combination of a multiplexer and a transponder. It is characterized by an N-to-1 re-

lationship between client and network ports. Several client signals are multiplexed together to obtain a higher speed line signal. 10G and 100G muxponders are the most commonly used today (note that the letters M and G are used to denote the data rate units Mb/s and Gb/s, respectively). The client ports can support different service types and data rates and may require specific mapping processes depending on the selected service. OAM functions allow the integrity of the service to be monitored and to take the necessary actions in case of faults or performance issues. As client ports generally run on independent clocks, an adaptation to the common clock of the line signal is required (for example by bit or byte stuffing). Multiplexing can be done in one or more stages and based on SDH/SONET, OTN, or other technologies. Aggregation of services at a packet level is also possible if the muxponder is equipped with packet processing functions.

An *add/drop multiplexer* is an extension of the muxponder where a second network port is available to enable the support of ring or protected point-to-point applications. The two network ports are often named *east* and *west* ports indicating that they are connected to two diverse fiber routes or ring directions. An add/drop multiplexer can pass traffic between the network ports (called pass-through traffic) and can terminate a portion of the traffic locally (called add/drop traffic) at each network port. A predefined mapping and multiplexing structure as well as a time slot approach is required to be able to isolate the client signals contained within the aggregate line signals. Client signals can be protected against link failures by switching the respective traffic from one to the other network port. In many cases, an add/drop multiplexer can also be operated as two independent muxponders, provided enough client ports are available to match the capacity of the line signal. While add/drop-multiplexers commonly operate on layer 1, adding and dropping of client signals can also be performed at layer 2 or 3, thereby providing oversubscription and statistical multiplexing capabilities.

A *cross-connect* allows a flexible switching between network and client ports, including hair-pinning for client-to-client connections. Multiple colored network ports allows multidegree network nodes to be supported, switching signals between more than two fiber directions. Client ports provide local add/drop access. Depending on the card architecture, the distinction between network and client port can be removed, and any port can serve as network or client port. Switching can be performed on multiple levels and is not restricted to simple port-to-port connectivity only. Higher order SDH and SONET cross-connects operate on a VC-4 (155M) and STS-1 (51M) granularity, respectively. OTN cross-connects can switch on ODU*k* (where if $k = 0$ then the data rate is 1.25G, if $k = 1$ 2.5G, if $k = 2/2e$ 10G, if $k = 3$ 40G, and if $k = 4$ 100G) and ODUflex (supporting multiples of 1.25G) levels. They can operate with tributary slot granularities of 1.25G, 2.5G, or 5G. If a cross-connect supports switching at the packet level, packet flows at Mb/s or lower granularities can also be supported. In principle, SDH/SONET, OTN and packet traffic require dedicated switching matrices. As the size of these matrices needs to be determined upfront, this either limits the flexibility leading to blocking of signals or requires an overprovisioning of switch capacity. A universal switch matrix can alleviate these disadvantages and is of particular interest if cross-connect functions are split across multiple cards [4.174]. In a packet-based universal switch, signals are segmented into packets at ingress, passed over the fabric, and then reassembled at egress. The functional unit performing the packetization is often called a segmentation and reassembly (SAR) block. Typical packet lengths are 128, 256, or 512 byte. However, care needs to be exercised so that client timing (phase and frequency) is maintained and latency as well as latency variations across the fabric are minimized. This can be achieved by variable length packets in which the packet size is slightly varied over time, for instance by lengthening or shortening the packet size by one byte [4.175]. Small cross-connects can be realized on a single line card, whereas large cross-connects can occupy a modular equipment shelf or even require multishelf configurations scaling up to tens of Tb/s of switching capacity.

4.6.2 The Anatomy of a Line Card

WDM platforms often use a generic line card approach. A common base design provides the functions necessary for the operation of any card. Customization is then performed to accommodate specific optical and electronic components for realization, for instance a transponder, an amplifier, or a wavelength-selective switch card. In multislot systems, a backplane is used to connect the line cards with the host platform. It provides power and control to the line cards and may include high-speed data plane connections between line cards or to a central switch fabric.

Common card functions include DC/DC converters, control logic, and the card processor. WDM platforms run off a −48 or +12 V DC (often redundant) power supply which feeds the line cards via the backplane. DC/DC converters on the card then derive all the necessary supply voltages from the base voltage. An onboard microprocessor connects the card to the shelf or network element controller; it is responsible for controlling and managing all card functions as well

as collecting associated fault and performance data. The microprocessor system may contain electronically erasable programmable read-only memory (EEPROM) and flash memory for storing the card calibration data and the card firmware, respectively. Often, the microprocessor is accompanied by additional control logic such as field-programmable gate arrays (FPGAs) which are used for more elaborate control tasks while complex programmable logic devices (CPLDs) are often employed whenever simple logic operations need to be performed in a hard-wired manner (for example laser safety actions).

Figure 4.60 shows the block diagram of a simple OTN transponder line card. In the text, the signal flow only in one direction will be described for clarity, while the other direction passes through the same functional blocks in the reverse order.

A gray transceiver on the client port receives the incoming optical signal, performs an optical to electrical conversion, and passes the signal on to an OTN mapper device. At lower speeds, the electrical interface between the transceiver and the mapper comprises only a single digital lane per direction. At higher speeds, the interface can be implemented over multiple parallel lanes. It then requires a multilane distribution (MLD) algorithm to distribute the transceiver signal over the lanes at one side and to reassemble it at the other side, taking into account the fact that delay differences between the lanes need to be compensated for by appropriate skew management. If enabled, client-side FEC is applied before the signal is passed onto the next step. Often, a non-intrusive monitoring of the client signal is provided to ensure that the incoming signal is of good quality and that the client side network connection is working properly. If the client signal is not already an OTN signal, the client mapper then maps the received signal into an optical payload unit (OPU). The mapping is client specific and can be performed employing a bit-synchronous (BMP), asynchronous (AMP), generic (GMP), or idle (IMP) mapping procedure [4.176]. With BMP, the line clock is phase-locked to the client signal. AMP, GMP, and IMP typically operate with a line clock that is independent from the client clock and therefore require stuffing of bits, bytes, or control characters to compensate for clock rate differences between the line and client clocks. Since its introduction in 2009, GMP has become the preferred mode for all new mappings due to its flexibility and capability to compensate large clock tolerances. Constant bit rate (CBR) signals can directly be mapped into OPU, while packetized signals must first undergo an adaptation step using the generic framing procedure (GFP) [4.177] or IMP. Timing transparent transcoding (TTT) [4.176] may be used as an intermediate step for CBR signals to reduce the data rate of client signals

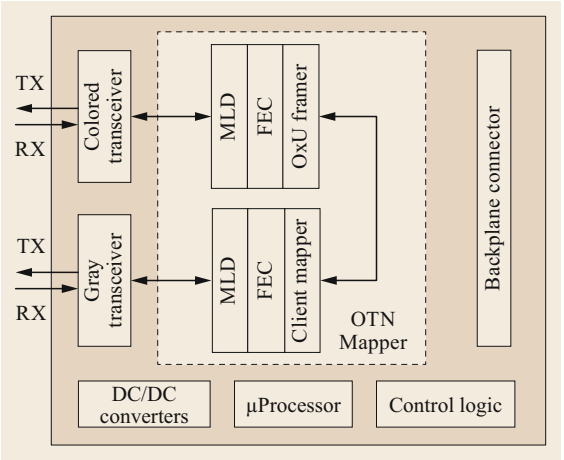


Fig. 4.60 Block diagram of a simple OTN transponder line card

by compressing the employed line coding overhead. In the OxU (OTU/ODU/OPU) framer, the OPU signal is mapped into an optical data unit (ODU) and the ODU is mapped into an optical transport unit (OTU). After adding FEC information, the signal is then sent to the colored transceiver on the line side (using MLD in the case of multiple electrical lanes) where the electrical signal is converted into an optical DWDM signal.

The client- and line-side transceivers may be realized as pluggable modules or discrete implementations on the transponder line card. They may contain clock-and-data recovery as well as basic equalization capabilities to compensate for the PCB trace losses between the transceivers and the mapper device. If the number of optical lanes and electrical lanes is different, they may also incorporate a multiplexer or gearbox IC. For higher data rates and longer reach applications, an enhanced FEC function may also be included. Coherent transceivers typically use a dedicated digital signal processor (DSP). In this case, some OTN mapping and multiplexing functions may be incorporated into the DSP, eliminating the need for a separate OTN mapper device in certain applications.

Extending the functions available to the transponder card described above, Fig. 4.61 shows a universal OTN line card which facilitates access to CBR, packet (for example Ethernet or IP/MPLS flows), TDM traffic (SDH/SONET) and OTN traffic from each of the optical interfaces. The interfaces may be implemented using pluggable transceivers. At data rates up to 10G, colored (line side) and gray (client side) transceivers may use the same pluggable module standard. If the protocol processing capabilities are then the same for each port, a flexible allocation of ports to network- and client-side traffic is possible. The OTN mapper device offers

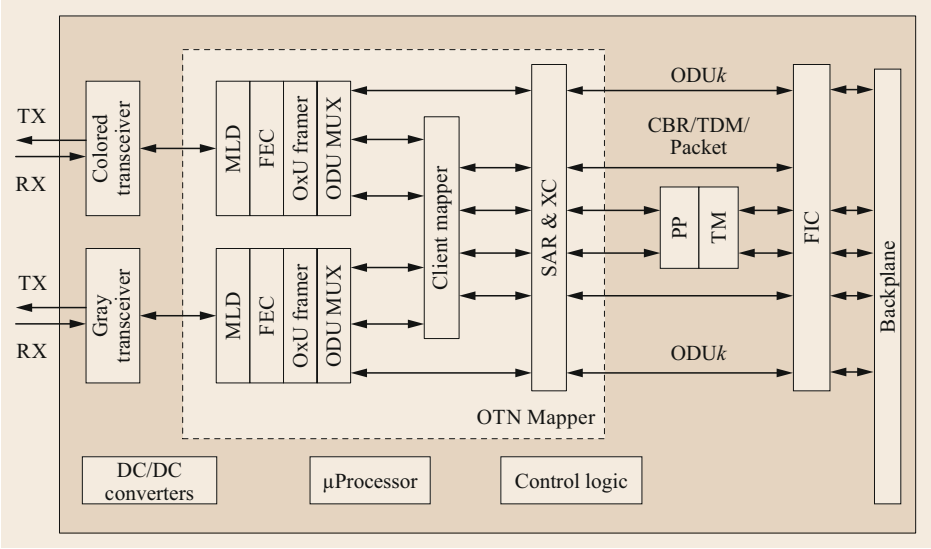


Fig. 4.61 Block diagram of a universal OTN line card (FEC: forward error correction, MLD: multi-lane distribution, TX: transmitter, OxU: OTU/ODU/ OPU, RX: receiver)

configurable protocol processing blocks which can either pass the frames through sequentially or make them accessible via device internal switches which allow individual signal paths for each port to be programmed. Assuming that each of the blocks can be bypassed if it is not required, Fig. 4.61 depicts a sequential arrangement of processing blocks. Incoming OTN signals from the transceivers are passed through the MLD and FEC blocks and then enter the OxU framer, where they are demapped and forwarded to the ODU multiplexer. If the OTN signal comprises a multiplex of lower order ODU*k* signals, then a demultiplexing is performed to access these lower order signals. While an OTN signal in principle may be multiplexed over multiple stages (for example ODU0 to ODU1 to ODU2 to ODU3 to ODU4), practical considerations and hardware constraints often limit the MUX block inside the mapper device to two multiplexing and several mapping stages which can be arbitrarily selected. At this point, multiple processing paths are possible: ODU*k* signals can directly be cross-connected (XC) to other ports of the same mapper device or passed on to a centralized switch matrix. An SAR device packetizes the OTN signals which are delivered to a universal packet-based switch fabric via a fabric interface chip (FIC). Typically, an Interlaken interface is used between the OTN mapper and the FIC. For smaller switch configurations, a direct interconnection of OTN mapper devices between paired cards is also sometimes possible, alleviating the need for an FIC in between. FEC and time stamping may be provided to improve the transmission and timing performance over the fabric. If CBR, packet, or TDM client signals inside an ODU*k* are to be accessed, the signals are passed on to a client mapper which extracts the traffic and cross-connects it either to another transceiver port or to the universal switch matrix via an SAR and an FIC. For advanced processing of packet traffic, an external packet processor (PP) and traffic manager (TM) is often used before the traffic is passed on to the switch matrix via the FIC. Packet traffic does not need to pass through the SAR block, as it is supported natively by the packet fabric. CBR, TDM, or packet traffic entering from a gray transceiver may pass the MLD and FEC blocks and is then directly processed in the client mapper, bypassing the OxU framer and ODU MUX stages. It can then be switched to another transponder port as in the case of the simple transponder of Fig. 4.60, directly fed to the switch fabric, or wrapped into an ODU structure first and then passed through the ODU MUX block.

4.6.3 Interface Standards

Standards are necessary to facilitate a line card design based on interoperable building blocks which can be procured from multiple sources. The most important areas of standardization include transceiver modules as well as their respective optical, electrical, and control interfaces. While transceivers historically were implemented using discrete optics and electronics, multisource agreements (MSA) for transceivers emerged in the late 1990s and allowed the use of integrated transceiver modules instead.

Hot pluggable, front-plate accessible transceiver modules were introduced starting with the gigabit interface converter (GBIC) in 1995, followed by the 10 Gb Ethernet transceiver package (XENPAK) and the small form-factor pluggable (SFP) transceiver in 2001, the 10 Gb small form-factor pluggable (XFP)

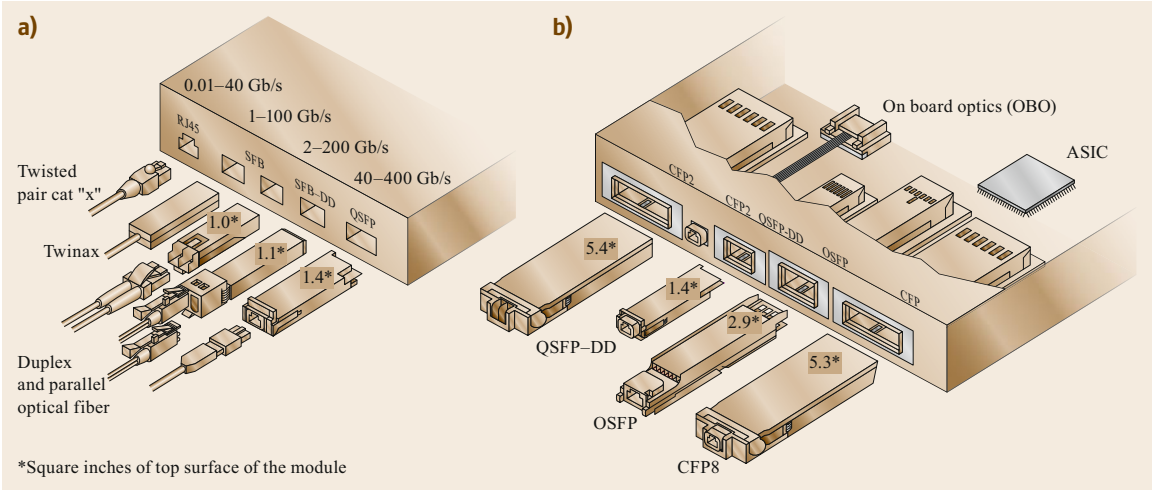


Fig. 4.62a,b Pluggable module form factors: **(a)** current modules, **(b)** modules in development for 400G (Graphics courtesy of Ethernet Alliance [4.178])

module in 2002, the quad small form-factor pluggable (QSFP) transceiver and the $1 \times 10\,\mathrm{Gb/s}$ small form-factor pluggable (SFP+) transceiver in 2006, as well as the $10\,\mathrm{Gb/s}$ 4 × pluggable transceiver (QSFP+) in 2009. SFP+ and QSFP+ modules are available in different speed classes. To distinguish them, the nominal electrical lane speed is often appended instead of the + symbol (for example SFP10 and QSFP10). The MSAs define the mechanics, the electrical connector, the power supply as well as the high- and low-speed signaling along with implementation and measurement descriptions. The specifications are maintained by the Storage Networking Industry Association (SNIA) small form-factor (SFF) Technology Affiliate Technical Work Group [4.179]. New developments include the QSFP double density (QSFP-DD) [4.180] and the octal SFP (OSFP) [4.181] targeting 400G and the SFP double density (SFP-DD) [4.182] targeting 100G applications.

Also the 40G/100G form-factor (CFP) MSA group [4.183] has developed specifications for 100G+ transceivers offering more space for component integration and a higher power dissipation than their SFP and QSFP counterparts. Since formation of the group in 2009, three generations of 100G small form-factor pluggable (CFP) modules (CFP, CFP2 and CFP4) were brought to market, while the CFP8 specification has just been released in 2017.

The most common front pluggable optical modules are illustrated in Fig. 4.62 with a twisted pair RJ45 copper cable connector for comparison, along with new module standards in development for 400G and beyond. QSFP-DD is by many the favored 400G interface standard, although it is the most challenging from an integration and thermal management point of view, as it possesses a smaller module surface area compared to the OSFP and CFP8.

In addition to front pluggable modules, MSAs have been drawn up for PCB mountable optical transceivers (on-board optics – OBO). This development started with 300-pin transceivers for 10G applications in a 5×7-in footprint which were first released in 2001. Subsequently, a smaller form-factor variant (2.2×3 in) and 40G support were added and the first 40G MSA modules became available in 2006. While the primary design target was colored WDM applications, MSA modules for gray interface applications were also offered. For 100G coherent long-haul applications, the OIF published an implementation agreement for a 5×7-in 168-pin transmission module in 2011 [4.184], which was augmented in 2015 by a second release targeting a smaller module footprint (4×5 in). Advances in CMOS technology and photonic integration since then have allowed integrating digital coherent 100G/200G transceivers also into CFP and CFP2 front pluggable modules (digital coherent optics or DCO modules). As an intermediate step to the CFP2-DCO module, a CFP2-ACO analog coherent optics transceiver was released which integrates all optics and RF electronics but not the coherent DSP. Further development directions include the definition of 400G and 800G on-board optical modules as currently being pursued by the Consortium on On-Board Optics (COBO) [4.185] as well as an integration of 400G coherent transceiver technology into next-generation optical modules (QSFP-DD, OSFP, CFP8 or coherent OBO).

Client-side optical interface standards are defined by the industry bodies responsible for the respective protocol specifications. These are Telcordia for SONET, the ITU-T Study Group 15 for SDH, OTN,

Table 4.5 Ethernet interface standards and related nomenclature (after [4.178]). Color code: *black* = IEEE standard, *blue* = non-IEEE standard but complies to IEEE electrical interfaces, *red*: under discussion in standardization

Standard	Electrical interface	MMF	Parallel SMF	2 km SMF	10 km SMF	40 km SMF
100BASE-		FX			LX	
1000BASE-		SX			LX	
10GBASE-	SFI, XFI, XSBI, XAUI	SR			LR	ER
25GBASE-	25GAUI	SR			LR	ER
40GBASE-	XLAUI	SR4	PSM4	FR	LR4	ER4
50GBASE-	LAUI-2/ 50GAUI-2 50GAUI-1	SR		FR	LR	ER
100GBASE-	CAUI-10	SR10		10 × 10	10 × 10	
	CAUI-4/	SR4	PSM4	CWDM4	LR4	ER4
	100GAUI-4	SR2	DR	CLR4	4WDM-10	4WDM-40
	100GAUI-2					
	100GAUI-1			100G-FR	100G-LR	
200GBASE-	200GAUI-4	SR4	DR4	FR4	LR4	
	200GAUI-2					ER2
400GBASE-	400GAUI-16	SR16	DR4	FR8	LR8	
	400GAUI-8					ER8
	400GAUI-4					ER4

and PON, the IEEE 802.3 group for Ethernet and Ethernet-PON, the International Committee for Information Technology Standards (Incits) Task Group 11 (T11) for Fiber Channel, the Infiniband Trade Association (IBTA) for Infiniband, the CPRI group for the Common Public Radio Interface, and the Society of Motion Picture and Television Engineers (SMPTE) for video distribution.

An overview of Ethernet interfaces over single-mode (SMF) and multimode (MMF) fiber along with their electrical host-side interfaces is shown in Table 4.5. Note that the numbers stated after the interface classifier indicate the number of parallel signaling lanes (with exception of the 4WDM interfaces where the numbers indicate the reach in km). NRZ signaling is commonly used up to 25G, and PAM4 signaling above this rate on optical and electrical interfaces.

Optical client-side interfaces for SDH/SONET and OTN signals are defined in the Telcordia GR-253 standard as well as the ITU-T G.957, G.691-693, G.695, and G.959.1 recommendations. Colored line-side interfaces are described in the ITU-T G.698 series of recommendations. Published ITU-T recommendations can be accessed online here [4.186].

The OIF Common Electrical I/O Implementation Agreement (CEI IA) defines electrical and jitter requirements for electrical chip-to-module, chip-to-chip, and backplane interfaces. Currently included are interfaces with 3G, 6G, 11G, and 28G signaling rates [4.187], and new revisions including 56G and 112G are already in development. The CEI forms the basis of the 10 to 400G attachment unit interfaces (AUI) as well as the SFP (SFI) and XFP (XFI) interfaces connecting Ethernet transceivers to the attached host board chip. For the AUI, a change of the speed classes is occurring translating from Arabic to Roman numerals (e.g., from CAUI where *C* denotes 100 to 100GAUI). Protocol-specific aspects such as MLD are defined in the IEEE 802.3 standard. The CEI IA also defines the electrical parameters for SDH/SONET and OTN module-to mapper/framer interfaces. At 10G, these interfaces typically need only a single serial electrical lane (SFI/XFI) and therefore do not require any MLD. At 40G, a group of four synchronous (STL256.4) and optical transport lanes (OTL3.4) as defined in ITU-T G.707 and G.709 are used for SDH/SONET and OTN signals, respectively. 100G OTN signals are transported over a 10 × 10G (OTL4.10) or a 4 × 25G (OTL4.4) interface. Beyond 100G module-to-framer interfaces can use multiple 100G optical transport lane (OTLC) groups or a FlexO interface (FOIC) as described in ITU-T G.Sup58 [4.188]. Future interface standards are expected to increase the signaling rate and reduce the number of lanes further.

Digital control of transceiver modules relies on either a two-wire serial (TWS) or a management data input/output (MDIO) interface as defined in IEEE 802.3. Currently, the management interface specification (MIS) forms part of module MSA documentation and is specifically developed for it. In the future, a common MIS (CMIS) may be created from which a module MIS can be derived.

References

4.1 F.A. Kingsley: *The Development of Radar Equipments for the Royal Navy, 1935–45* (Palgrave Macmillan, London 1995)

4.2 Telecommunication Standardization Sector (ITU-T): Recommendation G.691: Optical interfaces for single channel STM-64 and other SDH systems with optical amplifiers, available at https://www.itu.int/rec/T-REC-G/en

4.3 Telecommunication Standardization Sector (ITU-T): Recommendation G.694.1: Spectral grids for WDM applications: DWDM frequency grid, available at https://www.itu.int/rec/T-REC-G/en

4.4 Telecommunication Standardization Sector (ITU-T): Recommendation G.694.2: Spectral grids for WDM applications: CWDM wavelength grid, available at https://www.itu.int/rec/T-REC-G/en

4.5 C.H. Sterling: *Military Communications: From Ancient Times to the 21st Century* (ABC-CLIO, Santa Barbara 2007)

4.6 Telecommunication Standardization Sector (ITU-T): Recommendation G.698.2: Amplified multichannel dense wavelength division multiplexing applications with single channel optical interfaces, available at https://www.itu.int/rec/T-REC-G/en

4.7 M. Filer, J. Gaudette, M. Ghobadi, R. Mahajan, T. Issenhuth, B. Klinkers, J. Cox: Elastic optical networking in the Microsoft cloud, J. Opt. Commun. Netw. **8**(7), A45–A54 (2016)

4.8 R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel: Capacity limits of optical fiber networks, J. Lightwave Technol. **8**(7), 662–698 (2010)

4.9 J. Proakis: *Digital Communications* (McGraw-Hill, New York 2007)

4.10 Y. Sun, J.B. Judkins, A.K. Srivastava, L. Garrett, J.L. Zyskind, J.W. Sulhoff, C. Wolf, R.M. Derosier, A.P. Gnauck, R.W. Tkach, J. Zhou, R.P. Espindola, A.M. Vengsarkar, A.R. Chraplyvy: Transmission of 32-WDM 10-Gb/s channels over 640 km using broad-band, gain-flattened erbium-doped silica fiber amplifiers, IEEE Photonics Technol. Lett. **9**(12), 1652–1654 (1997)

4.11 N.S. Bergano, C.R. Davidson, M. Ma, A. Pilipetskii, S.G. Evangelides, H.D. Kidorf, M. Darcie, E. Golovchenko, K. Rottwitt, P.C. Corbett, R. Meges, M.A. Mills, B. Pedersen, D. Peckham, A.A. Abramov, A.M. Vengsarkar: 320 Gb/s WDM transmission (64×5 Gb/s) over 7,200 km using large mode fiber spans and chirped retum-to-zero signals. In: *Proc. Fiber Commun. Conf. (OFC'98), San Jose* (1998), Paper PD12

4.12 A.R. Pratt, P. Harper, S.B. Alleston, P. Bontemps, B. Charbonnier, W. Forysiak, L. Gleeson, D.S. Govan, G.L. Jones, D. Nesset, J.H.B. Nijhof, I.D. Phillips, M.F.C. Stephens, A.P. Walsh, T. Widdowson, N.J. Doran: 5,745 km DWDM transcontinental field trial using 10 Gbit/s dispersion managed solitons and dynamic gain equalization. In: *Proc. Opt. Fiber Commun. Conf., Atlanta* (2003), Paper PD26

4.13 I. Haxell, M. Ding, A. Akhtar, H. Wang, P. Farrugia: 52×12.3 Gb/s DWDM transmission over 3600 km of True Wave fiber with 100 km amplifier spans. In: *Proc. Opt. Amplif. Appl., Québec City* (2000), Paper OTuC7

4.14 S. Bigo, Y. Frignac, G. Charlet, W. Idler, S. Borne, H. Gross, R. Dischler, W. Poehlmann, P. Tran, C. Simonneau, D. Bayart, G. Veith, A. Jourdan, J. Hamaide: 10.2 Tbit/s (256×42.7 Gbit/s PDM/WDM) transmission over 100 km TeraLight fiber with 1.28 bit/s/Hz spectral efficiency. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2001), Paper PD25

4.15 S. Walklin, J. Conradi: Multilevel signaling for increasing the reach of 10 Gb/s lightwave systems, J. Lightwave Technol. **17**(11), 2235–2248 (1999)

4.16 A. Lender: The duobinary technique for high-speed data transmission, IEEE Trans. Commun. Technol. **82**(2), 214–218 (1963)

4.17 T. Ono, Y. Yano, K. Fukuchi, T. Ito, H. Yamazaki, M. Yamaguchi, K. Emura: Characteristics of optical duobinary signals in terabit/s capacity, high spectral efficiency WDM systems, J. Lightwave Technol. **16**(5), 788–797 (1998)

4.18 X. Gu, S.J. Dodds, L.C. Blank, D.M. Spirit, S.J. Pycock, A.D. Ellis: Duobinary technique for dispersion reduction in high capacity optical systems-modelling, experiment and field trial, IEE Proc. Optoelectron. **143**(4), 228–236 (1996)

4.19 K. Yonenaga, S. Kuwano: Dispersion tolerant optical transmission system using duobinary transmitter and binary receiver, J. Lightwave Technol. **15**(8), 1530–1537 (1997)

4.20 K. Yonenaga, A. Hirano, M. Yoneyama, Y. Miyamoto, K. Hagimoto, K. Noguchi: Expansion of tolerable dispersion range in a 40 Gbit/s optical transmission system using an optical duobinary signal, Electron. Lett. **34**(4), 385–386 (1998)

4.21 D. Penninckx, M. Chbat, L. Pierre, J.-P. Thiery: The phase-shaped binary transmission (PSBT): A new technique to transmit far beyond the chromatic dispersion limit, IEEE Photonics Technol. Lett. **9**(2), 259–261 (1997)

4.22 X. Zheng, F. Liu, P. Jeppesen: Receiver optimization for 40-Gb/s optical duobinary signal, IEEE Photonics Technol. Lett. **13**(7), 744–746 (2001)

4.23 H. Kim, C.X. Yu: Optical duobinary transmission system featuring improved receiver sensitivity and reduced optical bandwidth, IEEE Photonics Technol. Lett. **14**(8), 1205–1207 (2002)

4.24 I. Lyubomirsky, B. Pitchumani: Impact of optical filtering on duobinary transmission, IEEE Photonics Technol. Lett. **16**(8), 1969–1971 (2004)

4.25 G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini: Modulation formats suitable for ultrahigh spectral efficient WDM systems, IEEE J. Sel. Top. Quantum Electron. **10**(2), 321–328 (2004)

4.26 N.B. Pavlovic, A.V.T. Cartaxo: Optimized bandwidth limited duobinary coding format for ultra dense WDM systems. In: *Proc. Int. Conf.*

Transpar. Opt. Netw. (ICTON 2005), Barcelona (2005), https://doi.org/10.1109/icton.2005.1506180

4.27 H. Bissessur, G. Charlet, C. Simonneau, S. Borne, L. Pierre, C. De Barros, P. Tran, W. Idler, R. Dischler: 3.2 Tb/s (80×40 Gb/s) C-band transmission over 3×100 km with 0.8 bit/s/Hz efficiency. In: *Proc. 27th Eur. Conf. Opt. Commun., Amsterdam* (2001), Paper PD.M.1.11

4.28 M. Birk, C. Skolnick, R. Curto, R. Marlieb, T.J. Schmidt, R. Saunders, H. Zech, H. Schucht, J. Heinrich, C. Weiske: Field trial of a 40 Gbit/s PSBT channel upgrade to an installed 1700 km 10 Gbit/s system. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2005), Paper OTuH3

4.29 P.A. Humblet, M. Azizoglu: On the bit error rate of lightwave systems with optical amplifiers, J. Lightwave Technol. **9**(11), 1576–1582 (1991)

4.30 S.R. Chinn, D.M. Boroson, J.C. Livas: Sensitivity of optically preamplified DPSK receivers with Fabry-Perot filters, J. Lightwave Technol. **14**(9), 370–376 (1996)

4.31 G. Rayhnn, S. Chandrasekhar, J. Leuthold, C. Dnerr, L. Stulz, A. Agarwal, S. Banerjee, D. Grnsz, S. Hunsche, A. Kung, A. Marhelyuk, D. Maywar, M. Mnvassaghi, X. Liu, C. Xu, X. Wei, D.M. Gill: 2.5 Tb/s (64×42.7 Gb/s) transmission over 40×100 km NZDSF using RZ-DPSK format and all-Raman-amplified spans. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2002), Paper FC2

4.32 R.A. Griffin, A.C. Carter: Optical differential quadrature phase-shift key (oDQPSK) for high capacity optical transmission. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2002), Paper WX6

4.33 J. Wang, J.M. Kahn: Impact of chromatic and polarization-mode dispersions on DPSK systems using interferometric demodulation and direct detection, J. Lightwave Technol. **22**(2), 362–371 (2004)

4.34 A.H. Gnauck, P.J. Winzer: Optical phase-shift-keyed transmission, J. Lightwave Technol. **23**(1), 115–130 (2005)

4.35 P.J. Winzer, R.-J. Essiambre: Advanced optical modulation formats, Proc. IEEE **94**(5), 952–985 (2006)

4.36 D.A. Fishman, W.A. Thompson, L. Vallone: LambdaXtreme transport system: R&D of a high capacity system for low cost, ultra-long haul DWDM transport, Bell Labs Tech. J. **11**(2), 27–53 (2006)

4.37 A. Ohta, D. Tanimura, T. Kyakuno, S. Iio: 43-Gbps RZ-DQPSK transponder for long-haul optical transmission system, Yokogawa Tech. Rep. Engl. Ed. **46**, 3–6 (2008)

4.38 B. Mikkelsen, C. Rasmussen, P. Mamyshev, F. Liu: Partial DPSK with excellent filter tolerance and OSNR sensitivity tolerance, Electron. Lett. **42**(23), 1363–1364 (2006)

4.39 F. Derr: Coherent optical QPSK intradyne system: Concept and digital receiver realization, J. Lightwave Technol. **10**(9), 1290–1296 (1992)

4.40 M.G. Taylor: Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments, IEEE Photonics Technol. Lett. **16**(2), 674–676 (2004)

4.41 R. Noe: PLL-free synchronous QPSK polarization-multiplex/diversity receiver concept with digital I&Q baseband processing, IEEE Photonics Technol. Lett. **17**(4), 887–889 (2005)

4.42 S.J. Savory, A.D. Stewart, S. Wood, G. Gavioli, M.G. Taylor, R.I. Killey, P. Bayvel: Digital equalization of 40 Gbit/s per wavelength transmission over 2480 km of standard fibre without optical dispersion compensation. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Cannes* (2006), Paper Th2.5.5

4.43 C. Laperle, B. Villeneuve, Z. Zhang, D. McGhan, H. Sun, M. O'Sullivan: WDM performance and PMD tolerance of a coherent 40-Gbit/s dual-polarization QPSK transceiver, J. Lightwave Technol. **26**(1), 168–175 (2008)

4.44 C.R.S. Fludger, T. Duthel, D. van den Borne, C. Schulien, E.-D. Schmidt, T. Wuth, J. Geyer, E. De Man, G.-D. Khoe, H. de Waardt: Coherent equalization and POLMUX-RZDQPSK for robust 100-GE transmission, J. Lightwave Technol. **26**(1), 64–72 (2008)

4.45 S.J. Savory: Digital filters for coherent optical receivers, Opt. Express **16**(2), 804–817 (2008)

4.46 C. Laperle, B. Villeneuve, Z. Zhang, D. McGhan, H. Sun, M. O'Sullivan: Wavelength division multiplexing (WDM) and polarization mode dispersion (PMD) performance of a coherent 40Gbit/s dual-polarization quadrature phase shift keying (DP-QPSK) transceiver. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2007), Paper PDP16

4.47 P.J. Winzer: Beyond 100G Ethernet, IEEE Commun. Mag. **48**(7), 26–30 (2010)

4.48 K. Roberts, M. O'Sullivan, K.-T. Wu, H. Sun, A. Awadalla, D.J. Krause, C. Laperle: Performance of dual-polarization QPSK for optical transport systems, J. Lightwave Technol. **27**(16), 3546–3559 (2009)

4.49 R. Nagarajan, D. Lambert, M. Kato, V. Lal, G. Goldfarb, J.T. Rahn, J. McNicol, K.-T. Wu, M. Kuntz, J. Pleumeekers, A. Dentai, H.-S. Tsai, R. Malendevich, M. Missey, J. Tang, J.-M. Zhang, O. Khayam, T. Butrie, H.H. Sun, A.C. Nilsson, V. Dangui, M.L. Mitchell, M. Reffle, F. Kish, D. Welch: Five-channel, 114 Gbit/s per channel, dual carrier, dual polarisation, coherent QPSK, monolithic InP receiver photonic integrated circuit, Electron. Lett. **47**(9), 555–556 (2011)

4.50 M. Alfiad, S. Tibuleac: 100G Superchannel transmission using 4×28 Gb/s subcarriers on a 25-GHz grid, IEEE Photonics Technol. Lett. **27**(2), 157–160 (2015)

4.51 A.H. Gnauck, P.J. Winzer, S. Chandrasekhar, X. Liu, B. Zhu, D.W. Peckham: 10×224-Gb/s WDM transmission of 28-Gbaud PDM 16-QAM on a 50-GHz grid over 1,200 km of fiber. In: *Proc. Opt. Fiber Commun. (OFC), San Diego* (2010), Paper PDPB8

4.52 A. Ghazisaeidi, P. Tran, P. Brindel, O. Bertran-Pardo, J. Renaudier, G. Charlet, S. Bigo: Impact of tight optical filtering on the performance of 28 Gbaud Nyquist-WDM PDM-8QAM over 37.5 GHz grid. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2013), Paper OTu3B.6

4.53 B. Lavigne, M. Lefrançois, O. Bertran-Pardo, M. Le Monnier, L. Raddatz, S. Weisser, R. Peruta, G.A. Azz-

ini, L. Suberini: Real-time 200 Gb/s 8-QAM transmission over a 1800-km long SSMF-based system using add/drop 50 GHz-wide filters. In: *Proc. Opt. Fiber Commun. (OFC), Anaheim* (2016), Paper W3G.2

4.54 K. Roberts, Q. Zhuge, I. Monga, S. Gareau, C. Laperle: Beyond 100 Gb/s: Capacity, flexibility, and network optimization [invited], J. Opt. Commun. Netw. **9**(4), C12–C24 (2017)

4.55 Optical Internetworking Forum (OIF), 400G – ZR coherent interface, available at http://www.oiforum.com/

4.56 C. Li, Z. Zhang, J. Chen, T. Ding, Z. Xiao, F. Shah, J. Mitra, H. Xiang, X. Cui: Advanced DSP for single-carrier 400-Gb/s PDM-16QAM. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2016), Paper W4A.4

4.57 L. Kazovsky, S. Benedetto, A. Willner: *Optical Fiber Communication Systems* (Artech House, Norwood 1996)

4.58 R.S. Tucker: High-speed modulation of semiconductor lasers, IEEE Trans. Electron Dev. **32**(12), 2572–2584 (1985)

4.59 K. Czotscher, S. Weisser, A. Leven, J. Rosenzweig: Intensity modulation and chirp of 1.55-μm multiple-quantum-well laser diodes: modeling and experimental verification, IEEE J. Sel. Top. Quantum Electron. **5**(3), 606–612 (1999)

4.60 K. Sato, S. Kuwahara, Y. Miyamoto: Chirp characteristics of 40-Gb/s directly modulated distributed-feedback laser diodes, J. Lightwave Technol. **23**(11), 3799–3797 (2005)

4.61 N. Eiselt, H. Griesser, J. Wei, A. Dochhan, R. Hohenleitner, M. Ortsiefer, M. Eiselt, C. Neumeyr, J.J.V. Olmos, I.T. Monroy: Experimental demonstration of 56 Gbit/s PAM-4 over 15 km and 84 Gbit/s PAM-4 over 1 km SSMF at 1525 nm using a 25G VCSEL. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Amsterdam* (2016), Paper Th.1.C.1

4.62 I. Lyubomirsky, W.A. Ling: Advanced modulation for datacenter interconnect. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2016), Paper W4J.3

4.63 Y. Miyamoto, A. Hirano, K. Yonenaga, A. Sano, H. Toba, K. Murata, O. Mitomi: 320 Gbit/s (8×40 Gbit/s) WDM transmission over 367 km with 120-km repeater spacing using carrier-suppressed return-to-zero format, Electron. Lett. **35**(23), 2041–2042 (1999)

4.64 N.S. Bergano: Undersea communication systems. In: *Optical Fiber Telecommunication IV*, ed. by I. Kaminow, T. Li (Academic Press, New York 2002)

4.65 T. Saito, N. Henmi, S. Fujita, M. Yamaguchi, M. Shikada: Prechirp technique for dispersion compensation for a high-speed long span transmission, IEEE Photonics Technol. Lett. **3**(1), 74–76 (1991)

4.66 B. Bakhshi, M. Vaa, E.A. Golovchenko, W.W. Patterson, R.L. Maybach, N.S. Bergano: Comparison of CRZ, RZ and NRZ modulation formats in a 64×12.3 Gb/s WDM transmission experiment over 9000 km. In: *Proc. Opt. Fiber Commun. Conf. Int. Conf. Quantum Inf. (OFC), Anaheim* (2001), https://doi.org/10.1364/OFC.2001.WF4

4.67 P.J. Winzer, C. Dorrer, R.-J. Essiambre, I. Kang: Chirped return-to-zero modulation by imbalanced pulse carver driving signals, IEEE Photonics Technol. Lett. **16**(5), 1379–1381 (2004)

4.68 W.A. Atia, R.S. Bondurant: Demonstration of return-to-zero signaling in both OOK and DPSK formats to improve receiver sensitivity in an optically preamplified receiver. In: *IEEE Lasers Electro-Opt. Soc. Ann. Meet. Conf. Proc. (LEOS '99), Orlando* (1999) pp. 226–227, https://doi.org/10.1109/LEOS.1999.813561

4.69 P.J. Winzer, M. Pfennigbauer, M.M. Strasser, W.R. Leeb: Optimum filter bandwidths for optically preamplified RZ and NRZ receivers, J. Lightwave Technol. **19**(9), 1263–1273 (2001)

4.70 P.J. Winzer, A. Kalmar: Sensitivity enhancement of optical receivers by impulsive coding, J. Lightwave Technol. **17**(2), 171–177 (1999)

4.71 M. Pauer, P.J. Winzer, W.R. Leeb: Bit error probability reduction in direct detection optical receivers using RZ coding, J. Lightwave Technol. **19**(9), 1255–1262 (2001)

4.72 M. Pfennigbauer, M.M. Strasser, M. Pauer, P.J. Winzer: Dependence of optically preamplified receiver sensitivity on optical and electrical filter bandwidths-measurement and simulation, IEEE Photonics Technol. Lett. **14**(6), 831–833 (2002)

4.73 G. Bosco, R. Gaudino, P. Poggiolini: An exact analysis of RZ versus NRZ sensitivity in ASE noise limited optical systems. In: *Proc. 27th Eur. Conf. Opt. Commun., Amsterdam*, Vol. 4 (2001) pp. 526–527

4.74 G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini: On the use of NRZ, RZ, and CSRZ modulation at 40 Gb/s with narrow DWDM channel spacing, J. Lightwave Technol. **20**(9), 1694–1704 (2002)

4.75 J.L. Rebola, A.V.T. Cartaxo: Power penalty assessment in optically preamplified receivers with arbitrary optical filtering and signal-dependent noise dominance, J. Lightwave Technol. **20**(3), 401–408 (2002)

4.76 I. Lyubomirsky, S. Shetty, J. Roman, M.Y. Frankel: Optimum 10-Gb/s NRZ receiver bandwidths for ultradense WDM transmission systems, IEEE Photonics Technol. Lett. **14**(6), 870–872 (2002)

4.77 Telecommunication Standardization Sector (ITU-T): Recommendation G.975.1: Forward error correction for high bit-rate DWDM submarine systems, available at https://www.itu.int/rec/T-REC-G/en

4.78 G.P. Agrawal: *Fiber-Optic Communication System* (Wiley, Hoboken 2011)

4.79 R.S. Fyath, J.J. O'Reilly: Performance of optically preamplified direct-detection receivers in the presence of laser chirp. Part I. Ideal travelling-wave amplifier, IEE Proc. J. Optoelectron. **136**(5), 249–255 (1989)

4.80 G.J. Foschini, R.M. Jopson, L.E. Nelson, H. Kogelnik: The statistics of PMD-induced chromatic fiber dispersion, J. Lightwave Technol. **17**(9), 1560–1565 (1999)

4.81 E. Ip, J.M. Kahn: Digital equalization of chromatic dispersion and polarization mode dispersion, J. Lightwave Technol. **25**(8), 2033–2043 (2007)

4.82 R.M. Jopson, L.E. Nelson, G.J. Pendock, A.H. Gnauck: Polarization-mode dispersion impairment in return-to-zero and non return-to-zero systems. In: *Proc. Opt. Fiber Commun. Conf. (OFC), San Diego* (1999), Paper WE3

4.83 J.M. Gene, P.J. Winzer: First-order PMD outage prediction based on outage maps, J. Lightwave Technol. **28**(13), 1873–1881 (2010)

4.84 P.J. Winzer, H. Kogelnik, K. Ramanan: Precise outage specifications for first-order PMD, IEEE Photonics Technol. Lett. **16**(2), 449–451 (2004)

4.85 Telecommunication Standardization Sector (ITU-T): Recommendation G.Sup39: Optical system design and engineering considerations, available at https://www.itu.int/rec/T-REC-G/en

4.86 P.J. Winzer, H. Kogelnik, C.H. Kim, H. Kim, R.M. Jopson, L.E. Nelson, K. Ramanan: Receiver impact on first-order PMD outage, IEEE Photonics Technol. Lett. **15**(10), 1482–1484 (2003)

4.87 C. Xie, S. Shen, L. Möller: Effects of transmitter imperfection on polarization mode dispersion impairments, IEEE Photonics Technol. Lett. **15**(4), 614–616 (2003)

4.88 P. Poggiolini, G. Bosco: Impact of chromatic dispersion on DPSK and DQPSK direct-detection optical systems, Annal. Telecommun. **62**(5/6), 531–549 (2007)

4.89 K. Tanaka, T. Tsuritani, N. Edagawa, M. Suzuki: 320 Gbit/s (32×10.7 Gbit/s) error-free transmission over 7280 km using dispersion flattened fibre link with standard SMF and slope compensating DCF, Electron. Lett. **35**(21), 1860–1862 (1999)

4.90 A.H. Gnauck, R.W. Tkach, A.R. Chraplyvy, T. Li: High-capacity optical transmission systems, J. Lightwave Technol. **26**(9), 1032–1045 (2008)

4.91 J. Rebola, A. Cartaxo: Optimization of level spacing in quaternary optical communication systems, Proc. SPIE **4087**, 49–59 (2000)

4.92 J.M. Kahn, K.-P. Ho: Spectral efficiency limits and modulation/detection techniques for DWDM systems, J. Sel. Top. Quantum Electron. **10**(2), 259–272 (2004)

4.93 G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini: Quantum limit of direct detection optically preamplified receivers using duobinary transmission, IEEE Photonics Technol. Lett. **15**(1), 102–104 (2003)

4.94 M. Joindot, G. Bosco, A. Carena, V. Curri, P. Poggiolini: Fundamental performance limits of optical duobinary, Opt. Express **16**(24), 19600–19614 (2008)

4.95 X. Gu, L.C. Blank: 10 Gbit/s unrepeated optical transmission over 100 km of standard fibre, Electron. Lett. **29**(25), 2209–2211 (1993)

4.96 K. Yonenaga, S. Kuwano, S. Norimatsu, N. Shibata: Optical duobinary transmission system with no receiver sensitivity degradation, Electron. Lett. **3**(4), 302–304 (1995)

4.97 A.J. Price, L. Pierre, R. Uhel, V. Havard: 210 km repeaterless 10 Gb/s transmission experiment through nondispersion-shifted fiber using partial response scheme, IEEE Photonics Technol. Lett. **7**(10), 1219–1221 (1995)

4.98 D. Penninckx, M. Chbat, L. Pierre, J.-P. Thiery: The phase-shaped binary transmission (PSBT): A new technique to transmit far beyond the chromatic dispersion limit. In: *Proc. Eur. Conf. Opt. Commun. (ECOC'96), Oslo*, Vol. 2 (1996) pp. 173–176

4.99 G. Bosco, A. Carena, V. Curri, P. Poggiolini: Best optical filtering for duobinary transmission. In: *Optical Communication Theory and Techniques* (Springer, Norwell 2005) pp. 21–28

4.100 E. Ip, J.M. Kahn: Feedforward carrier recovery for coherent optical communications, J. Lightwave Technol. **25**(9), 2675–2692 (2007)

4.101 K.-P. Ho: *Phase-Modulated Optical Communication Systems* (Springer, New York 2005)

4.102 P.J. Winzer, S. Chandrasekhar, H. Kim: Impact of filtering on RZ-DPSK reception, IEEE Photonics Technol. Lett. **15**(6), 840–842 (2003)

4.103 P. Poggiolini, S. Benedetto: Theory of polarization spreading techniques—Part I, IEEE Trans. Commun. **42**(5), 2105–2118 (1994)

4.104 G. Bosco, P. Poggiolini: On the joint effect of receiver impairments on direct-detection DQPSK systems, J. Lightwave Technol. **24**(3), 1323–1333 (2006)

4.105 P.J. Winzer, H. Kim: Degradations in balanced DPSK receivers, IEEE Photonics Technol. Lett. **15**(9), 1282–1284 (2003)

4.106 H. Kim, P.J. Winzer: Robustness to laser frequency offset in direct detection DPSK and DQPSK systems, J. Lightwave Technol. **21**(9), 1887–1891 (2003)

4.107 K.-P. Ho: The effect of interferometer phase error on direct-detection DPSK and DQPSK signals, IEEE Photonics Technol. Lett. **16**(1), 308–310 (2004)

4.108 G. Bosco, P. Poggiolini: The impact of receiver imperfections on the performance of optical direct-detection DPSK, J. Lightwave Technol. **23**(2), 842–848 (2005)

4.109 K.-P. Ho, H.-W. Cuei: Generation of arbitrary quadrature signals using one dual-drive modulator, J. Lightwave Technol. **23**(2), 764–770 (2005)

4.110 M. Seimetz: *High Order Modulation for Optical Fiber Transmission* (Springer, Berlin, Heidelberg 2009)

4.111 P.S. Cho, J.B. Khurgin, I. Shpantzer: Closed-loop bias control of optical quadrature modulator, IEEE Photonics Technol. Lett. **18**(21), 2209–2211 (2006)

4.112 T. Yoshida, T. Sugihara, K. Uto, H. Bessho, K. Sawada, K. Ishida, K. Shimizu, T. Mizuochi: A study on automatic bias control for arbitrary optical signal generation by dual-parallel Mach-Zehnder modulator. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Torino* (2010), Paper Tu.3.A.6

4.113 H. Kawakami, E. Yoshida, Y. Miyamoto: Asymmetric dithering technique for bias condition monitoring in optical QPSK modulator, Electron. Lett. **46**(6), 430–431 (2010)

4.114 M. Sotoodeh, Y. Beaulieu, J. Harley, D.L. McGhan: Modulator bias and optical power control of optical complex E-field modulators, J. Lightwave Technol. **29**(15), 2235–2248 (2011)

4.115 Y. Yang, C. Lu, A.P.T. Lau, Y. Yao, Y. Sun, J.J. Xiao, H.Y. Tam, P.K.A. Wai: A robust and dither-free technique for controlling driver signal amplitude for

stable and arbitrary optical phase modulation, Opt. Express **19**(27), 26353–26358 (2011)
4.116 R. Cigliutti, A. Nespola, D. Zeolla, G. Bosco, A. Carena, V. Curri, F. Forghieri, Y. Yamamoto, T. Sasaki, P. Poggiolini: 16 × 125 Gb/s Quasi-Nyquist DAC-generated PM-16QAM transmission over 3590 km of PSCF, IEEE Photonics Technol. Lett. **24**(23), 2143–2146 (2012)
4.117 R. Schmogrow, M. Meyer, P.C. Schindler, A. Josten, S. Ben-Ezra, C. Koos, W. Freude, J. Leuthold: 252 Gbit/s real-time Nyquist pulse generation by reducing the oversampling factor to 1.33. In: *Proc. Opt. Fiber Commun. Conf., Anaheim* (2013), https://doi.org/10.1364/OFC.2013.OTu2I.1
4.118 A. Nespola, S. Straullu, G. Bosco, A. Carena, J. Yanchao, P. Poggiolini, F. Forghieri, Y. Yamamoto, M. Hirano, T. Sasaki, J. Bauwelinck, K. Verheyen: 1306-km 20 × 124.8-Gb/s PM-64QAM transmission over PSCF with Net SEDP 11,300 (b · km)/s/Hz using 1.15 samp/symb DAC, Opt. Express **22**(1), 1796–1805 (2014)
4.119 G. Khanna, B. Spinnler, S. Calabrò, E. De Man, N. Hanik: A robust adaptive pre-distortion method for optical communication transmitters, IEEE Photonics Technol. Lett. **28**(7), 752–755 (2016)
4.120 A.W. Davis, M.J. Pettitt, P.J. King, S. Wright: Phase diversity techniques for coherent optical receivers, J. Lightwave Technol. **5**(4), 561–572 (1987)
4.121 T. Okoshi, K. Kikuchi: *Coherent Optical Fiber Communications* (Springer, Dordrecht 1988)
4.122 K. Kikuchi: Fundamentals of coherent optical fiber communications, J. Lightwave Technol. **34**(1), 157–179 (2016)
4.123 G. Bosco, V. Curri, A. Carena, P. Poggiolini, F. Forghieri: On the performance of Nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers, J. Lightwave Technol. **29**(1), 53–61 (2011)
4.124 X. Liu, S. Chandrasekhar, P. Winzer: Digital signal processing techniques enabling multi-Tb/s superchannel transmission, IEEE Signal Process. Mag. **31**(2), 16–24 (2014)
4.125 S. Chandrasekhar, X. Liu, B. Zhu, D.W. Peckham: Transmission of a 1.2-Tb/s 24-carrier no-guard-interval coherent OFDM superchannel over 7200-km of ultra-large-area fiber. In: *Proc. 35th Eur. Conf. Opt. Commun. (ECOC), Vienna* (2009), Paper PD2.6
4.126 G. Bosco, A. Carena, V. Curri, P. Poggiolini, F. Forghieri: Performance limits of Nyquist-WDM and CO-OFDM in high-speed PM-QPSK systems, IEEE Photonics Technol. Lett. **22**(15), 1129–1131 (2010)
4.127 G. Bosco, D. Pilori, P. Poggiolini, A. Carena, F. Guiomar: Scalable modulation technology and the tradeoff of reach, spectral efficiency, and complexity, Proc. SPIE (2017), https://doi.org/10.1117/12.2249160
4.128 R. Cigliutti, E. Torrengo, G. Bosco, N.P. Caponio, A. Carena, V. Curri, P. Poggiolini, Y. Yamamoto, T. Sasaki, F. Forghieri: Transmission of 9 × 138 Gb/s prefiltered PM-8QAM signals over 4000 km of pure silica-core fiber, J. Lightwave Technol. **29**(15), 2310–2318 (2011)
4.129 J.-X. Cai: 100G transmission over transoceanic distance with high spectral efficiency and large capacity, J. Lightwave Technol. **30**(24), 3845–3856 (2012)
4.130 D.G. Foursa, H.G. Batshon, H. Zhang, M. Mazurczyk, J.-X. Cai, O. Sinkin, A. Pilipetskii, G. Mohs, N.S. Bergano: 44.1 Tb/s transmission over 9,100 km using coded modulation based on 16QAM signals at 4.9 bits/s/Hz spectral efficiency. In: *Proc. Eur. Conf. Opt. Commun., London* (2013), Paper PD3.E.1
4.131 G. Bosco: Spectral shaping in ultra-dense WDM systems: Optical vs. electrical approaches. In: *Proc. Opt. Fiber Commun. Conf., Los Angeles* (2012), Paper OM3H.1
4.132 D. Hillerkuss, R. Schmogrow, M. Meyer, S. Wolf, M. Jordan, P. Kleinow, N. Lindenmann, P.C. Schindler, A. Melikyan, X. Yang, S. Ben-Ezra, B. Nebendahl, M. Dreschmann, J. Meyer, F. Parmigiani, P. Petropoulos, B. Resan, A. Oehler, K. Weingarten, L. Altenhain, T. Ellermeyer, M. Moeller, M. Huebner, J. Becker, C. Koos, W. Freude, J. Leuthold: Single-laser 32.5 Tbit/s Nyquist WDM transmission, J. Opt. Commun. Netw. **4**(10), 715–723 (2012)
4.133 X. Zhou, L.E. Nelson, P. Magill, R. Isaac, B. Zhu, D.W. Peckham, P.I. Borel, K. Carlson: PDM-Nyquist-32QAM for 450-Gb/s per-channel WDM transmission on the 50 GHz ITU-T grid, J. Lightwave Technol. **30**(4), 553–559 (2012)
4.134 X. Zhou, L. Nelson, P. Magill: Rate-adaptable optics for next generation long-haul transport networks, IEEE Commun. Mag. **51**(3), 41–49 (2013)
4.135 J.-X. Cai, H. Zhang, H.G. Batshon, M. Mazurczyk, O.V. Sinkin, D.G. Foursa, A.N. Pilipetskii, G. Mohs, N.S. Bergano: 200 Gb/s and dual wavelength 400 Gb/s transmission over transpacific distance at 6.0 bit/s/Hz spectral efficiency, J. Lightwave Technol. **32**(4), 832–939 (2014)
4.136 G. Bosco, S.M. Bilal, A. Nespola, P. Poggiolini, F. Forghieri: Impact of the transmitter IQ-skew in multi-subcarrier coherent optical systems. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2016), Paper W4A.5
4.137 G. Bosco: Spectrally efficient multiplexing: Nyquist-WDM. In: *Enabling Technologies for High Spectral-efficiency Coherent Optical Communication Networks* (Wiley, Hoboken 2016) pp. 123–156
4.138 A. Sano, T. Kobayashi, S. Yamanaka, A. Matsuura, H. Kawakami, Y. Miyamoto, K. Ishihara, H. Masuda: 102.3-Tb/s (224 × 548-Gb/s) C- and extended L-band all-Raman transmission over 240 km using PDM-64QAM single carrier FDM with digital pilot tone. In: *Proc. Opt. Fiber Commun. Conf., Los Angeles* (2012), Paper PDP5C.3
4.139 R. Schmogrow, M. Winter, M. Meyer, D. Hillerkuss, B. Nebendahl, J. Meyer, M. Dreschmann, M. Huebner, J. Becker, C. Koos, W. Freude, J. Leuthold: Real-time Nyquist pulse modulation transmitter generating rectangular shaped spectra of 112 Gbit/s 16QAM signals. In: *Proc. Adv. Photonics (SPPCom), Toronto* (2011), https://doi.org/10.1364/SPPCOM.2011.SPMA5

4.140 M. Mazurczyk: Spectral shaping in long haul optical coherent systems with high spectral efficiency, J. Lightwave Technol. **32**(16), 2915–2924 (2014)

4.141 Telecommunication Standardization Sector (ITU-T): Recommendation G.8251: The control of jitter and wander within the optical transport network (OTN), available at https://www.itu.int/rec/T-REC-G/en

4.142 C.R.S. Fludger, T. Duthel, P. Hermann, T. Kupfer: Jitter tolerant clock recovery for coherent optical receivers. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2013), Paper OTh1F.3

4.143 P. Winzer: High-spectral-efficiency optical modulation formats, J. Lightwave Technol. **30**(24), 3824–3835 (2012)

4.144 O. Gerstel, M. Jinno, A. Lord, S.J. Ben Yoo: Elastic optical networking: A new dawn for the optical layer?, IEEE Signal Process. Mag. **50**(2), s12–s20 (2012)

4.145 M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, S. Matsuoka: Spectrum-efficient and scalable elastic optical path network: Architecture, benefits, and enabling technologies, IEEE Signal Process. Mag. **47**(11), 66–72 (2009)

4.146 M. Qiu, Q. Zhuge, M. Chagnon, Y. Gao, X. Xu, M. Morsy-Osman, D.V. Plant: Digital subcarrier multiplexing for fiber nonlinearity mitigation in coherent optical communication systems, Opt. Express **22**(15), 18770–18777 (2014)

4.147 W. Shieh, H. Bao, Y. Tang: Coherent optical OFDM: Theory and design, Opt. Express **16**(2), 841–859 (2008)

4.148 Q. Zhuge, M. Morsy-Osman, X. Xu, M. Chagnon, M. Qiu, D. Plant: Spectral efficiency-adaptive optical transmission using time domain hybrid QAM for agile optical networks, J. Lightwave Technol. **31**(15), 2621–2628 (2013)

4.149 F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, W. Idler: Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration, J. Lightwave Technol. **34**(7), 1599–1609 (2016)

4.150 A.J. Lowery, L.B. Du, J. Armstrong: Performance of optical OFDM in ultralong-haul WDM lightwave systems, J. Lightwave Technol. **25**(1), 131–138 (2007)

4.151 J. Armstrong: OFDM for optical communications, J. Lightwave Technol. **27**(3), 189–204 (2009)

4.152 L. Nadal, M.S. Moreolo, J.M. Fabrega, A. Dochhan, H. Grießer, M. Eiselt, J.-P. Elbers: DMT modulation with adaptive loading for high bit rate transmission over directly detected optical channels, J. Lightwave Technol. **32**(21), 3541–3551 (2014)

4.153 T. Takahara, T. Tanaka, M. Nishihara, Y. Kai, L. Li, Z. Tao, J. Rasmussen: Discrete multi-tone for 100 Gb/s optical access networks. In: *Proc. Opt. Fiber Commun. Conf. (OFC), San Francisco* (2014), Paper M2I.1

4.154 P. Dong, J. Lee, K. Kim, Y. Chen, C. Gui: Ten-channel discrete multi-tone modulation using silicon microring modulator array. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2016), Paper W4J.4

4.155 K. Zhong, X. Zhou, T. Gui, L. Tao, Y. Gao, W. Chen, J. Man, L. Zeng, A.P.T. Lau, C. Lu: Experimental study of PAM-4, CAP-16, and DMT for 100 Gb/s short reach optical transmission systems, Opt. Express **23**(2), 1176–1189 (2015)

4.156 R. Schmogrow, S. Wolf, B. Baeuerle, D. Hillerkuss, B. Nebendahl, C. Koos, W. Freude, J. Leuthold: Nyquist frequency division multiplexing for optical communications. In: *Proc. Conf. Lasers Electro-Opt., San Jose* (2012), Paper CTh1H.2

4.157 L. Dou, X. Su, Y. Fan, H. Chen, Y. Zhao, Z. Tao, T. Tanimura, T. Hoshida, J. Rasmussen: 420Gbit/s DP-64QAM Nyquist-FDM single-carrier system. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2016), Paper Tu3A.5

4.158 B. Baeuerle, A. Josten, M. Eppenberger, D. Hillerkuss, J. Leuthold: FPGA-based realtime receivers for Nyquist-FDM. In: *Proc. Adv. Photonics. Signal Proc. Photonic Commun., New Orleans* (2017), https://doi.org/10.1364/SPPCOM.2017.SpM3F.3

4.159 P. Poggiolini, A. Nespola, Y. Jiang, G. Bosco, A. Carena, L. Bertignono, S.M. Bilal, S. Abrate, F. Forghieri: Analytical and experimental results on system maximum reach increase through symbol rate optimization, J. Lightwave Technol. **34**(8), 1872–1885 (2016)

4.160 P.C. Schindler, R. Schmogrow, S. Wolf, B. Baeuerle, B. Nebendahl, C. Koos, W. Freude, J. Leuthold: Full flex-grid asynchronous multiplexing demonstrated with Nyquist pulse-shaping, Opt. Express **22**, 10923–10937 (2014)

4.161 D. Krause, A. Awadalla, A.S. Karar, H. Sun, K.-T. Wu: Design considerations for a digital subcarrier coherent optical modem. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Los Angeles* (2017), Paper Th1D.1

4.162 B. Baeuerle, A. Josten, R. Bonjour, D. Hillerkuss, J. Leuthold: Effect of transmitter impairments on Nyquist-FDM signals with increasing subband granularity. In: *Proc. Adv. Photonics. Signal Proc. Photonic Commun., Vancouver* (2016), https://doi.org/10.1364/SPPCOM.2016.SpW3F.4

4.163 X. Zhou, L.E. Nelson, P. Magill, R. Isaac, B. Zhu, D.W. Peckham, P.I. Borel, K. Carlson: High spectral efficiency 400 Gb/s transmission using PDM time-domain hybrid 32–64 QAM and training-assisted carrier recovery, J. Lightwave Technol. **31**(7), 999–1005 (2013)

4.164 F. Buchali, L. Schmalen, K. Schuh, W. Idler: Optimization of time-division hybrid-modulation and its application to rate adaptive 200 Gb transmission. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Cannes* (2014), https://doi.org/10.1109/ECOC.2014.6963823

4.165 V. Curri, A. Carena, P. Poggiolini, R. Cigliutti, F. Forghieri, C.R. Fludger, T. Kupfer: Time-division hybrid modulation formats: Tx operation strategies and countermeasures to nonlinear propagation. In: *Proc. Opt. Fiber Commun. Conf. (OFC), San Francisco* (2014), Paper Tu3A.2

4.166 W. Idler, F. Buchali, L. Schmalen, K. Schuh, H. Buelow: Hybrid modulation formats outperforming 16QAM and 8QAM in transmission distance and filtering with cascaded WSS. In: *Proc. Opt.*

Fiber Commun. Conf. (OFC), Los Angeles (2015), Paper M3G.4
4.167 R. Dar, M. Feder, A. Mecozzi, M. Shtaif: On shaping gain in the nonlinear fiber-optic channel. In: *Proc. IEEE Int. Symp. Inf. Theory (ISIT), Honolulu* (2014) pp. 2794–2798
4.168 M.P. Yankov, D. Zibar, K.J. Larsen, L.P.B. Christensen, S. Forchhammer: Constellation shaping for fiber-optic channels with QAM and high spectral efficiency, IEEE Photonics Technol. Lett. **26**(23), 2407–2410 (2014)
4.169 B. Smith, F. Kschischang: A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications, J. Lightwave Technol. **30**(13), 2047–2053 (2012)
4.170 T. Fehenberger, G. Böcherer, A. Alvarado, N. Hanik: LDPC coded modulation with probabilistic shaping for optical fiber systems. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Los Angeles* (2015), Paper Th2A.23
4.171 D. Pilori, L. Bertignono, A. Nespola, F. Forghieri, G. Bosco: Comparison of probabilistically shaped 64QAM with lower-cardinality uniform constellations in long-haul optical systems, J. Lightwave Technol. **36**(2), 501–509 (2017)
4.172 F.P. Guiomar, L. Bertignono, D. Pilori, A. Nespola, G. Bosco, A. Carena, F. Forghieri: Comparing Different Options for Flexible Networking: Probabilistic Shaping vs. Hybrid Subcarrier Modulation. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Goteborg* (2017), Paper Th1.E.3
4.173 F. Steiner, G. Boecherer: Comparison of geometric and probabilistic shaping with application to ATSC 3.0. In: *Proc. 11th Int. ITG Conf. Syst. Commun. Coding (SCC 2017), Hamburg* (2017)
4.174 S. Gringeri, N. Bitar, R. Egorov, B. Basch, C. Suitor, H. Peng: Optimizing transport systems to integrate TDM and packet services. In: *Proc. Nat. Fiber Opt. Eng. Conf. (NFOEC), Anaheim* (2007), Paper NTuA4
4.175 Optical Internetworking Forum (OIF): OTN Over Packet Fabric Protocol (OFP). Implementation Agreement, available at http://www.oiforum.com/
4.176 Telecommunication Standardization Sector (ITU-T): Recommendation G.709: Interfaces for the optical transport network, available at https://www.itu.int/rec/T-REC-G/en
4.177 Telecommunication Standardization Sector (ITU-T): Recommendation G.7041: Generic framing procedure, available at https://www.itu.int/rec/T-REC-G/en
4.178 Ethernet Alliance: https://ethernetalliance.org/
4.179 Storage Networking Industry Association (SNIA) Small Form Factor (SFF) Technology Affiliate Technical Work Group: https://ta.snia.org/higherlogic/ws/public
4.180 Quad Small Form Factor Pluggable Double Density (QSFP-DD) Multi Source Agreement (MSA) Group: http://www.qsfp-dd.com/
4.181 Octal Small Form Factor Pluggable (OSFP) Multi Source Agreement (MSA) Group: http://osfpmsa.org/index.html
4.182 Small Form Factor Pluggable Double Density (SFP-DD): http://sfp-dd.com/
4.183 Multi Source Agreement Group: 100 G Formfaktor Pluggable (CFP), available at http://www.cfp-msa.org/
4.184 Optical Internetworking Forum (OIF): OIF-MSA-100GLH-EM-02.1 – Multisource Agreement for Generation 2.0 100G Long-Haul DWDM Transmission Module – Electromechanical (March 2015), available at http://www.oiforum.com/
4.185 Consortium on On-Board Optics (COBO): http://onboardoptics.org/
4.186 Telecommunication Standardization Sector (ITU-T): Standardization Sector, G series recommendations, available at https://www.itu.int/rec/T-REC-G/en
4.187 Optical Internetworking Forum (OIF): OIF-CEI-03.1 – Common Electrical I/O (CEI) – Electrical and Jitter Interoperability agreements for 6G+ bps, 11G+ bps, 25G+ bps I/O (February 2014), available at http://www.oiforum.com/
4.188 Telecommunication Standardization Sector (ITU-T): Recommendation G.Sup58: Optical transport network (OTN) module framer interfaces, available at https://www.itu.int/rec/T-REC-G/en

Gabriella Bosco

Dept. of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
gabriella.bosco@polito.it

Gabriella Bosco is an Associate Professor at Politecnico di Torino, Italy, where she received her PhD degree in Electronic and Communication Engineering in 2002. Her current research interests include the performance analysis and design of high-capacity optical transmission systems and subsystems. She is Senior Member IEEE, Fellow OSA and Deputy Editor of the IEEE/OSA Journal of Lightwave Technology.

Jörg-Peter Elbers

Advanced Technology
ADVA Optical Networking SE
Martinsried, Germany
jelbers@adva.com

Jörg-Peter Elbers is Senior Vice President of Advanced Technology, Standards and IPR at ADVA Optical Networking. Before joining ADVA, he was Director of Technology in the Optical Product Unit of Marconi (now Ericsson) and Director of Network Architecture in the Optical Networks Business Unit of Siemens. He received his Dr-Ing degree in electrical engineering from the Technical University Dortmund, Germany.

5. Optical Transponder Components

Xi Chen, Sethumadhavan Chandrasekhar

This chapter introduces the general architecture of an optical transponder and describes the three critical optical components that comprise the transponder, the laser, the optical modulator, and the photodetector. Following this, the subsystem consisting of the optical transmitter and the coherent receiver that are typically used for generating and detecting dual-polarization complex-modulated signals are explained. The typical characteristics of the components used in the transponder are presented both from current-generation products, as well as from more recent research demonstrations.

In this chapter, we focus on the line-side optics, as they are the key to understanding the components in a transponder. The transmitter at the line side consists of a number of high-performance optical components, such as narrow-linewidth lasers, and high-speed modulators with linear driver amplifiers. The receiver at the line side consists of 90° optical hybrids, high-dynamic-range linear balanced photodetectors (BPDs), and high-gain low-noise transimpedance amplifiers (TIAs). The digital electronic functions are in application-specific integrated circuits (ASICs), which include digital-to-analog converters (DACs) for the transmitter to generate optical signals, analog-to-digital converters (ADCs) for the receiver to recover the optical signals. Both ends need digital signal processing (DSP) engines to condition and process the signals. The client side optics, which are not covered in this chapter, are usually a subset or a simpler version of those used in the line-side part of the transponder.

An optical transponder consists of a client side and a line side. The client side collects traffic from various client applications such as text, images, audio, video, etc. The line side connects to the optical core network and has high-speed interfaces. The transponder collects low-speed optical signals from clients, down-converts the optical signals to electrical signals, and multiplexes all the information collected onto one optical carrier that carries the high-speed signal. This high-speed signal will then be sent to the core network. The core network is often referred to as the line system. An optical transponder also performs the reverse function, namely, de-multiplexing the high-speed signal received from the line system to multiple low-speed client optical signals and sends them to the clients. An optical transponder is usually characterized by the information

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_5

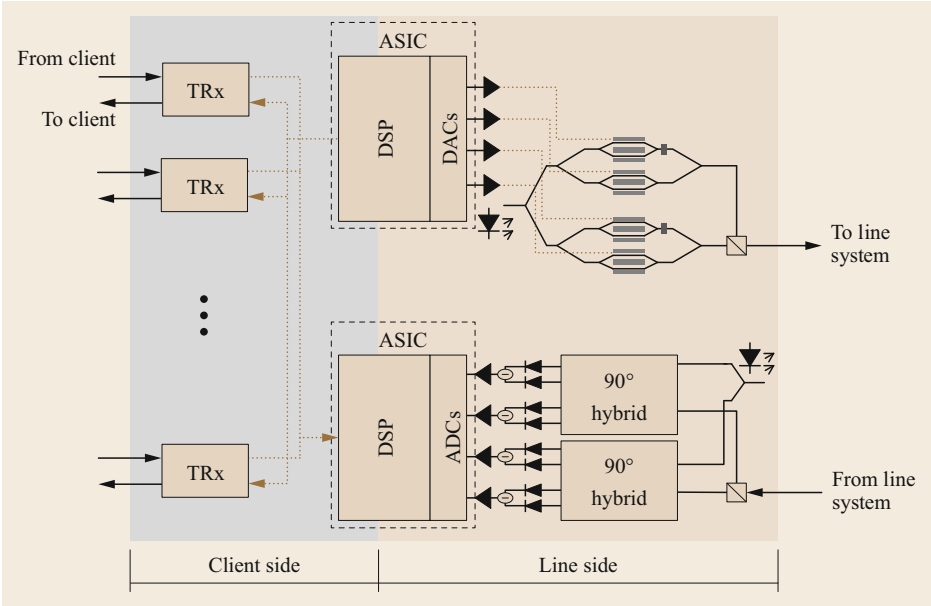


Fig. 5.1 Schematic architecture of an optical transponder (TRx: transceiver)

rate that faces the line side. For instance, a 100 Gb/s transponder could collect either ten 10 Gb/s signals or four 25 Gb/s to form 100 Gb/s and send them to the line system. A schematic hardware architecture of an optical transponder is shown in Fig. 5.1.

In the following sections, we first explain three critical components of an optical transponder, namely, the laser, the optical modulator, and the photodetector. We then describe the architecture of the transmitter and receivers and their other related components.

5.1 Lasers

The word *laser* stands for *light amplification by stimulated emission of radiation*. It is a device that emits light through the process of optical amplification based on stimulated emission of electromagnetic radiation, and the device under this condition is sometimes referred to as *lasing*. A laser, different from other light sources such as light bulbs and light-emitting diodes (LEDs), emits coherent light with a very narrow optical spectrum, which makes it particularly suitable for telecommunication applications.

In order to achieve *lasing*, the laser needs a cavity and a gain medium in the cavity. Depending on the type of gain medium, lasers can be categorized as solid-state lasers, liquid lasers, gas lasers, and semiconductor lasers. The most common type of laser for telecommunications is the semiconductor laser, due to its low cost, compact size, and low power consumption. Semiconductor lasers are also known as *laser diodes*, as a pn junction of a semiconductor diode forms the gain medium. The physics behind the stimulated emission is important for understanding lasers. The stimulated emission is the process that generates coherent photons, resulting in a highly coherent and directional light beam. A semiconductor laser needs a pump current to move the electrons to their higher-energy state so that stimulated emission can occur. The details of these processes are not covered in this chapter but can be found in other textbooks [5.1, 2]. In this chapter, we explain the basics of how lasers work based on different cavity structures. Depending on the cavity structure, lasers can be categorized as Fabry–Pérot (FP) lasers, distributed feedback (DFB) lasers, or external cavity lasers (ECLs).

5.1.1 Fabry–Pérot Laser

The FP laser, though not used in optical transponders, has important features, and by studying these features we can learn about lasers in general. Figure 5.2 shows the schematic structure of an FP laser. An FP resonator forms the cavity with one mirror at each end. A gain medium is placed inside the FP resonator.

The laser starts with some weak spontaneous emission. At this stage, the light is broadband and incoherent. The two mirrors, one with very high reflection

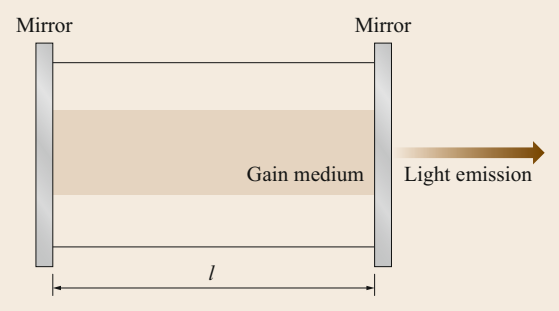


Fig. 5.2 Schematic architecture of an FP laser

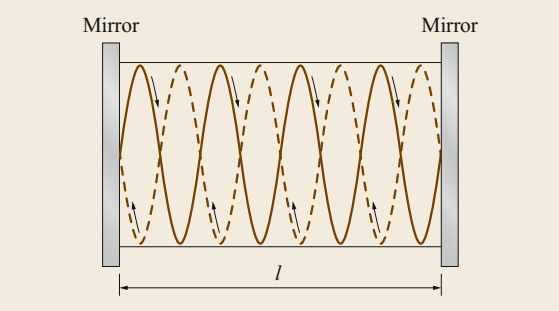


Fig. 5.3 An example of a cavity mode

($\approx 100\%$) and the other with a slightly lower reflectivity ($\approx 99\%$), cause the broadband light to bounce back and forth between the mirrors. One can imagine that each round-trip of the bouncing generates a new wave that can be superimposed on the original wave. The new wave will be constructively added to the old wave when they are in phase. This basically requires the wave to repeat itself after every round trip. An example of such repetition is shown in Fig. 5.3.

To achieve such repetition, the wavelength must equal the cavity round-trip distance divided by an integer number. In other words, assuming the cavity has a length of l, the wavelength λ needs to satisfy

$$N\lambda = 2l\,, \tag{5.1}$$

where N is an integer. From (5.1), it can be seen that only certain frequencies can constructively add and extract the maximum gain from the cavity. These frequencies are also known as longitudinal modes of the cavity, or equivalently, cavity modes. Equation (5.1) suggests that an infinite number of modes can be supported by the cavity. However, the gain medium provides gain over a limited bandwidth (BW), and, therefore, only the modes within the gain medium's bandwidth can achieve *lasing*. A schematic spectrum of the gain profile and its limitation on the cavity modes are shown in Fig. 5.4.

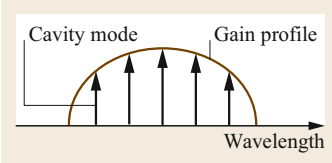


Fig. 5.4 Schematic spectrum of cavity modes and the gain medium's gain profile

From (5.1), one can also note that a change in the laser cavity length l can result in a change in the lasing wavelength(s). This is how lasers, not only FP lasers but essentially all lasers, achieve their wavelength tunability. For instance, in FP lasers, devices with different cavity lengths will have different emission wavelengths.

In a laser cavity, the quantum fluctuations associated with the lasing process affect the phase of the optical field emitted. A change in phase leads to a shift in the lasing frequency. These fluctuations effect the spectral coherence of the emitted light. A measure of this spectral coherence is the laser linewidth. The linewidth is defined as the full width at half maximum (FWHM) of the power spectral density. This is an important figure of merit for lasers. In an optical transponder, the smaller the laser linewidth, the lower the degradation to data modulation and detection. Typical linewidths of telecom lasers will be mentioned later in this section.

5.1.2 Distributed Feedback (DFB) Laser

FP lasers emit multiple wavelengths (as shown in Fig. 5.4), which are not suitable for optical transponders. An optical transponder needs lasers to emit only one wavelength (one cavity mode). One way to achieve this single mode operation is to use a Bragg grating, as depicted in Fig. 5.5. The Bragg grating, with a periodic variation of the refractive index, reflects light along the cavity and functions as a distributed mirror. This type of laser is known as a distributed feedback (DFB) laser. Additionally, as there is typically only one mode that is favored by the grating in terms of loss, the grating also functions as a wavelength selector. In a practical implementation, the mirrors at the two ends are still used, as the grating gives only $\approx 80\%$ reflectivity. The wavelength tunability of a DFB laser is achieved by thermally changing the period of the grating. A typical DFB laser diode can be tuned continuously over a few nanometers by changing the temperature of the device.

As shown in Fig. 5.5, in the DFB laser, the gain region overlaps with the wavelength selection element. This is inconvenient, as the tuning of wavelength affects the lasing process. Additionally, as the cavity length is relatively short, the linewidth of a DFB laser is usu-

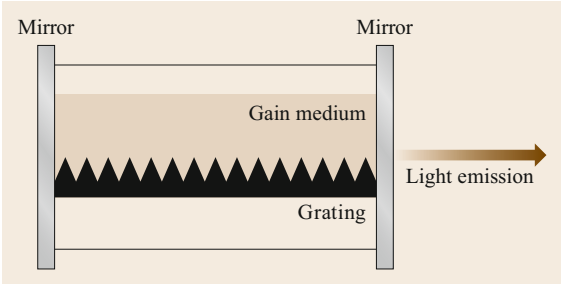


Fig. 5.5 Schematic architecture of a DFB laser

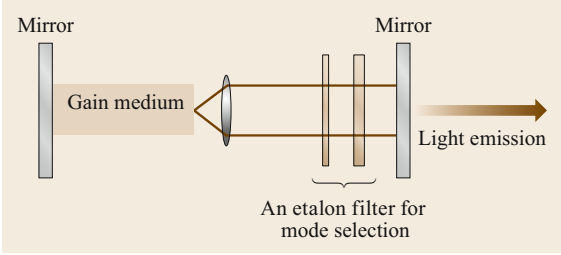


Fig. 5.6 Schematic structure of a free-space etalon filter-based ECL

ally large, e.g., a few MHz. Due to this relatively large linewidth, DFB lasers are used in some of the client-side optics but not in the line-side optics.

5.1.3 External Cavity Laser (ECL)

A better way to achieve a wider range of tunability and a narrower linewidth is to use an external cavity structure. The external cavity means that part of the laser cavity is outside the gain medium. A laser that uses an external cavity is known as an external cavity laser (ECL).

An example of an external cavity laser is shown in Fig. 5.6. The light generated from the gain medium (which consists of multiple cavity modes) is collimated and passed through an etalon filter. The etalon filter is an interference-type filter that usually consists of two thin-film plates. The transmission spectrum of the filter is adjustable by varying the distance between the two plates. This type of filter provides very narrow bandpass feature and is used to select the desired mode/wavelength. Such mode selection is illustrated in Fig. 5.7. The wavelength tunability of this type of ECL is achieved by adjusting the position of the laser cavity mirrors and the transmission spectrum of the etalon filter.

The external-cavity-based wavelength selection can also be achieved with gratings as shown in Fig. 5.8. The ECL can also be made with DFB structures. For instance, sampled Bragg gratings can be added

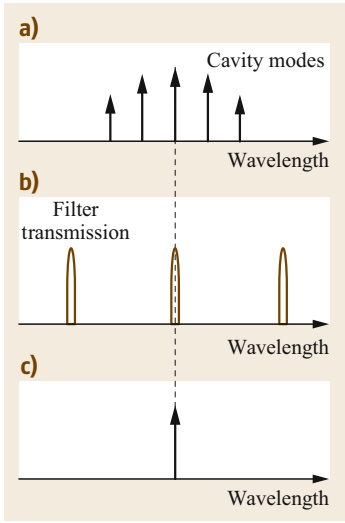


Fig. 5.7a–c Mode selection process in an ECL using an etalon filter. **(a)** Spectrum of laser cavity modes without the etalon filter; **(b)** transmission spectrum of the etalon filter; **(c)** generated laser spectrum

to the two sides of the gain region as depicted in Fig. 5.9 [5.3].

ECLs have much a narrower linewidth than stand-alone laser diodes, because the cavity length in an ECL is much longer than that of a solitary laser diode. The short gain medium in combination with a long cavity causes phase fluctuations to decrease dramatically [5.4]. With the different kinds of free-space or sampled grating ECLs, the laser linewidth is of the order of 100 kHz. The tuning range of such ECLs can be ≈ 40 nm. These ECLs are the most popular lasers used in modern optical transponders. More information about these tunable lasers can be found in [5.5].

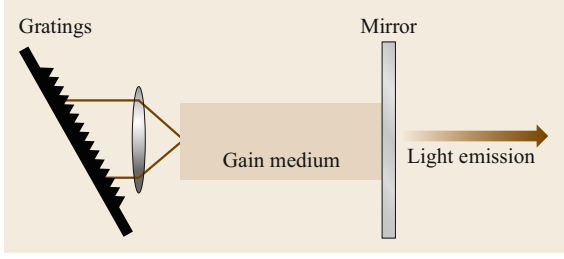


Fig. 5.8 Schematic structure of a free-space and grating-based ECL

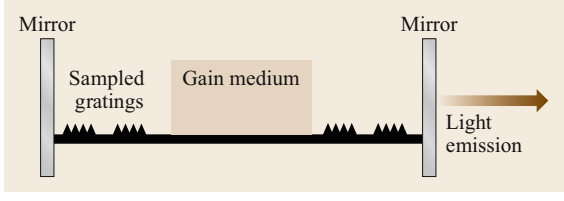


Fig. 5.9 Schematic structure of a sampled grating-based ECL

A further reduction of linewidth can be achieved with fiber lasers, where the cavity is formed by a special fiber Bragg grating. Fiber lasers can have linewidths in the region of several kHz [5.6, 7]. Very advanced fiber lasers may have linewidths of ≈ 100 Hz. Even lower linewidths, e.g., less than 1 Hz, can be achieved via self-injection locking to an optical microresonator. These ultra-low linewidth lasers are usually single frequency and not tunable, and therefore not used in optical transponders. Nevertheless, such narrow linewidth lasers could be useful in the future for generation and detection of ultra-high-order modulation formats (which are sensitive to laser phase noise).

5.1.4 Performance Merits of a Laser

There are a number of important parameters to consider when selecting a laser for optical transponders:

1. Wavelength/frequency. A telecom laser should operate at a wavelength around 1310 or 1550 nm.
2. Tunability. According to the International Telecommunication Union (ITU) standardization sector, lasers for dense wavelength-division multiplexing (DWDM) networks should have a continuous wavelength tunability across the C band (1530–1565 nm) [5.8].
3. Output power. A typical telecom laser should produce 4–40 mW of output power to have enough power for data modulation and fiber transmission.
4. Linewidth. A typical line-side optical transponder laser should have a linewidth of no more than a few hundreds of kHz. Different modulation formats have different tolerances to the laser linewidth. The linewidth's impact on data modulation and detection are rather complex. More information can be found in the literature [5.9, 10].

5.2 Modulators

A modulator is a device that can modulate a beam of light. *Modulate* means manipulating the parameter(s) of the light using an electrical signal. Depending on which parameter is manipulated, modulators can be categorized as phase modulators, amplitude modulators, polarization modulators, etc. Modulators can also be categorized by the mechanism with which the modulation is achieved. For instance, modulation can be done by direct-modulated lasers (DMLs) where the laser itself serves as a modulator; by electro-absorption modulators (EAMs), where the material absorption spectrum is changed by an applied electric field; or by electro-optic modulators where an electro-optic crystal is used, and the phase of the optical field in the crystal can be changed by an electric field. In this section, we explain DML, EAM, and one basic type of electro-optic phase modulator. Based on the phase modulator, we describe the operation of Mach–Zehnder modulators (MZMs), in-phase/quadrature (I/Q) modulators, and dual-polarization I/Q modulators.

5.2.1 Direct-Modulated Laser (DML)

A DML is a device that performs the function of both light emission and data modulation. In a DML, the laser output power is changed by changing the pump current in the laser gain medium. This pump current is controlled via an electrical driving signal, as shown in Fig. 5.10. The DML eliminates the need for coupling light between a laser and a modulator, which reduces the cost of fabrication and packaging. DMLs are suitable for applications requiring compact design and low power consumption.

The output power as a function of pump current for a typical DML is shown in Fig. 5.11. This curve is known as the light–current (L–I) characteristics of a semiconductor laser. One can observe from the L–I curve that there is a threshold current (I_{th}) required to *turn on* the laser (the onset of lasing). With a current that is larger than the threshold, the laser output power increases linearly with the pump current. However, after a certain current level, there is a *kink* on the L–I curve, indicating the existence of transitions to higher-order modes or transitions from TE to TM mode [5.2]. Operation of the DML beyond the kink is generally avoided. Therefore, as shown in Fig. 5.11, the DML has a limited current range available for data modulation. As a result, the ratio between maximum and minimum modulated optical power is limited. This power ratio is known as the modulator's extinction ratio. Typical DMLs have an extinction ratio of the order of several dBs. The ex-

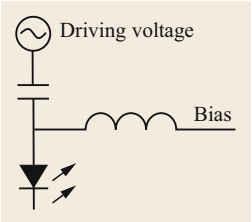


Fig. 5.10 Schematic structure of a DML

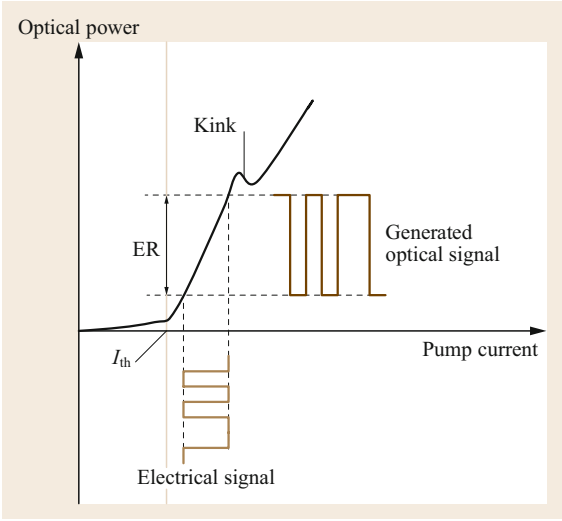


Fig. 5.11 An exemplary light–current curve of a laser. ER: extinction ratio

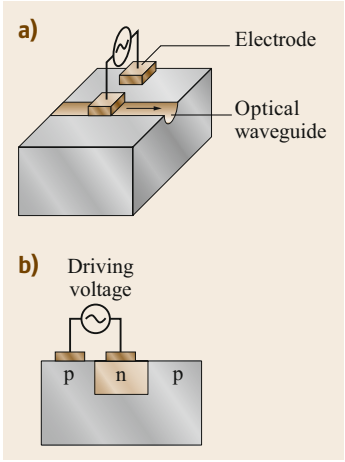


Fig. 5.12 **(a)** Schematic structure of an EAM; **(b)** cross-section of the EAM

tinction ratio is a very important figure of merit for modulators, as the data reception relies on the separation of different optical power/amplitude levels. A small extinction ratio affects the quality of the data detection.

The other characteristic associated with DMLs is the modulation chirp. The chirp of an optical pulse is the time dependence of its instantaneous emission frequency. In a DML, the pump current that modulates the laser introduces a shift in the emission frequency, typically resulting in spectral broadening. The modulation chirp degrades the signal and usually needs to be minimized.

Commercial DMLs typically modulate signals at $\approx 10\,\mathrm{Gb/s}$. Recent research demonstrations have shown DMLs with bandwidths beyond 50 GHz [5.11, 12] and capable of modulating data at $>100\,\mathrm{Gb/s}$ [5.12]. Though high-speed operation can be achieved, due to the limited extinction ratio and other constraints that originate from the fact that the modulation changes the laser properties directly (e.g., the modulation chirp), DMLs are mainly used for short reach transmission ($< 10\,\mathrm{km}$). In other words, they are used in client-side optics or data center interconnections, but are not used in a transponder's line-side optics.

5.2.2 Electro-Absorption Modulator (EAM)

External modulators are needed to achieve a better extinction ratio. An external modulator is a modulator that is independent of a laser. Such a modulator typically has a CW optical input, an RF port for adding the electrical drive signal, and an optical output for coupling the modulated light out.

An EAM is an example of an external modulator and is often made in the form of a semiconductor waveguide (a pn junction) with electrodes for applying an electric field in a direction perpendicular to the propagation direction of the light beam (Fig. 5.12). The semiconductor's optical absorption spectrum is changed by the electric field, known as the Franz–Keldysh effect [5.13]. The length of the EAM device determines the frequency response of the modulator. The device is typically short, in order to achieve high-speed operation. The electrodes that apply the field to an EAM are generally much shorter than the operating electrical wavelength. Such electrodes are referred to as a lumped electrode design.

A modulator based on electro-absorption is usually very efficient. EAMs can operate with drive voltages as low as $\approx 2\,\mathrm{V}$ and can provide bandwidths of more than 100 GHz [5.14]. Another important feature of EAMs is that they can be integrated with DFB laser diodes on the same chip, a device referred to as the electro-absorption modulated laser (EML).

Though EAMs are compact and have a large bandwidth, they can still produce a modulation chirp, and the only parameter they can modulate is the optical power. To have more flexibility in modulation and also to achieve chirp-free modulation, phase modulators and Mach–Zehnder modulators have to be used.

5.2.3 Phase Modulators

A phase modulator is another type of external modulator where the optical phase can be changed according to an applied electric voltage. To change the optical phase, an electro-optic crystal is often used. In electro-optic crystals the refractive index of the crystal can be modified via an electrical voltage, thus the phase delay of the propagating light can be manipulated. This effect

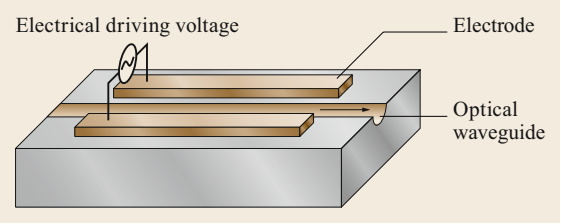


Fig. 5.13 Schematic structure of a phase modulator

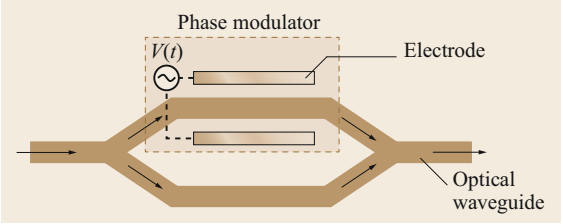


Fig. 5.14 Schematic structure of an MZI

is known as the Pockels effect. Typical crystal materials that exhibit the Pockels effect are lithium niobate ($LiNbO_3$), gallium arsenide (GaAs), and indium phosphide (InP). Devices that make use of the Pockels effect are referred to as Pockels cells.

An important property of a Pockels cell is its half-wave voltage V_π. This is the voltage required for inducing a phase change of π on the light beam. The phase-modulated light C is given by

$$C = A e^{j\omega_c t} e^{j\frac{\pi}{V_\pi}V(t)}, \tag{5.2}$$

where A and ω_c are the amplitude and the angular frequency of the laser beam, and $V(t)$ is the applied electric voltage. For simplicity, the instantaneous phase of the CW light and the waveguide propagation delay and loss are ignored.

Phase modulators may use lumped electrode design as EAMs. However, to modulate the light more efficiently, a traveling-wave configuration is often adopted. In traveling-wave configurations the electrodes are placed along the optical waveguide, and the velocity of the driving electrical wave is designed to match the velocity of the optical wave. This way, the light not only experiences phase change at one local position but accumulates it along the waveguide, and the modulation efficiency can be improved. An exemplary structure of the traveling-wave layout is shown in Fig. 5.13. Traveling-wave phase modulators typically have half-wave voltages of 3–5 V.

5.2.4 Mach–Zehnder Modulator (MZM)

Phase modulators can be used to construct power/amplitude modulators. A very popular type of amplitude modulator is the MZM.

A simple Mach–Zehnder interferometer (MZI) structure is shown in Fig. 5.14. The MZI splits the light into two arms and, after phase modulating one of the arms, recombines the two paths at the output. In this example, the two arms have exactly the same length, and the upper interferometer arm incorporates a phase modulator to add extra phase relative to the other arm. By writing the extra phase term as $e^{j(\pi/V_\pi)V(t)}$, the optical field C at the output can be expressed as

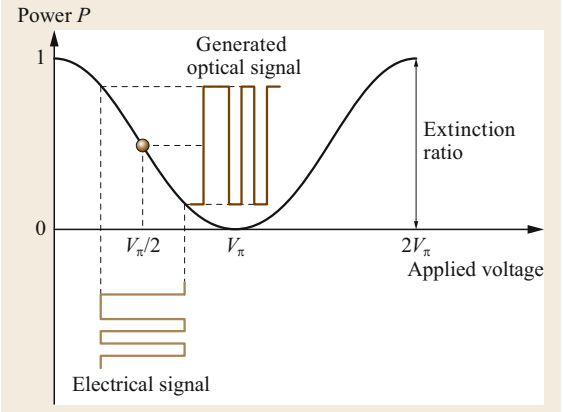


Fig. 5.15 The power transfer function of an MZM. The *vertical axis* is the MZM's output optical power normalized to the input optical power

$$C = \frac{1}{2}Ae^{j\omega_c t}\left(e^{j\frac{\pi}{V_\pi}V(t)} + 1\right) \tag{5.3}$$

and the optical power can be written as

$$P = |C|^2 = \frac{1}{2}A^2\left\{\cos\left[\frac{\pi}{V_\pi}V(t)\right] + 1\right\}. \tag{5.4}$$

It can be seen from (5.4) that the optical power can be modulated by the electrical voltage $V(t)$. The relationship between the optical power P and the electrical voltage $V(t)$ is known as the power transfer function of the modulator. This power transfer function is shown in Fig. 5.15. The power transfer function of an MZM is sinusoidal.

Theoretically, an MZM can have an infinite extinction ratio, as the minimum power should be zero when the two arms of the MZM are perfectly balanced. In practice, the extinction ratio is finite due to factors such as fabrication errors resulting in incomplete extinction. A typical telecommunication MZM has an extinction ratio between 25 and 35 dB.

An important property of an MZM is its operating point. This operating point is the point on the transfer curve around which the modulation signal $V(t)$ is

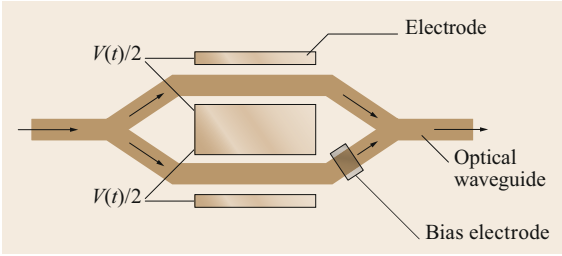


Fig. 5.16 Schematic structure of an MZM

applied and is controlled by a bias voltage. The bias voltage is a DC voltage that is added via an additional pair of lumped electrodes on one of the MZM arms, as shown in Fig. 5.16. The DC voltage induces an extra phase shift φ_{bias} to one of the MZM arms, as described in (5.5), thus shifting the operating point. The purpose of shifting the operating point is to operate the MZM in a desired region within its sinusoidal transfer function. For example, according to the power transfer function shown in Fig. 5.15, operation around $V_\pi/2$ gives the maximum linear power modulation range

$$P = |C|^2 = \frac{1}{2}A^2\left\{\cos\left[\frac{\pi}{V_\pi}V(t) + \varphi_{\text{bias}}\right] + 1\right\}. \tag{5.5}$$

As the $V_\pi/2$ bias voltage induces a 90° (a quarter of 2π rad) relative phase delay between the two arms, this bias point is known as *quadrature bias point*. The quadrature bias point is typical for client-side MZMs used for simple optical power modulation. A different bias point is needed for line-side MZMs for generation of more advanced modulation, as will be explained shortly.

In practice, inserting a phase modulator in only one of two the arms would result in a modulation chirp (this is not explained in this chapter but can be found in [5.15, 16]). In order to avoid this, most practical MZMs have phase modulation on both upper and lower arms. By applying equal voltages on the two phase modulators, the electric field in the two arms undergo phase shifts of equal magnitude. The relative phase offset between the two arms can be controlled by the DC bias, resulting in the sinusoidal transfer function. Operation of the MZM with equal drive voltages but with opposite signs is known as a push–pull operation, an intrinsically chirp-free condition. For example, if $V(t)/2$ and $-V(t)/2$ are the voltages applied to the upper and lower arms, respectively (Fig. 5.16), the optical field generated (normalized to the input CW frequency) can then be expressed as

$$C_{\text{Norm}} = \frac{1}{2}A\left(e^{j\varphi_{\text{bias}}}e^{j\frac{\pi}{V_\pi}\frac{V(t)}{2}} + e^{-j\frac{\pi}{V_\pi}\frac{V(t)}{2}}\right). \tag{5.6}$$

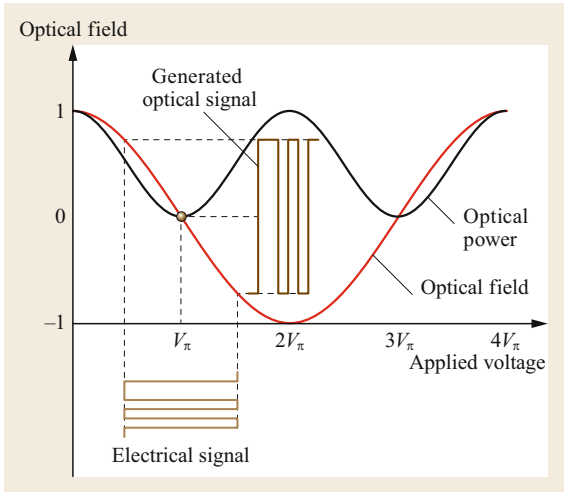


Fig. 5.17 Optical field and power transfer function of an MZM biased at null position. The *vertical axis* is the output optical field normalized to input optical amplitude

For the bias phase φ_{bias}, only the relative phase between the two arms matters, and we can rewrite (5.6) as

$$\begin{aligned} C &= \frac{1}{2}A\left(e^{j\frac{\varphi_{\text{bias}}}{2}}e^{j\frac{\pi}{V_\pi}\frac{V(t)}{2}} + e^{j-\frac{\varphi_{\text{bias}}}{2}}e^{-j\frac{\pi}{V_\pi}\frac{V(t)}{2}}\right) \\ &= A\cos\left(\frac{\pi}{V_\pi}\frac{V(t)}{2} + \frac{\varphi_{\text{bias}}}{2}\right). \end{aligned} \tag{5.7}$$

As the modern optical transmitter modulates the optical field, and the coherent receiver also recovers the field, it is more important to understand the MZM's field transfer function rather than the power transfer function. The field transfer function describes the relationship between the optical field and the electrical driving voltage. For instance, (5.3) (with normalization to input CW frequency) is the field transfer function for the MZM shown in Fig. 5.14. It is a complex-valued function, and its dependence on the driving voltage is difficult to depict in two dimensions. However, with (5.7), the field transfer function becomes real valued and can easily be plotted as shown below (Fig. 5.17). For comparison, the optical power is also plotted in Fig. 5.17.

It can be seen that the best linear range for field modulation is achieved when operating around V_π. This is the bias point that line-side MZMs use. As this bias voltage corresponds to zero optical power when there is no electrical signal, the bias position is also known as the *null bias position*.

Most MZMs used in optical transponders are typically $LiNbO_3$ MZMs. The optical waveguide in those MZMs is made via ion diffusion, and the waveguide is usually bulky (> 5 cm long). The driving electrodes have to be placed far from each other, resulting in less

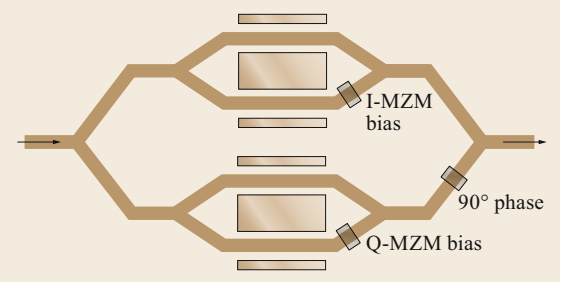


Fig. 5.18 Schematic structure of an I/Q modulator

efficient modulation at high frequencies. Such $LiNbO_3$ MZMs typically have a 3 dB bandwidth of ≈ 35 GHz and a V_π of ≈ 3.5 V [5.17]. There are new MZMs using thin-film $LiNbO_3$ that have a much higher refractive index, and the electrodes can be placed closer, which offers ultra-high modulation bandwidth. Research demonstrations have shown thin-film based $LiNbO_3$ MZMs with 3 dB bandwidths beyond 100 GHz and a V_π voltage of about 4 V [5.18]. Ultra-high speed modulators have also been demonstrated on other platforms such as InP [5.19] and silicon-organic hybrid materials [5.20].

5.2.5 Dual-Polarization I/Q Modulators

So far, we have mostly discussed optical power/amplitude modulators. However, a light wave, like all other electromagnetic waves, is characterized not only by amplitude but also by phase. The amplitude and phase are two orthogonal dimensions and can be multiplexed to encode more information onto one light wave (more information can be found in [5.21–24]). The amplitude and phase can also be represented as the real and imaginary parts of the complex field. The real and imaginary parts are known as sine and cosine components, or alternatively, in-phase and quadrature (I/Q) parts of the optical field.

A modulator that modulates the in-phase and quadrature parts of the optical field is known as an I/Q modulator. (*Nested MZM*, *vector modulator*, and *Cartesian modulator* are some other names used to describe the I/Q modulator.) The I/Q modulator consists of two MZM-based amplitude modulators for modulating the real and imaginary parts, and the modulated light from the two MZMs are combined with a 90° phase difference to form the complex field. The architecture of an I/Q modulator is shown in Fig. 5.18. Besides the two bias electrodes for the two MZMs, there is a third electrode on one of the MZM outputs to adjust the 90° phase for the field.

Like the I and Q parts, the two orthogonal polarizations of an electric field can also be used for information encoding. To modulate the optical field on both polarizations, dual-polarization I/Q modulators are needed. Dual-polarization I/Q modulators are made by splitting the optical carrier and modulating two independent complex streams via two I/Q modulators, and then combining the two modulated fields via a polarization beam combiner (PBC). A schematic architecture of a dual-polarization I/Q modulator is shown in Fig. 5.19. Such dual-polarization I/Q modulators are used in advanced line-side optical transponders.

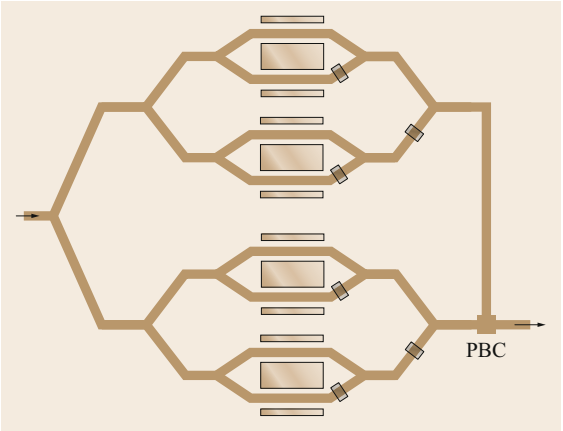


Fig. 5.19 Schematic structure of a dual-polarization I/Q modulator

5.2.6 Figure of Merits of a Modulator

In addition to the bandwidth (BW) and the V_π, the insertion loss (IL) and the footprint are also important figures of merit that characterize modulators for an optical transponder. The IL includes the optical waveguide propagation loss and fiber coupling loss. The insertion loss is measured by comparing the input and output light power when the modulator is biased for maximum transmission. An acceptable insertion loss is ≈ 5 dB.

In traditional commercial line-side optical transmitters, bulk $LiNbO_3$ modulators are used. They have 3 dB bandwidths around 35 GHz and have been used for 100, 200, and 400 Gb/s data modulation. The device length of the such modulators is typically 5−7 cm. To reduce the footprint and to increase the bandwidth, modulators that are fabricated on other platforms or with different materials are being investigated. For instance, thin-film $LiNbO_3$ modulators have been demonstrated with very high bandwidths [5.18]. Modulators made with InP and SiPh technologies offer a very small footprint [5.25, 26], and commercial products are available. A comparison of the characteristics of the different modulator technologies is shown in Table 5.1. The values shown in this table are typical, as quoted from the references, and are constantly being improved.

Table 5.1 Comparison of different technologies for phase modulators and MZMs

	3 dB bandwidth (GHz)	V_π (V)	Insertion loss (dB)	Length (cm)	Reference
$LiNbO_3$	≈ 35	≈ 3.5	≈ 5	≈ 5	[5.27]
Thin-film $LiNbO_3$	≈ 100	≈ 4.4	0.5[a]	0.5	[5.18]
InP	> 70	≈ 1.5	≈ 9	0.8	[5.25]
SiPh	≈ 25	≈ 5	≈ 6[a]	0.4	[5.26]

[a] On-chip insertion loss (fiber coupling loss) is not reported. (The other two insertion losses reported in this table include both fiber coupling loss and in-chip loss.)

5.3 Photodetectors

A photodetector (PD) converts an optical signal into an electrical signal via the process of photon absorption followed by electron–hole pair generation. There are several types of PDs, including the Schottky-barrier (SB) diode, the metal-semiconductor-metal (MSM) photodiode, the p-i-n photodiode, the avalanche photodiode (APD), and the heterojunction phototransistor (HPT). The most commonly used PD in optical transponders is the p-i-n photodiode. Common semiconductor material(s) used to fabricate the p-i-n photodiode are silicon (Si), GaAs, germanium (Ge), InP, and indium gallium arsenide (InGaAs). The material characteristics determine the photodiode's spectral response. Generic performance metrics of a photodiode include responsivity, bandwidth, dark current, and nonlinearity.

The responsivity is an expression that reflects how efficiently the PD can convert the optical power into electrical current. In other words, the responsivity is defined as the ratio of this photocurrent divided by the total incident light power falling upon the PD. The symbol for responsivity is R, and the unit is Ampere/Watt. The responsivity of the PD for each of the different materials listed above is typically a function of the wavelength (of frequency) of the incident optical signal.

The bandwidth of a PD defines how fast it responds to a time-varying incoming optical signal. When a light pulse impinges on a PD, the time needed for the photocurrent to go from 10 to 90% of the final output is defined as the response time. In frequency domain, the frequency at which the detected RF power drops to half its value at zero frequency is termed the 3 dB bandwidth of the PD. The response parameters strongly depend on the physical dimensions of the PD to a first order, as they determine the device parasitics such as capacitance, resistance, and transit time. Higher-bandwidth devices would generally be smaller in size.

As the PD has a junction (a p-n junction in the semiconductor case), the imperfections at the junction result in a leakage current. This current flowing through a PD even in the absence of light is termed dark current. The unit of dark current is A. The dark current manifests itself as noise and degrades the detection process. All well-designed PDs have dark currents in the pA to nA range. The frequency domain representation of the dark current is known as the noise spectrum of the PD.

Nonlinearity is a measure of the response (specifically the responsivity) of the PD as a function of the incident optical power (or incident photons) at a given modulation frequency. When the incident optical power is small, the response is generally linear. However, when the incident power is large, the large number of photon-generated carriers build up a space charge within the PD, resulting in a saturation of the response. The onset of saturation depends on the device structure and the frequency of the modulated signal. Generally, p-i-n photodiodes are more linear than APDs, where the multiplication process introduces nonlinearity in the response.

5.3.1 Normal-Incidence PIN Photodiodes

In Fig. 5.20, the p-i-n photodiode is a p-doped semiconductor and an n-doped semiconductor with an intrinsic

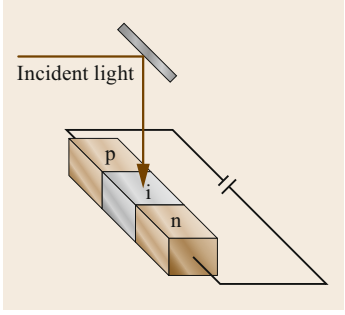


Fig. 5.20 Schematic structure of a normal-incident p-i-n photodiode

layer of the semiconductor sandwiched between them. A reverse bias is applied to the p-n junction, providing an electric field that sweeps out the photo-generated carriers to the contacts, where it appears as photocurrent.

In a conventional PD design, the light enters perpendicular to the semiconductor surface (Fig. 5.20) passing through the substrate. Such light coupling is called normal-incidence coupling. Absorption of light in normal-incidence PDs takes place in the thin intrinsic layer. High-speed response of these PDs is obtained by reducing the thickness of the intrinsic region (so that the carrier transit time can be minimized). However, this comes at the expense of two impairments. The capacitance of the PD increases when the intrinsic layer thickness is reduced, resulting in an RC limited frequency response. In addition, the volume for light absorption is reduced, resulting in a decrease in the responsivity. As a result, normal-incidence PDs have an inherent tradeoff between the responsivity and the bandwidth [5.28]. A responsivity of 0.5 A/W with a 50 GHz bandwidth is typical for such a PD.

5.3.2 Waveguide PIN Photodiodes

A method to circumvent the responsivity-bandwidth tradeoff seen in normal incidence PDs is to partially decouple the responsivity from the device parasitics. This is achieved using waveguide coupled p-i-n PDs. Bandwidths higher than 50 GHz have been demonstrated using such a waveguide coupling structure [5.29–31]. A waveguide PD utilizes an input optical waveguide with embedded absorbing layers, as shown in Fig. 5.21. The light is absorbed when it propagates through the intrinsic layer. The propagating length in such devices can be made much larger than the thickness of the intrinsic layer in a normal incidence PD, resulting in large responsivity. At the same time, the thickness of the waveguide layer can be designed to be small to reduce the transit time of the photon-generated carriers, within some bounds. The combined design approach minimizes the bandwidth-responsivity tradeoff seen in normal incidence PDs, yielding higher bandwidth PDs. Waveguide PDs with 3 dB bandwidths greater than 100 GHz and with responsivity of about 0.5 A/W have been demonstrated.

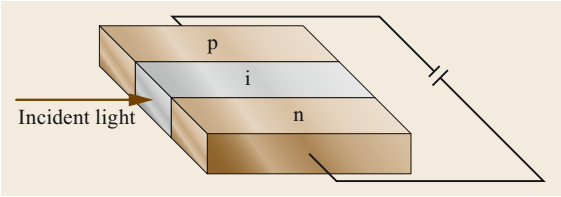


Fig. 5.21 Schematic structure of a waveguide photodiode

5.3.3 Avalanche Photodiodes (APD)

The APD is a special case of the p-i-n PD where one of the photo-generated carriers (usually electrons) is accelerated by a strong electric field to result in a multiplication, leading to an avalanche of carriers. The APD thus provides gain as compared to a simple p-i-n PD through the avalanche multiplication process. An APD is typically operated near its reverse bias break-down voltage (for a high electric field) to realize gains between 5 and 10 (in linear units). APDs are generally used where high receiver sensitivity is needed.

5.4 Transmitter Hardware

An optical transmitter, as shown in Fig. 5.22, comprises a laser that generates a light beam and a dual-polarization I/Q modulator for imprinting the information onto the light beam. The information, before optical modulation, is represented by electrical voltages. The voltages are generated by DACs, which receive the information from the DSP engine. There are linear RF amplifiers between the DACs and the modulator to amplify the DAC output (which is less than 1 V) to the modulator's required voltage (3–5 V). The laser is typically an ECL as described in Sect. 5.1. The dual-polarization I/Q modulator has the same architecture as described in Sect. 5.2.5. In this section, we explain the DACs and their drivers.

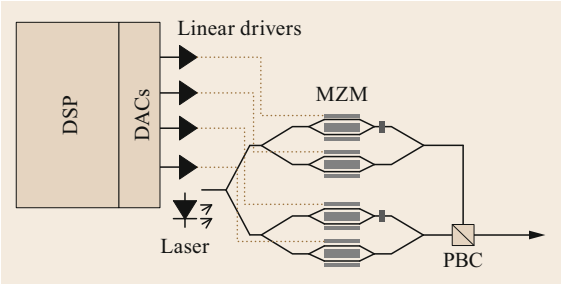


Fig. 5.22 Schematic architecture of an optical transmitter for the line side

5.4.1 Digital to Analog Converters (DAC)

A DAC converts a finite-precision number into a physical quantity. In the telecommunications context, a DAC converts a finite-precision binary series to a continuously time-varying voltage. In modern optical communications, DACs are the key to the generation of arbitrary electrical waveforms that are needed for advanced multilevel modulation formats.

To understand DACs, we start with the simple 2 bit DAC shown in Fig. 5.23a. To represent a ramp signal, e.g., the dashed waveform shown in Fig. 5.23b, we can describe the waveform (at each time instance) as two binary values (Fig. 5.23c). The 2 bit DAC will take two values (D_0 and D_1) at a time and convert them to a voltage. The voltage is generated and held until the next voltage is produced. An example of the 2 bit DAC generated ramp signal is shown as the solid line in Fig. 5.23b. As this DAC uses 2 bit at a time to produce one voltage, it has a digital-to-analog conversion resolution of 2 bit.

Typical DACs in deployed line-side transmitters have resolution from 6 to 8 bit. The more resolution a DAC has, the better it can represent an arbitrary waveform. A limited number of bits results in quantization noise. The quantization noise is a fundamental digital-to-analog conversion noise, representing the information loss when using finite-precision binary numbers to represent an arbitrary waveform. The more quantization bits a DAC has, the lower the quantization noise. Besides the quantization errors, there are other noise and distortions in a DAC that are induced by electronic circuits. Such noise and distortions, along with the fundamental quantization noise, lead to the definition of the effective number of bits (ENOB). The ENOB is the equivalent number of bits for practical DACs that includes quantization noise and electric circuit-induced noise and distortions. Typical ENOB values range between 5 and 6 bit. The ENOB is also frequency dependent, as electronic circuit noise changes with frequencies. In general, ENOB decreases with frequency.

Besides the resolution, the sampling speed is another figure of merit of DACs. The sampling speed tells how fast the DAC can generate a new voltage. In the example shown in Fig. 5.23, the DAC can generate a new voltage every Δt seconds. This means the sampling rate of the DAC is $Sa = 1/\Delta t$ samples/s (Sa/s). According to the sampling theorem [5.32, 33], the sampling rate Sa bounds the highest possible analog bandwidth of the DAC ($Sa/2$ Hz for this example). Quite often, instead of the sampling rate, the electronic circuits are more likely to limit the actual bandwidth. Practical DACs realized from complementary metal oxide semiconductor (CMOS) technologies generally have 3 dB bandwidths that are smaller than $Sa/2$.

DACs with a higher sampling rate usually have lower resolutions due to the challenges in electronic circuit design for higher frequencies. For example, DACs with a sampling speed lower than 10 GSa/s can have more than 10 bit resolution. DACs that are being used in line-side transponders, which operate at $\approx$ 100 GSa/s, usually have a nominal resolution of 8 bit and an ENOB of 5–6 bit. For commercially available CMOS-integrated DACs, the 3 dB bandwidths are usually around 20 GHz. Signals with more than 20 GHz bandwidth can still be generated with such DACs if proper pre-emphasis via digital signal processing is used, at the expense of ENOB. Silicon germanium (SiGe)-based DACs can have a higher speed, e.g., with a 3 dB bandwidth of $\approx$ 40 GHz. In laboratory demonstrations where discrete components (binary sources, electrical multiplexers, etc.) are used, DACs can have sampling rate of up to 204 GSa/s with 1 bit resolution [5.34, 35]. Comparable sampling speeds but higher resolutions can be achieved via electrical frequency multiplexing. With such multiplexing (also referred to as the digital band interleaving technique), DACs with a sampling rate of 240 GSa/s and a bandwidth of 100 GHz with a nominal resolution of 8 bit can be achieved [5.36, 37].

Besides the bandwidth and resolutions, there are other figures of merit associated with DACs, such as the spurious-free dynamic range, signal-to-noise ratio, and

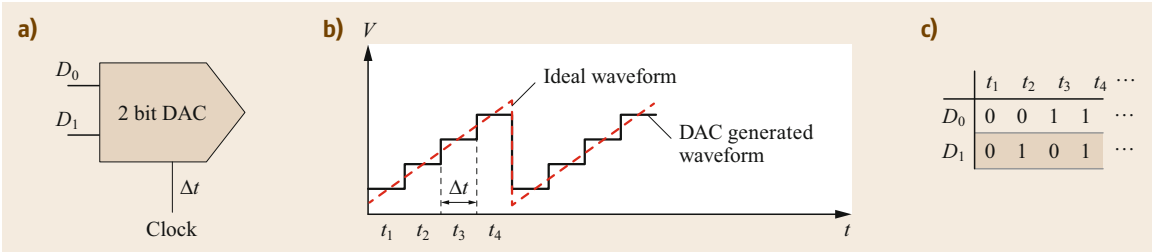


Fig. 5.23a–c An example of a 2 bit DAC. **(a)** Schematic diagram for a 2 bit DAC. **(b)** example of an ideal ramp signal, and a 2 bit DAC generated ramp signal. **(c)** description of the ramp signal in the form 2 binary bits

quantization errors, all of which are extensively treated in textbooks [5.38, 39].

5.4.2 Linear Driver Amplifier

Linear RF amplifiers between the DACs and the optical modulator amplify the DAC output voltage to the voltage required to drive the modulator. The typical gain of these drivers is $\approx 20\,\mathrm{dB}$. Linearity is an important aspect for these drivers, as the waveforms generated from the DACs are arbitrary with many levels, and the amplifiers should preserve the characteristics of the waveforms. Departure from linearity would result in distortions that are passed on to the modulated optical signal. In the case that the amplifiers are not linear enough, mitigation is possible in the DSP engine that precedes the DAC, albeit at the expense of the available DAC resources (dynamic range, ENOB, etc.).

As described in Sect. 5.2.5, the dual-polarization I/Q modulator, typically used for line-side signal generation, has four MZMs and requires four input RF waveforms (namely, Ix, Qx, Iy, and Qy). Therefore, four-in-one linear driver arrays are commonly used for such applications. The quad-drivers must have well-matched characteristics such as gain, analog bandwidth, and noise figure. The ability to control the gain of each driver is also important for the optimization of the symmetry of the I–Q constellations.

5.5 Coherent Receiver Hardware

The modern optical modulator modulates both amplitude and phase of the light, as we learnt in Sect. 5.2.5. This requires the receiver to be capable of recovering both the amplitude and phase. However, one PD with power detection squares the received optical field $C_{\mathrm{SIG}} = A\mathrm{e}^{\mathrm{j}\varphi}\mathrm{e}^{\mathrm{j}\omega_{\mathrm{c}}t}$ as

$$E = |C_{\mathrm{SIG}}|^2 = \left|A\mathrm{e}^{\mathrm{j}\varphi}\mathrm{e}^{\mathrm{j}\omega_{\mathrm{c}}t}\right|^2 = A^2\,, \tag{5.8}$$

where ω_{c} is the angular frequency of the optical carrier; A and φ are, respectively, the amplitude and phase of the modulated light. Both A and φ contain information and need to be recovered at the receiver. It can be seen that single PD power detection captures only the power/amplitude of the light and loses the phase information. To receive the phase information, various receivers can be implemented. For instance, one can encode the signal differentially at the transmitter and use a 1 bit delay receiver at the receiver. One can also use self-coherent receivers where CW light is transmitted with the signal and is used to mix with the signal at the receiver to recover the phase.

More conveniently, for high-speed optical transponders, coherent receivers are used. The coherent receiver uses an optical local oscillator (OLO) $C_{\mathrm{LO}} = R\mathrm{e}^{\mathrm{j}\omega_{\mathrm{LO}}t}$ to mix with the signal. The mixed light is detected by the PD,

$$\begin{aligned} E &= \left|\frac{C_{\mathrm{SIG}} + C_{\mathrm{LO}}}{\sqrt{2}}\right|^2 \\ &= \frac{R^2}{2} + \mathrm{Re}\left(RA\mathrm{e}^{\mathrm{j}\varphi}\mathrm{e}^{\mathrm{j}\Delta\omega t}\right) + \frac{A^2}{2}\,, \end{aligned} \tag{5.9}$$

where R and ω_{LO} are the amplitude and angular frequency of the OLO. The instantaneous phase of the OLO is ignored for simplicity; $\Delta\omega = \omega_{\mathrm{c}} - \omega_{\mathrm{LO}}$ is the angular frequency difference of the signal carrier and the OLO; $\mathrm{Re}(\cdot)$ stands for the real part of a complex number. The factor $1/\sqrt{2}$ is brought by a 3 dB power loss from an optical coupler. In (5.9), the first term is a DC term that is easy to remove using a DC block, and the second term is the signal of interest. The third term is a square of the signal and is an unwanted term. Note that the third term is a broadband distortion and spectrally overlaps with the signal of interest, thus can not be removed by a DC block. Keeping the power of the OLO much larger than the signal, the second term will dominate over the third term.

In this section, we start with the simplest coherent receiver that realizes (5.9). We then describe more complex versions of coherent receivers that can improve the performance. In the end, we will explain coherent receivers being used in most advanced coherent transponders.

5.5.1 2×2-Coupler-Based Coherent Receivers

Combining an OLO with the signal can be realized by a simple 2×2 optical coupler followed by a PD (Fig. 5.24), and is known as a 2×2-coupler-based coherent receiver. A PD placed at one output of the coupler performs the detection, while the second output can be used to provide feedback for phase locking the OLO to the incoming signal.

Adding a second PD at the other output and combining the two PDs outputs in a differential manner, as shown in Fig. 5.25, can remove the first and third term of (5.9), leaving only the second term (with twice the magnitude compared to (5.9)) yielding a better per-

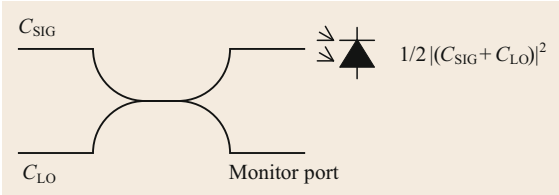


Fig. 5.24 Schematic architecture of a 2×2-coupler-based coherent receiver

formance. With this structure, the power ratio between the OLO and the signal can therefore be reduced compared to the architecture shown in Fig. 5.24. This PD pair is known as a balanced photodetector (BPD). The 2×2 coupler, as it mixes the signal and OLO and produces the second copy of the OLO with a 180° phase shift, can be viewed as a 180° optical hybrid

$$E = \left|\frac{C_{SIG}+C_{LO}}{\sqrt{2}}\right|^2 - \left|\frac{C_{SIG}-C_{LO}}{\sqrt{2}}\right|^2 = 2\mathrm{Re}\left\{RAe^{j\varphi}e^{j\Delta\omega t}\right\}. \quad (5.10)$$

To explain the required number of PDs and ADCs for the received signal, we rewrite the detected signal in (5.10) as $E = 2RAe^{j\varphi}e^{j\Delta\omega t} + 2RAe^{-j\varphi}e^{-j\Delta\omega t}$. From this expression, it can be seen that the detected signal is composed of the down-converted signal and its conjugate. According to the sampling theorem, given only one ADC, the signal and its conjugate should be spectrally separated to avoid loss of information. This means that the $\Delta\omega$ then has to be larger than half the bandwidth of the signal. In other words, this *spectral separation* means the OLO should be placed outside the signal spectrum. (More explanations on the sampling theorem can be found in the literature [5.28, 29].) Such a type of coherent receiver, with the OLO placed outside the signal spectrum, is known as a *heterodyne receiver*. A heterodyne receiver requires the PD and ADC's bandwidths be equal to (or larger than) the optical signal's bandwidth.

5.5.2 90° Optical Hybrid Based Coherent Receivers

The other way to avoid a loss of information is to use a 90° hybrid plus a pair of lower-speed PDs and ADCs. In this setting, the real and imaginary parts (or I and Q parts) of the complex signal can be received by two separate PDs and ADCs. The 90° hybrid is a special type of coupler that has two input ports and four output ports. As shown in Fig. 5.26, the 90° hybrid mixes the incoming signal with the four quadrature states associated with the OLO. The four mixed signals are, respectively, $(C_{SIG}+C_{LO})/2$, $(C_{SIG}-C_{LO})/2$, $(C_{SIG}+jC_{LO})/2$, $(C_{SIG}-jC_{LO})/2$, and they are grouped into two pairs and sent to two BPDs. The output of the two BPDs, E_1 and E_2, can be expressed as the following

$$E_1 = \left|\frac{C_{SIG}+C_{LO}}{2}\right|^2 - \left|\frac{C_{SIG}-C_{LO}}{2}\right|^2 = 2\mathrm{Re}\left\{RAe^{j\varphi}e^{j\Delta\omega t}\right\}, \quad (5.11)$$

$$E_2 = \left|\frac{C_{SIG}+jC_{LO}}{2}\right|^2 - \left|\frac{C_{SIG}-jC_{LO}}{2}\right|^2 = 2\mathrm{Im}\left\{RAe^{j\varphi}e^{j\Delta\omega t}\right\}, \quad (5.12)$$

where $\mathrm{Im}\{\cdot\}$ stands for the imaginary part of a complex value. It can be seen that the two BPDs receive the real and imaginary parts of the optical field. In this case, $\Delta\omega$ is close to zero, and this type of coherent receiver is known as an intradyne receiver. In a very special case, when the $\Delta\omega$ equals zero, the receiver is called the homodyne receiver.

5.5.3 Polarization-Diversity Coherent Receiver

In the approach described earlier, the signal was considered to have one polarization state. For efficient mixing, the OLO's polarization must be aligned with that of the signal. One can use a feedback mechanism to track the signal polarization and tune the OLO accordingly. Such tracking is usually fairly complex. An alternate option is to extend the coherent receiver design to a polarization diversity design, as shown in Fig. 5.27. With this approach, there is no need to track the polarization states. Moreover, such a receiver is even capable of receiving polarization-multiplexed signals.

As shown in Fig. 5.27 in a polarization diversity coherent receiver, a polarization beam splitter (PBS) splits the incoming signal into the two orthogonal polarizations (namely, x and y), while a 3 dB coupler splits the

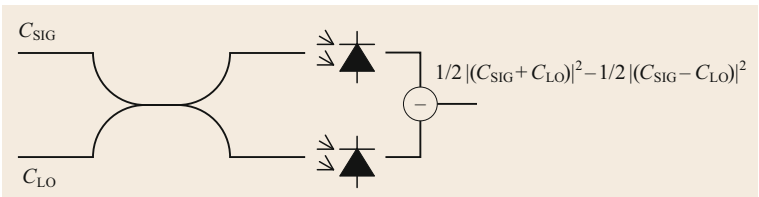


Fig. 5.25 Schematic architecture of a 2×2-coupler-based coherent receiver with a balanced photodetector

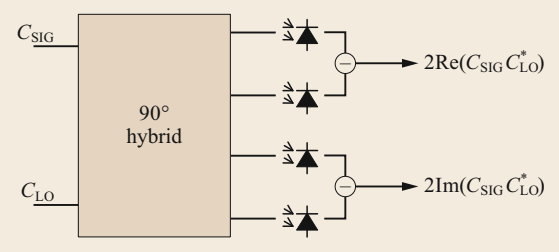


Fig. 5.26 Schematic architecture of a 90° hybrid based coherent receiver

OLO into two copies. Two sets of 90° hybrids are then used to recover the x and y electric fields, respectively. The two sets of the hybrid, the PBS and the OLO power splitter, together are known as the *polarization diversity hybrid*. This hybrid has eight output ports, four for each polarization. A total of four BPDs (eight PDs) in addition to their amplifiers complete the receiver's optical-to-electrical (O/E) front-end.

It is important to mention that this polarization diversity coherent receiver is capable of receiving all four multiplexing dimensions (namely, two quadratures and two polarizations) of an optical signal. This is then the universal coherent receiver that is used to recover any type of information encoding of the optical field.

5.5.4 Transimpedance Amplifier (TIA)

The electrical signal at the output of a BPD (or a PD as the case may be) is typically small. The BPD is usually followed by a low-noise electronic amplifier to raise the electrical signal to a sufficient level for ADCs. This amplifier transfers the detected photocurrent I_{ph} from PDs into a proportionate output voltage V_{pp}, and is known as a transimpedance amplifier (TIA). A TIA is characterized by its transimpedance Z, given by $Z = V_{pp}/I_{ph}$. Typical values for Z are in the range from 1 to 3 MΩ. TIAs also typically have automatic gain control (AGC) to allow for delivering a fixed output voltage independent of the input optical power to the photodiode. Since the PDs are linear in detecting multilevel waveforms (as described in Sect. 5.3), the following electrical amplifiers should also preserve that linearity. TIAs are also characterized by their gain, bandwidth, and electronic circuit noise. Typical TIAs that are used in coherent optical receivers have bandwidth of 30–40 GHz.

5.5.5 Analog to Digital Converter (ADC)

The signal detected by the coherent receiver is an analog signal that needs to be digitized and then digitally processed to recover the encoded information. ADCs are the devices that convert the analog signals to digital signals. This device essentially performs the reverse function of the DAC described in Sect. 5.4.1. It converts a continuous-time and continuous-amplitude analog signal to discrete-time and discrete-amplitude digital samples. Similar to the DACs, the main characteristics of ADCs are the sampling rate, analog bandwidth, and ENOB. For a coherent receiver, four channel ADCs are integrated on a single chip, allowing for synchronized reception.

CMOS-integrated ADCs nowadays typically have ≈ 128 GSa/s sampling rate with 8 bit nominal resolution. Higher speeds have been achieved with other technologies such as InP. For instance, InP-based ADCs in the form of real-time sampling scope can have

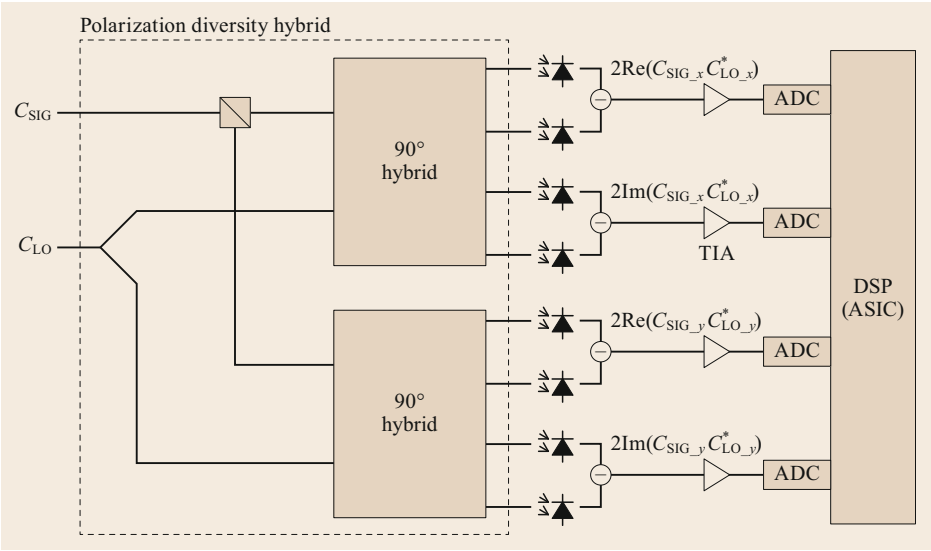


Fig. 5.27 Schematic architecture of a 90° hybrid based coherent receiver with polarization diversity

256 GSa/s with nominal bits of 8 and an ENOB of ≈ 6 bit [5.30]. CMOS ADCs with a digital band interleaving technique and also in the form of real-time sampling scope can have 240 GSa/s [5.33]. More reviews of the ADCs and their role in the signal reception can be found in [5.40].

5.6 Integrated Transmitters and Receivers

When the optical transmitter and receiver are implemented with discrete components, there are many optical and electrical interfaces. For instance, in an optical transmitter, there are 3 optical interfaces (laser output, modulator input, and modulator output) and at least 16 RF interfaces (4 DAC outputs, 4 driver inputs, 4 driver outputs, and 4 modulator inputs). In a polarization diversity coherent receiver, there are a total of 18 optical interfaces (2 optical hybrid inputs, 8 hybrid outputs, and 8 PD inputs). This large number of optical and RF interfaces makes the task of minimizing impairments (reflections, delay mismatch, and bandwidth loss) very challenging when the system is operating at high speed. These impairments degrade the transponder's ability to transmit/recover complex modulation formats with good performance. A few of these impairments, such as delay mismatch could be precompensated in the DSP engine. Reflections and bandwidth loss are difficult to address without sacrificing complexity and energy. However, when all the components are integrated in a small form factor, and the number of interfaces are minimized, these impairments can be mitigated to a great extent. Integrated coherent transmitters (ICT) and integrated coherent receivers (ICR) are typically realized via one of three available technologies. These are all-InP technology, micro-optics with InP technology, and silicon photonics (SiPh) technology with external laser sources. More information about these technologies can be found in papers and textbooks [5.41–43].

The integrated transceivers are packaged into a small-form factor module. Modern line-side optical transponders use C form-factor pluggable (CFP) modules. The CFP is a multisource agreement to produce a common form-factor for the transmission of high-speed digital signals. There are a number of variations of CFP, including CFP, CFP2, and CFP4. Among them, the CFP2 module is being used for 100 Gb/s and 200 Gb/s transceivers. It measures 41.5 mm (width) × 12.4 mm (height) × 107.5 mm (depth) and allows for a maximum of 12 W power usage. The CFP2 module can be a complete coherent transceiver that has the analog front-end and DSP module. This kind of CFP2 transceiver is called CFP2-DCO, where the DCO stands for *digital coherent optics*. The CFP2-DCO can be plugged into a slot that is designed for a CFP and communicates through the socket with digital signals. A CFP2 module can include only the analog parts and not have a DSP module. This type of module is called the CFP2-ACO, where ACO stands for *analog coherent optics*. The CFP2-ACO communicates with the circuit board by sending analog signals across a special connector, and the DSP engine resides outside the CFP2 module.

References

5.1 G.P. Agrawal, K.D. Niloy: *Semiconductor Lasers* (Springer, Berlin, Heidelberg 2013)

5.2 H.C. Casey Jr., M.B. Panish: *Heterostructure Lasers, Part A: Fundamental Principles* (Academic Press, New York 1978)

5.3 V. Jayaraman, Z.M. Chuang, L.A. Coldren: Theory, design, and performance of extended tuning range semiconductor lasers with sampled gratings, IEEE J. Quantum Electron. **29**, 1824–1834 (1993)

5.4 M. Fleming, A. Mooradian: Spectral characteristics of external-cavity controlled semiconductor lasers, IEEE J. Quantum Electron. **17**, 44–59 (1981)

5.5 J. Buus, E.J. Murphy: Tunable lasers in optical networks, J. Lightwave Technol. **24**, 5–11 (2006)

5.6 C. Spiegelberg, J. Geng, Y. Hu, Y. Kaneda, S. Jiang, N. Peyghambarian: Low-noise narrow-linewidth fiber laser at 1550 nm, J. Lightwave Technol. **22**, 57–62 (2004)

5.7 J. Geng, S. Staines, Z. Wang, J. Zong, M. Blake, S. Jiang: Highly stable low-noise Brillouin fiber laser with ultranarrow spectral linewidth, IEEE Photonics Technol. Lett. **18**, 1813–1815 (2006)

5.8 ITU-T G.694.1: Spectral grids for WDM applications: DWDM frequency grid (2012)

5.9 M. Fleming, A. Mooradian: Spectral characteristics of external-cavity controlled semiconductor lasers, IEEE J. Quantum Electron. **17**, 44–59 (1981)

5.10 T. Pfau, S. Hoffmann, R. Noé: Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations, J. Lightwave Technol. **27**, 989–999 (2009)

5.11 W. Kobayashi, T. Ito, T. Yamanaka, T. Fujisawa, Y. Shibata, T. Kurosaki, M. Kohtoku, T. Tadokoro, H. Sanjoh: 50-Gb/s direct modulation of 1.3-μm InGaAlAs based DFB laser with ridge waveguide structure, IEEE J. Sel. Top. Quantum Electron. **19**, 1500908 (2013), https://doi.org/10.1109/jstqe.2013.2238509

5.12 Y. Matsui, T. Pham, W.A. Ling, R. Schatz, G. Carey, H. Daghighian, T. Sudo, C. Roxlo: 55-GHz bandwidth short-cavity distributed reflector laser and its application to 112-Gb/s PAM-4. In: *Proc. Opt. Fiber Commun. Conf.* (2016), Paper PDP Th5B.4

5.13 J.I. Pankove: *Optical Processes in Semiconductors* (Prentice Hall, Englewood Cliffs 1971), Chap. 2-C-2

5.14 M. Chacinski, U. Westergren, B. Stoltz, L. Thylen, R. Schatz, S. Hammerfeldt: Monolithically integrated 100 GHz DFB-TWEAM, J. Lightwave Technol. **27**, 3410–3415 (2009)

5.15 F. Koyama, K. Iga: Frequency chirping in external modulators, J. Lightwave Technol. **6**, 87–93 (1988)

5.16 H. Kim, A.H. Gnauck: Chirp characteristics of dual-drive. Mach–Zehnder modulator with a finite DC extinction ratio, IEEE Photonics Technol. Lett. **14**, 298–300 (2002)

5.17 T. Kawanishi, T. Sakamoto, M. Izutsu: High-speed control of lightwave amplitude, phase, and frequency by use of electrooptic effect, IEEE J. Sel. Top. Quantum Electron. **13**, 79–91 (2007)

5.18 M. Zhang, C. Wang, X. Chen, M. Bertrand, A. Shams-Ansari, S. Chandrasekhar, P. Winzer, M. Lončar: Ultra-high bandwidth integrated Lithium Niobate modulators with record-low V_π. In: *Proc. Opt. Fiber Commun. Conf.* (2018), Paper Th4A.5

5.19 Y. Ogiso, T. Yamada, J. Ozai, N. Kashio, N. Kikuchi, E. Yamada, H. Mawatari, H. Tanobe, S. Kanazawa, H. Yamazaki, Y. Ohiso, T. Fujii, M. Ishikawa, M. Kohtoku: Ultra-high bandwidth InP IQ modulator with 1.5 V_π. In: *Eur. Conf. Opt. Commun.* (2016), Paper Tu.3.A.2

5.20 S. Wolf, W. Hartmann, M. Lauermann, H. Zwickel, Y. Kutuvantavida, C. Kieninger, W. Freude, C. Koos: High-speed silicon-organic hybrid (SOH) modulators. In: *Eur. Conf. Opt. Commun.* (2016), Paper Tu.3.A.3

5.21 O. Takanori, K. Kikuchi: *Coherent Optical Fiber Communications*, Vol. 4 (Springer, New York 1988)

5.22 D. Ly-Gagnon, S. Tsukamoto, K. Katoh, K. Kikuchi: Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation, J. Lightwave Technol. **24**, 12–21 (2006)

5.23 P.J. Winzer, R.-J. Essiambre: Advanced modulation formats for high-capacity optical transport networks, J. Lightwave Technol. **24**, 4711–4728 (2006)

5.24 P.J. Winzer: High-spectral-efficiency optical modulation formats, J. Lightwave Technol. **30**, 3824–3835 (2012)

5.25 Y. Ogiso, J. Ozaki, Y. Ueda, N. Kashio, N. Kikuchi, E. Yamada, H. Tanobe, S. Kanazawa, H. Yamazaki, Y. Ohiso, T. Fujii, M. Kohtoku: Over 67 GHz bandwidth and 1.5 Vp; InP-based optical I/Q modulator with n-i-p-n heterostructure, J. Lightwave Technol. **35**, 1450–1455 (2017)

5.26 P. Dong, L. Chen, Y.-K. Chen: High-speed low-voltage single-drive push-pull silicon Mach–Zehnder modulators, Opt. Express **20**, 6163–6169 (2012)

5.27 E.L. Wooten, K.M. Kissa, A. Yi-Yan, E.J. Murphy, D.A. Lafaw, P.F. Hallemeier, D. Maack, D.V. Attanasio, D.J. Fritz, G.J. McBrien, D.E. Bossi: A review of lithium niobate modulators for fiber-optic communications systems, IEEE J. Sel. Top. Quantum Electron. **6**, 69–82 (2000)

5.28 G. Lukovsky, R.F. Schwarz, R.B. Emmons: Transit-time considerations in p-i-n diodes, J. Appl. Phys. **35**, 622–628 (1964)

5.29 K. Kato, A. Kozen, Y. Muramoto, Y. Itaya, T. Nagatsuma, M. Yaita: 110-GHz, 50%-efficiency mushroom-mesa waveguide p-i-n photodiode for a 1.55-μm wavelength, IEEE Photonics Technol. Lett. **6**, 719–721 (1994)

5.30 A. Umbach, D. Trommer, G.G. Mekonnen, W. Ebert, G. Unterborsch: Waveguide integrated 1.55 μm photodetector with 45 GHz bandwidth, Electron. Lett. **32**, 2143–2145 (1996)

5.31 J.C. Campbell, A. Beling: High-speed, waveguide photodiodes. In: *Proc. Int. Top. Meet. Microwave* (2008) pp. 51–54

5.32 C. Shannon: Communication in the presence of noise, Proc. IRE **37**(1), 10–21 (1949)

5.33 A.J. Jerri: The Shannon sampling theorem – Its various extensions and applications: A tutorial review, Proc. IEEE Inst. Electr. Electron. Eng. **65**, 1565–1596 (1977)

5.34 G. Raybon, A. Adamiecki, J. Cho, F. Jorge, A. Konczykowska, M. Riet, B. Duval, J.-Y. Dupuy, N. Fontaine, P.J. Winzer, S. Chandrasekhar, X. Chen: 180-GBaud all-ETDM single-carrier polarization multiplexed QPSK transmission over 4480 km. In: *Proc. Opt. Fiber Commun. Conf.* (2018), Paper Th4C.3

5.35 H. Mardoyan, F. Jorge, O. Ozolins, J.M. Estaran, A. Udalcovs, A. Konczykowska, M. Riet, B. Duval, V. Nodjiadjim, J.-Y. Dupuy, X. Pang, U. Westergren, J. Chen, S. Popov, S. Bigo: 204-GBaud on-off keying transmitter for inter-data center communications. In: *Proc. Opt. Fiber Commun. Conf.* (2018), Paper Th4A.4

5.36 P. Pupalaikis: High speed arbitrary waveform generator, U.S. Patent 7,535,394 (2009)

5.37 X. Chen, S. Chandrasekhar, S. Randel, G. Raybon, A. Adamiecki, P. Pupalaikis, P.J. Winzer: All-electronic 100 GHz bandwidth digital-to-analog converter generating PAM signals up to 190 GBaud, J. Lightwave Technol. **35**, 411–417 (2017)

5.38 D.H. Sheingold: *Analog-Digital Conversion Handbook*, Vol. 16 (Prentice-Hall, Englewood Cliffs 1986)

5.39 R.J. Van de Plassche: *CMOS Integrated Analog-to-Digital and Digital-to-Analog Converters*, Vol. 742 (Springer, Berlin, Heidelberg 2013)

5.40 C. Laperle, M. O'Sullivan: Advances in high-speed DACs, ADCs, and DSP for optical coherent transceivers, J. Lightwave Technol. **32**, 629–643 (2014)

5.41 G. Lifante: *Integrated Photonics: Fundamentals* (Wiley, New York 2003)

5.42 R. Nagarajan, M. Kato, J. Pleumeekers, P. Evans, S. Corzine, S. Hurtt, A. Dentai, S. Murthy, M. Missey,

R. Muthiah, R.A. Salvatore, C. Joyner, R. Schneider, M. Ziari, F. Kish, D. Welch: InP photonic integrated circuits, IEEE J. Sel. Top. Quantum Electron. **16**, 1113–1125 (2010)

5.43 C.R. Doerr: Silicon photonic integration in telecommunications, Front. Phys. (2015), https://doi.org/10.3389/fphy.2015.00037

Xi Chen

Optical Transmission Research
Nokia Bell Labs
Holmdel, NJ, USA
xi.v.chen@nokia-bell-labs.com

Xi Chen received her PhD degree in 2012 from The University of Melbourne, Australia. From 2013 to 2015, Dr Chen was a Research Fellow at The University of Melbourne, where she received the Discovery Early Carrier Research Award (DECRA) grant. Dr Chen is currently a Member of Technical Staff in Nokia Bell Labs, New Jersey. Her research interests include fiber transmission and advanced digital signal processing for high-speed optical systems.

Sethumadhavan Chandrasekhar

Optical Transmission Research
Nokia Bell Labs
Holmdel, NJ, USA
s.chandrasekhar@ieee.org

Dr S. Chandrasekhar joined Nokia Bell Labs in 1986. He initially worked on semiconductor devices and optoelectronic integrated circuits (OEICs). Since 1999, he has been involved in WDM optical networking research at 100Gb/s and beyond. His current interests include coherent optical transmission systems for high spectral efficiency transport, multicarrier superchannels, and software-defined transponders for efficient end-to-end optical networking.

6. DSP for Optical Transponders

Seb J. Savory, David S. Millar

This chapter outlines the principles of the digital signal processing (DSP) used in modern optical transceivers. The historic developments that have led to the emergence of DSP being applied in optical transceivers is reviewed, including the high-speed complementary metal oxide semiconductor (CMOS) analog to digital converters (ADC) that have facilitated the creation of the application-specific integrated circuit (ASIC) which underpins digital coherent transceivers. Following on from this, the mathematics associated with finite impulse response (FIR) filters is reviewed, including the Wiener and least-squares design of FIR filters. The mathematics associated with the adaptive multiple-input-multiple-output (MIMO) filter employed in the receiver is also discussed, including derivation of the stochastic descent algorithm based on differentiation with respect to a complex vector. Subsequently, we provide an overview of DSP algorithms, before detailing both those required for equalization and synchronization. Following a summary of error correction used in a digital transceiver, we reflect on the current research trends and future opportunities for DSP in optical transceivers.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_6

6.1 Historical Background

The key to realizing digital signal processing (DSP) in optical transceivers has been the development of high-speed analog to digital converters (ADC) that can be realized in CMOS technology. Historically, optical fiber communication systems operated at the technological limits of the electronics available with data encoded as binary intensity modulation at a rate commensurate with the bandwidth of the electronics available. As such, optical line rates were far in excess of those that could be generated or sampled digitally. Consequentially, it seemed unlikely that DSP would be applied to optical fiber communication systems.

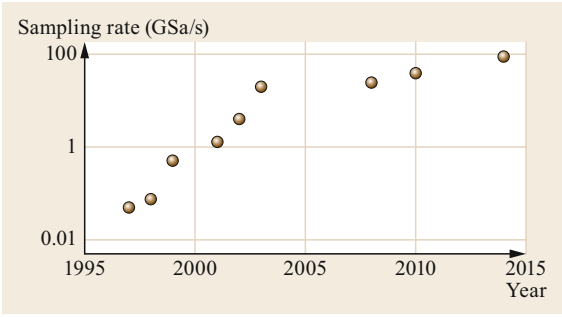


Fig. 6.1 Progress in the maximum CMOS ADC sampling rate reported at the International Solid-State Circuits Conference (ISSCC)

This situation changed significantly at the turn of the century when 40 Gb/s optical systems were being developed. While 40 Gb/s could have been deployed in the early 2000s, the dot com crash delayed commercial deployment by several years, such that the optical line rate stagnated at 10 Gb/s for a decade (from 1997–2007). At the same time, as can be seen in Fig. 6.1, research into CMOS-based ADCs made staggering improvements, with sampling speeds increasing by two orders of magnitude in just 5 years (1997–2002). Critically, in 2003 when 20 GSa/s CMOS ADCs first became available [6.1], it allowed the current 10 Gb/s optical line rate to be digitized. The emergence of high-speed CMOS data converters that could be integrated with the digital signal processing (DSP) into a single application-specific integrated circuit (ASIC) enabled a digital revolution in optical transceiver design.

While DSP was initially applied to the direct detection systems of the time, attention quickly moved toward digital coherent transceivers. Rather than merely directly detecting the power, as in a traditional system, a phase and polarization diverse coherent receiver was employed to linearly detect the electrical field, as illustrated in Fig. 6.2.

As we will discuss in subsequent sections, typically the DSP first equalizes the incoming signal prior to synchronizing the receiver with the transmitter, with

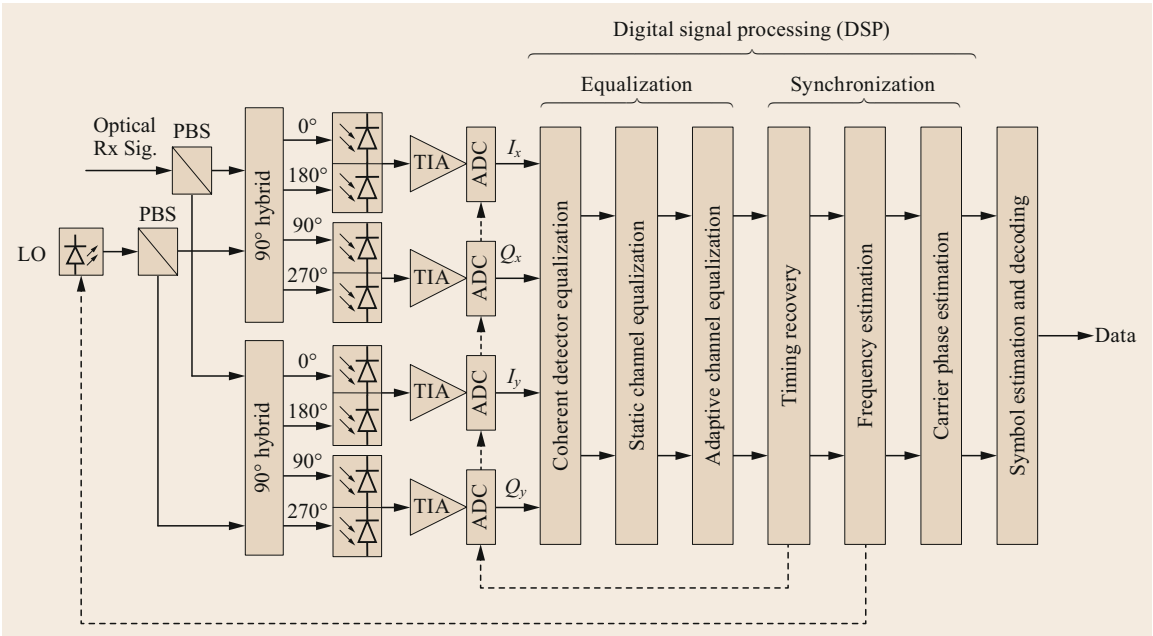


Fig. 6.2 Phase and polarization diverse coherent receiver and associated DSP (TIA: transimpedance amplifier, PBS: polarization beam splitter)

feedback provided to the frequency of the local oscillator laser and the ADC.

The first commercially available digital coherent transceiver emerged in 2008 [6.2], with the Nortel transceiver, which used dual polarization quadriphase shift keying (DP-QPSK) at a line rate of 11.5 GBd to achieve a net data rate of 40 Gb/s. The resulting DSP was realized in a single ASIC using 90 nm CMOS incorporating four 23 GSa/s ADCs with 6 bits of resolution to digitize the 46 Gb/s signal (with the power per bit of resolution reducing from 1.1 W in 0.18 μm CMOS [6.1] to less than 0.5 W in 90 nm CMOS [6.3]). Within a year of the 40 Gb/s product being released a 100 Gb/s product was released in 2009 [6.4], as a dual carrier transceiver with the 46 Gb/s product overclocked at 56 Gb/s, such that from the two carriers a total of 112 Gb/s could be transmitted. This allowed 100 Gb/s coherent transceivers to be realized a year earlier than the single-carrier 100 Gb/s solutions that emerged in 2010 [6.5]. Since then progress has been rapid, with 400 Gb/s systems now commercially deployed and research continuing to 1 Tb/s and beyond.

In this chapter, we will review the DSP currently being applied to optical transceivers. Given the investment required, at present this is limited to coherent transceivers, which will be the focus of this chapter. Nonetheless, as we outline later in this chapter, it seems likely that DSP may progress beyond coherent transceiver to direct detection transceivers. Before going into detail regarding algorithms, we begin by reviewing the underpinning mathematics associated with finite impulse response (FIR) filters that arise in many of the algorithms employed.

6.2 Preliminaries

6.2.1 FIR Filters

A key building block of DSP in optical transceivers is the finite impulse response (FIR) filter. Herein we discuss the underlying mathematics of FIR filters and their optimization (including differentiation with respect to a complex vector). As illustrated in Fig. 6.3, a finite impulse response (FIR) filter is a nonrecursive filter with tap weights $h[n]$, where $n \in \{0, 1, \ldots, N-1\}$, and the time step between taps given by T_s.

In the time domain the taps may be written as

$$h(t) = \sum_{n=0}^{N-1} h[n]\delta(t - nT_s) \,. \tag{6.1}$$

Alternatively, we can write

$$h(t) = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{\delta}_s \tag{6.2}$$

where superscript T denotes the transpose operation and $\boldsymbol{h}$ and $\boldsymbol{\delta}_s$ are defined as

$$\boldsymbol{h}^{\mathrm{T}} = [h[0], h[1].. \, h[N-1]] \tag{6.3}$$

and

$$\boldsymbol{\delta}_s^{\mathrm{T}} = [\delta(t), \delta\,(t - T_s)\,, .. \,\delta\,(t - (N-1)T_s)] \,. \tag{6.4}$$

Hence, the Fourier transform is given by

$$\begin{aligned} H(\omega) &= \int_{-\infty}^{\infty} h(t)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t = \sum_{n=0}^{N-1} h[n]\mathrm{e}^{-\mathrm{j}n\omega T_s} \\ &= \boldsymbol{h}^{\mathrm{T}}\boldsymbol{e}(\omega)\,, \end{aligned} \tag{6.5}$$

where we have defined

$$\boldsymbol{e}(\omega) = \left[1,\ \mathrm{e}^{-\mathrm{j}\omega T_s}, \mathrm{e}^{-2\mathrm{j}\omega T_s}, \ldots \mathrm{e}^{-\mathrm{j}\omega(N-1)T_s}\right]^{\mathrm{T}} \tag{6.6}$$

and we have assumed throughout this chapter the Fourier transform pair

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t \,, \tag{6.7}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\mathrm{e}^{\mathrm{j}\omega t}\mathrm{d}\omega \,. \tag{6.8}$$

Given the above definitions, if we define $\omega_s = 2\pi/T_s$, it readily follows that since $\mathrm{e}^{-\mathrm{j}nm2\pi} = 1$ for $m \in \mathbb{Z}$, that $H(\omega + m\omega_s) = H(\omega)$ indicating the sampled spectrum is aliased to multiples of the sampling frequency, as expected from sampling theory.

In order to consider the effect of filtering we define an input signal vector $\boldsymbol{x}$ of N samples as

$$\boldsymbol{x}^{\mathrm{T}} = (x[n], x[n-1], .. \, x[n-(N-1)]) \,. \tag{6.9}$$

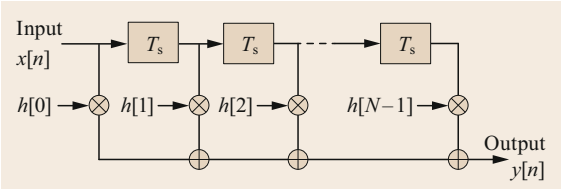


Fig. 6.3 Finite impulse response filter with output $y[n] = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x} = \sum_{i=0}^{N-1} h[i]x[n-i]$

Then the process of generating an output signal $y[n]$ by filtering an input vector may be written as

$$y[n] = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x} = \sum_{i=0}^{N-1} h[i]x[n-i]\,, \tag{6.10}$$

indicating that the output is obtained via the discrete time convolution of the input vector with the tap weights. This convolution may be realized in the frequency domain as multiplication, underpinning computationally efficient techniques such as the overlap and save method that we will discuss at the end of this section.

6.2.2 Differentiation with Respect to a Complex Vector

If z is a complex number such that $z = x + \mathrm{j}y$, then we can define differentiation with respect to a complex number as via the Wirtinger derivative given by

$$\frac{\partial}{\partial z} = \frac{1}{2}\frac{\partial}{\partial x} - \frac{\mathrm{j}}{2}\frac{\partial}{\partial y}\,, \tag{6.11}$$

with

$$\frac{\partial}{\partial z^*} = \left(\frac{\partial}{\partial z}\right)^* = \frac{1}{2}\frac{\partial}{\partial x} + \frac{\mathrm{j}}{2}\frac{\partial}{\partial y}\,, \tag{6.12}$$

which in turn gives

$$\frac{\partial z}{\partial z} = 1 \quad \frac{\partial z^*}{\partial z^*} = 1 \quad \frac{\partial z^n}{\partial z} = nz^{n-1}\,, \tag{6.13}$$

which is in keeping with expectations from usual calculus. However, it also follows that

$$\frac{\partial z^*}{\partial z} = \frac{\partial z}{\partial z^*} = 0\,, \tag{6.14}$$

indicating that a complex variable and its conjugate may be considered as independent variables, insofar as differentiation is concerned. If we extend the concept to a vector, such that

$$\frac{\partial}{\partial \boldsymbol{z}} = \frac{1}{2}\frac{\partial}{\partial \boldsymbol{x}} - \frac{\mathrm{j}}{2}\frac{\partial}{\partial \boldsymbol{y}}\,, \tag{6.15}$$

where $\partial/\partial\boldsymbol{x} = (\partial/\partial x_0, \partial/\partial x_1, \ldots \partial/\partial x_{N-1})^{\mathrm{T}}$ etc., then it follows that

$$\frac{\partial \boldsymbol{z}^{\mathrm{T}}}{\partial \boldsymbol{z}} = \frac{\partial}{\partial \boldsymbol{z}}\boldsymbol{z}^{\mathrm{T}} = \mathbf{I} = \frac{\partial \boldsymbol{z}^{*\mathrm{T}}}{\partial \boldsymbol{z}^*} \tag{6.16}$$

and

$$\frac{\partial \boldsymbol{z}^{*\mathrm{T}}}{\partial \boldsymbol{z}} = \mathbf{0} = \frac{\partial \boldsymbol{z}^{\mathrm{T}}}{\partial \boldsymbol{z}^*}\,, \tag{6.17}$$

where $\mathbf{I}$ and $\mathbf{0}$ are the $N \times N$ identity and null matrices, respectively.

6.2.3 Least-Squares Tap Weights

The least-squares formulation gives the tap weights that minimize the squared error ϵ^2 between the desired response $H_{\mathrm{d}}(\omega)$ and the obtained $H(\omega)$ over a frequency range $\omega \in (-\omega_{\mathrm{s}}/2, \omega_{\mathrm{s}}/2)$, such that [6.6]

$$\epsilon^2 = \int_{-\frac{\omega_{\mathrm{s}}}{2}}^{\frac{\omega_{\mathrm{s}}}{2}} |H(\omega) - H_{\mathrm{d}}(\omega)|^2\,\mathrm{d}\omega\,. \tag{6.18}$$

Substituting our definition for $H(\omega) = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{e}(\omega)$ we obtain

$$\frac{\mathrm{d}\epsilon^2}{\mathrm{d}\boldsymbol{h}} = \int_{-\frac{\omega_{\mathrm{s}}}{2}}^{\frac{\omega_{\mathrm{s}}}{2}} \left[\boldsymbol{h}^{\mathrm{T}}\boldsymbol{e}(\omega) - H_{\mathrm{d}}(\omega)\right]^* \boldsymbol{e}(\omega)\mathrm{d}\omega = \mathbf{0}\,, \tag{6.19}$$

which, noting that $\boldsymbol{e}^{\mathrm{T}}(\omega)\,\boldsymbol{h} = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{e}(\omega)$, can be rewritten as

$$\left[\int_{-\frac{\omega_{\mathrm{s}}}{2}}^{\frac{\omega_{\mathrm{s}}}{2}} \boldsymbol{e}^*(\omega)\boldsymbol{e}^{\mathrm{T}}(\omega)\mathrm{d}\omega\right]\boldsymbol{h} = \int_{-\frac{\omega_{\mathrm{s}}}{2}}^{\frac{\omega_{\mathrm{s}}}{2}} \boldsymbol{e}^*(\omega)H_{\mathrm{d}}(\omega)\mathrm{d}\omega\,. \tag{6.20}$$

Noting that $\int_{-\omega_{\mathrm{s}}/2}^{\omega_{\mathrm{s}}/2} \boldsymbol{e}^*(\omega)\boldsymbol{e}^{\mathrm{T}}(\omega)\mathrm{d}\omega = \omega_{\mathrm{s}}\mathbf{I}$, it follows that

$$\boldsymbol{h}_{\mathrm{opt}} = \frac{1}{\omega_{\mathrm{s}}}\int_{-\frac{\omega_{\mathrm{s}}}{2}}^{\frac{\omega_{\mathrm{s}}}{2}} \boldsymbol{e}^*(\omega)H_{\mathrm{d}}(\omega)\mathrm{d}\omega\,, \tag{6.21}$$

where the tap weights $\boldsymbol{h}_{\mathrm{opt}}$ are optimal in a least-squares sense.

6.2.4 Application to Stochastic Gradient Algorithms

In an adaptive equalizer, we frequently have a cost function whose gradient is stochastically estimated and used to update the tap weights. For a complex valued set of taps $\boldsymbol{h}$, the stochastic gradient algorithm is applied to the real ($\mathrm{Re}(\boldsymbol{h})$) and imaginary ($\mathrm{Im}(\boldsymbol{h})$) components of the taps independently, and hence may be written as

$$\mathrm{Re}(\boldsymbol{h}) \leftarrow \mathrm{Re}(\boldsymbol{h}) - \frac{\mu}{2}\left(\frac{\partial\epsilon^2}{\partial\mathrm{Re}(\boldsymbol{h})}\right) \tag{6.22}$$

and

$$\mathrm{Im}(\boldsymbol{h}) \leftarrow \mathrm{Im}(\boldsymbol{h}) - \frac{\mu}{2}\left(\frac{\partial\epsilon^2}{\partial\mathrm{Im}(\boldsymbol{h})}\right), \tag{6.23}$$

where $\leftarrow$ denotes the assignment operation, such that $x \leftarrow y$ indicates that x is assigned to be the value y. These two equations may, however, be written more compactly in terms of the conjugate derivative as

$$\boldsymbol{h} \leftarrow \boldsymbol{h} - \mu \frac{\partial \epsilon^2}{\partial \boldsymbol{h}^*} \,. \tag{6.24}$$

To illustrate the approach, we consider the least mean-squares equalizer whose error term is given by $\epsilon = d - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}$, where d is the desired output, which gives the squared error term as $|\epsilon|^2 = \left|d - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}\right|^2 = \left(d - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}\right)^* \left(d - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}\right)$, and hence using the product rule for differentiation and the relationships related to differentiation with respect to a complex vector and its conjugate, we obtain

$$\frac{\partial |\epsilon|^2}{\partial \boldsymbol{h}^*} = -\epsilon \boldsymbol{x}^* \,. \tag{6.25}$$

Hence, the tap weight adaption algorithm is given by

$$\boldsymbol{h} \leftarrow \boldsymbol{h} + \mu \epsilon \, \boldsymbol{x}^* \tag{6.26}$$

Equally, for the constant modulus algorithm (CMA) [6.7], the cost function is

$$\epsilon = 1 - \left|\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}\right|^2 \,. \tag{6.27}$$

Similarly, we find

$$\frac{\partial \epsilon^2}{\partial \boldsymbol{h}^*} = -2\epsilon \, \boldsymbol{x}^* \left(\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}\right), \tag{6.28}$$

giving the update algorithm for the CMA as

$$\boldsymbol{h} \leftarrow \boldsymbol{h} + 2\mu\epsilon \, \boldsymbol{x}^* \left(\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}\right). \tag{6.29}$$

6.2.5 The Wiener Filter

We have already shown that the error term is given by $\epsilon = d - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}$, such that

$$\frac{\partial |\epsilon|^2}{\partial \boldsymbol{h}^*} = -\epsilon \boldsymbol{x}^* = -\boldsymbol{x}^* \epsilon^{\mathrm{T}} = -\boldsymbol{x}^* \left(d - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{h}\right). \tag{6.30}$$

While we have shown how it is possible to iteratively solve this using a stochastic gradient technique, an alternative is to solve this analytically by setting the expected value of derivative to zero so as to give

$$\langle \boldsymbol{x}^* d \rangle = \langle \boldsymbol{x}^* \boldsymbol{x}^{\mathrm{T}} \boldsymbol{h} \rangle \,, \tag{6.31}$$

where $\langle \cdot \rangle$ is the expectation operator. Hence, if we define $\boldsymbol{P} = \langle \boldsymbol{x}^* d \rangle$ as the cross-correlation vector between the desired signal and the distorted signal, and $\mathbf{R}_{xx} = \langle \boldsymbol{x}^* \boldsymbol{x}^{\mathrm{T}} \rangle$ to be the autocorrelation matrix of the distorted signal, then we can write the Wiener filter solution of the tap weights as

$$\boldsymbol{h} = \mathbf{R}_{xx}^{-1} \boldsymbol{P} \,. \tag{6.32}$$

Often in a coherent optical communication system the situation is further complicated by the presence of phase noise or the frequency difference between the signal and the local oscillator. Nevertheless, it may be readily be applied in a simulation environment where frequency offset correction and carrier recovery is not required.

6.2.6 Frequency Domain Implementation of FIR Filters

For a long-haul communication system, a large number of FIR filter taps N_{f} are typically required to compensate for the accumulated chromatic dispersion. In such a case, the filtering can be done more efficiently in the frequency domain with $\mathcal{O}(\log N_{\mathrm{f}})$ complex multiplications. The algorithm we will outline is traditionally referred to as overlap and save, although overlap and discard would be a more appropriate name. The first key concept is that the convolution is done on a block-by-block basis, with the output of each block combined to give the overall output. A second key concept is that $\mathrm{FFT}^{-1}\{\mathrm{FFT}(x)\mathrm{FFT}(h)\}$ results in the circular convolution between the arrays x and h (rather than the desired linear convolution). If we define the size of the FFT to be N, then the overlap and save algorithm is as follows [6.8]:

1. Append $N - N_{\mathrm{f}}$ zeros to the tap weights to get an array of length N. This we transform into the frequency domain (using an N point FFT). Since this transformation is done just once, it can be neglected insofar as the computational cost (being fundamentally just an alternative way of representing the tap weights) is concerned.
2. Take the block of data of length N and FFT. An N point FFT requires $(N/2)\log_2 N$ complex multiplications.
3. The taps in the frequency domain from stage (1) are multiplied by the data in the frequency domain from part (2). This requires a further N complex multiplications
4. The resulting vector is then transformed back into the time domain using an N point inverse fast Fourier transform IFFT requiring a further $(N/2)\log_2 N$ complex multiplication.
5. Discard the first $N_{\mathrm{f}} - 1$ samples and output the last $N - N_{\mathrm{f}} + 1$ samples.

6. Take the next block of data samples, which starts $N - N_f + 1$ after the preceding block.

The net result is that the number of complex multiplications per output sample is

$$N_{CM} = \frac{N \log_2 N + N}{N - N_f + 1} . \quad (6.33)$$

In general, the size of the FFT can be optimized to minimize the number of complex multiplications to give N_{opt}, with the optimal values plotted in Fig. 6.4.

Given that direct linear convolution would require N_f complex multiplications per sample, it can be shown that the overlap and save is more efficient for $N_f \geq 8$ (using a 32-point FFT), yielding $N_{CM} = 7.7$. Using alternative FFT implementations, such as the radix 4 rather than radix 2, can reduce the computation complexity still further.

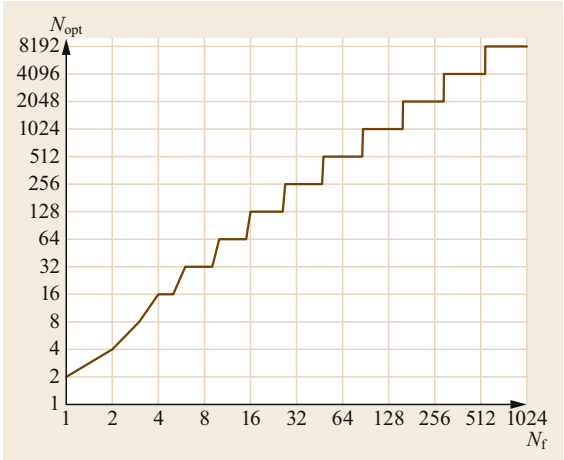


Fig. 6.4 Optimum FFT size to minimize the number of complex multiplications

6.3 Overview of DSP Algorithms

Herein, we will focus on receiver-based digital signal processing; however, many methods are equally applicable to the transmitter. As illustrated in Fig. 6.2, these can be broadly partitioned into equalization and synchronization algorithms with the typical arrangement being to equalize the signal before attempting to perform synchronization.

6.3.1 Equalization Algorithms

Equalization algorithms correct for the channel response and include the following:

- Orthonormalization and de-skew algorithms – these correct for imperfections in the coherent detection, such as the 90° hybrids not being ideal, as well as timing and amplitude differences between the different paths.
- Static channel equalization algorithms – these correct for the accumulated chromatic dispersion (which varies slowly). At the receiver this can also include pulse shaping and matched filtering, assuming that similar processing has been applied at the transmitter.
- Adaptive channel equalization algorithms – these correct for dynamic variations. For example, the polarization axes at the transmitter will be rotated relative to the receiver. These algorithms also compensate for PMD.

6.3.2 Synchronization Algorithms

Once the channel has been equalized, the synchronization algorithms correct for the difference between the oscillators at the transmitter and the receiver, both optical and electrical. In principle, synchronization could be achieved prior to equalization, however, many algorithms perform suboptimally in the presence of distortion and, as such, it is usual to equalize the channel prior to synchronizing the receiver and transmitter. Typical algorithms include:

- Timing recovery algorithms – these ensure that the sample clock at the receiver matches that at the transmitter, by both adjusting the ADC sampling phase and rate, as well as digitally interpolating to give samples at the correct time interval.
- Frequency estimation algorithms – these detect the difference in frequency between the incoming carrier frequency on to which the data is modulated and the local oscillator. The local oscillator wavelength can be adjusted and the residual frequency error corrected digitally.
- Carrier phase estimation algorithms – both the transmitter and local oscillator laser will have phase noise, which follows a random walk (Brownian motion, also known as a Wiener process). These algorithms track the mean phase evolution to reduce the impact of phase noise.

6.4 Orthonormalization and De-skew Algorithms

The aim of these algorithms is to compensate for the imperfections in the optical front end, where optical path lengths may not be perfectly matched or the hybrid may not be exactly 90°. The orthonormalization process compensates for imbalances in amplitude and nonideal 90° hybrids after which de-skew compensates for the path length mismatches between the signals, synchronizing the digital signals with respect to one another in time.

6.4.1 Orthogonalization Algorithms

In order to recover the original signal using a receiver with a suboptimal hybrid it is often necessary to use orthogonalization algorithms. The most common approach is to use the Gram–Schmidt algorithm [6.9], which creates a set of mutually orthogonal vectors, taking the first vector as a reference against which all subsequent vectors are orthogonalized. In a coherent receiver, the primary cause of nonorthogonality is due to a nonideal 90° hybrid. If r_I and r_Q are the received in-phase and quadrature components, then the orthogonalized components are given by

$$I = r_\mathrm{I} \,, \tag{6.34}$$

$$Q = r_\mathrm{Q} - \frac{\langle r_\mathrm{I} r_\mathrm{Q} \rangle}{\langle r_\mathrm{I}^2 \rangle} r_\mathrm{I} \,, \tag{6.35}$$

After the signals have been orthogonalized, they are subsequently normalized to unit power.

6.4.2 De-skew

In order to temporally align the channels any timing skew should be corrected for. This can be implemented using a short FIR filter to interpolate between samples in order to create a temporal delay. Alternatively, it can be readily implemented via the frequency domain, where the temporal delay corresponds to a linear phase response. Since skew is fundamentally a timing offset, the methods employed are similar to those discussed in Sect. 6.7.1 for timing recovery. As such, we will not go into more detail here other than to note that care should be taken to ensure that if the skew is tracked the interaction with the channel is considered [6.10]. While robust estimators exist, such as [6.10], an alternative approach is to perform an initial calibration for the skew, using, for example, a 4×4 equalizer (discussed in Sect. 6.6.7), since such an equalizer can correct for timing skew [6.11].

6.5 Static Equalization

In modern digital coherent transceivers, the static equalizer is used to implement pulse shaping and matched filtering, as well as equalizing for the significant intersymbol interference due to the accumulated chromatic dispersion. Since the application of pulse shaping is well covered by digital communications textbooks such as *Proakis* and *Salehi* [6.12], in this section, we focus on the compensation of chromatic dispersion, since this is specific to optical fiber communication systems.

6.5.1 Nature of Chromatic Dispersion

Chromatic dispersion is a consequence of the frequency-dependent group delay in the optical fiber. If two wavelengths are separated by $\Delta\lambda$ nm, then the temporal spread Δt (in ps) is given by [6.13]

$$\Delta t = D z \Delta\lambda \,, \tag{6.36}$$

where D is the dispersion coefficient of the fiber given in ps/(nm km), and z is the length of the link in km. Given $c = f\lambda$, it follows that $\Delta\lambda = -\Delta f \times \lambda / f$, hence for $f \approx 193$ THz ($\lambda \approx 1553$ nm), then $\Delta\lambda/\Delta f = 8$ pm/GHz. Hence, a signal occupying 35 GHz has a spectral width of 0.28 nm. If the dispersion coefficient is 16.7 ps/(nm km), then for every 1000 km of fiber, the signal disperses by at least 165 symbol periods, with the minimum value obtained by assuming a rectangular spectrum. The number of taps depends on the ADC sampling rate, for example at 40 GSa/s 189 taps per 1000 km of SMF are required, increasing to 330 taps per 1000 km of SMF if two samples per symbol are used with an ADC sampling rate of 70 GSa/s.

6.5.2 Modeling of Chromatic Dispersion in an Optical Fiber

In the absence of fiber nonlinearity, the effect of chromatic dispersion on the envelope $A(z,t)$ of a pulse may be modeled by the following partial differential equation (which is based on the electronic engineer's definition of phase compared with the physicist's con-

vention [6.13])

$$\frac{\partial A(z,t)}{\partial z} = \frac{\mathrm{j}\beta_2}{2}\frac{\partial^2 A(z,t)}{\partial t^2}\,, \tag{6.37}$$

where z is the distance of propagation, t is time variable in a frame moving with the pulse $A(z,t)$, and β_2 the group delay dispersion of the fiber. Taking the Fourier transform of (6.37) and solving gives the frequency domain transfer function $G(z,\omega)$, given by

$$G(z,\omega) = \exp\left(-\frac{\mathrm{j}\beta_2}{2}\omega^2 z\right). \tag{6.38}$$

The dispersion compensating filter is, therefore, given by the all-pass filter $G(-z,\omega)$, which can be approximated using an FIR filter.

6.5.3 Truncated Impulse Response

Herein, we discuss a simple but intuitive means of designing the chromatic dispersion compensating FIR filter [6.14], providing a basis for the discussion of more complex techniques. In contrast to a frequency domain approach, not only does this give a simple closed-form solution for the tap weights, but also it provides bounds on the number of taps required for a given value of dispersion. We begin by obtaining the impulse response $g(z,t)$ of the dispersive fiber by applying the inverse Fourier transform to the frequency domain transfer function $G(z,\omega)$ to give

$$g(z,t) = \frac{1}{\sqrt{2\pi\mathrm{j}\beta_2 z}}\exp\left(\frac{\mathrm{j}}{2\beta_2 z}t^2\right). \tag{6.39}$$

For an arbitrary input, the output can be obtained by convolving this impulse response with the input, and as expected, the impulse response itself satisfies (6.37). By inverting the sign of the chromatic dispersion, we obtain the impulse function of the chromatic dispersion compensating filter $g_c(z,t)$, given by

$$g_c(z,t) = \frac{1}{\sqrt{-2\pi\mathrm{j}\beta_2 z}}\exp[-\mathrm{j}\phi(t)]\,, \tag{6.40}$$

where

$$\phi(t) = \frac{t^2}{2\beta_2 z}\,. \tag{6.41}$$

The impulse response given by (6.40) presents a number of issues for digital implementation. Not only is it infinite in duration, but since it passes all frequencies for a finite sampling frequency, aliasing will occur. The solution to all of these problems is to truncate the impulse response to a finite duration. To determine the length of the truncation window, we note that if we sample every T_s s, then aliasing will occur for frequencies which exceed the Nyquist frequency, given by $\omega_N = \pi/T$, and that the impulse response may be considered as a rotating vector, whose angular frequency is given by

$$\omega = \frac{\partial\phi(t)}{\partial t} = \frac{t}{\beta_2 z}\,. \tag{6.42}$$

When the magnitude of this frequency exceeds the Nyquist frequency, aliasing will occur, giving the criterion that $|\omega| < \omega_N$ and hence,

$$-|\beta_2|\,z\frac{\pi}{T_s} \le t \le |\beta_2|\,z\frac{\pi}{T_s}\,. \tag{6.43}$$

Since the impulse response is of finite duration, this can be implemented digitally using a finite impulse response (FIR) filter. If we assume that the number of taps is large, then the sampled impulse response will approximate the continuous time impulse response. Hence, if we consider a filter with N_{TI} taps, then the tap weights will be given by

$$h_{TI}[n] = \frac{1}{\sqrt{\rho}}\exp\left[-\frac{\mathrm{j}\pi}{\rho}\left(n - \frac{N_{TI}-1}{2}\right)^2\right], \tag{6.44}$$

where

$$\begin{aligned} &\rho = 2\frac{\pi\beta_2 z}{T_s^2}\,,\\ &N_{TI} = \lfloor|\rho|\rfloor\,,\\ \text{and}\quad &n \in [0,1,2,\ldots,N_{TI}-1]\,, \end{aligned} \tag{6.45}$$

where $\lfloor x\rfloor$ is the integer part of x rounded towards minus infinity. These tap weights form the basis for the compensation of chromatic dispersion using an FIR filter.

As an example, we consider the application of this truncated impulse FIR filter design to the compensation of chromatic dispersion for a 35 GBd signal (with a root raised cosine shape with $\beta = 0.01$). As can be seen in Fig. 6.5, for a QPSK signal, the penalty is less than 0.1 dB when the distance exceeds 250 km. Given that the first digital coherent transceivers employed QPSK and operated over thousands of kilometers, this design approach was adequate for the first generation of systems. Nonetheless, as future systems increased the cardinality of the modulation format, moving towards 16 QAM and 64 QAM, this approach proved suboptimal, with for example a 2 dB penalty observed using 64 QAM at 250 km. This resulted in moving to an alternative formulation, namely least-squares (LS) filter design.

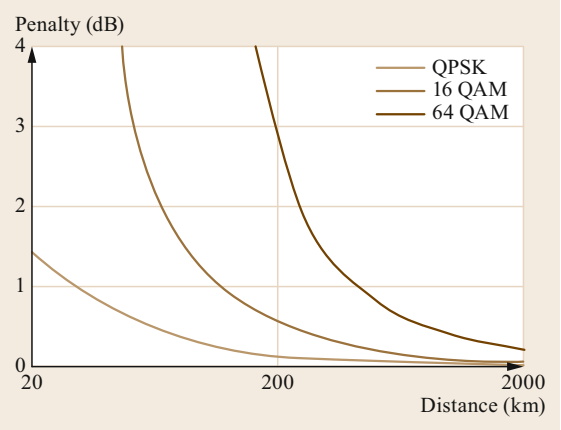


Fig. 6.5 Penalty at BER = 2% for a 35 GBd signal ($\beta = 0.01$) with a truncated impulse response design. The single-mode fiber has a chromatic dispersion of $D = 17\,\mathrm{ps/(nm\,km)}$

6.5.4 LS FIR Filter Design

In order to improve the design of the chromatic dispersion compensation, first, a least-squares approach can be utilized. As previously discussed, the least-squares criterion may be applied to the design of a complex FIR filter, giving optimal tap weights in a least-squares sense

$$\boldsymbol{h}_{\mathrm{LS}} = \frac{1}{\omega_{\mathrm{s}}} \int_{-\frac{\omega_{\mathrm{S}}}{2}}^{\frac{\omega_{\mathrm{S}}}{2}} \boldsymbol{e}^{*}(\omega) H_{\mathrm{d}}(z,\omega)\mathrm{d}\omega \,. \tag{6.46}$$

If, as in the previous section, we define our taps symmetrically such that

$$\boldsymbol{e}(\omega) = \mathrm{e}^{\mathrm{j}\omega(N-1)T_{\mathrm{s}}/2} \times \left[1, \mathrm{e}^{-\mathrm{j}\omega T_{\mathrm{s}}}, \mathrm{e}^{-2\mathrm{j}\omega T_{\mathrm{s}}}, \ldots \mathrm{e}^{-\mathrm{j}\omega(N-1)T_{\mathrm{s}}}\right]^{\mathrm{T}} ,$$

and so as to neglect the combination of any subsequent filtering, we choose $H_{\mathrm{d}}(z,\omega) = 1/G(z,\omega) = G(-z,\omega)$, then we obtain [6.15]

$$h_{\mathrm{LS}}[n] = w[n] \times h_{\mathrm{TI}}[n] \,, \tag{6.47}$$

where

$$w[n] = \frac{1}{2\mathrm{j}} \operatorname{erfi}\left[\sqrt{\frac{\mathrm{j}\pi}{\rho}}\left(n - \frac{N-1}{2} + \frac{\rho}{2}\right)\right] - \frac{1}{2\mathrm{j}} \operatorname{erfi}\left[\sqrt{\frac{\mathrm{j}\pi}{\rho}}\left(n - \frac{N-1}{2} - \frac{\rho}{2}\right)\right], \tag{6.48}$$

where $\operatorname{erfi}(x)$ is the imaginary error function given by $\operatorname{erfi}(x) = -\mathrm{j}\operatorname{erf}(\mathrm{j}x)$ where

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{0}^{z} \mathrm{e}^{-t^2}\mathrm{d}t \,. \tag{6.49}$$

It can be shown that $h_{\mathrm{LS}}[n]$ corresponds to the sampled version of the bandlimited impulse response $g_{\mathrm{c_{bl}}}(z,t)$, given by

$$g_{\mathrm{c_{bl}}}(z,t) = \frac{1}{2\pi} \int_{-\frac{\omega_{\mathrm{S}}}{2}}^{\frac{\omega_{\mathrm{S}}}{2}} \exp\left(\frac{\mathrm{j}\beta_2}{2}\omega^2 z\right) \exp(\mathrm{j}\omega t)\mathrm{d}\omega \,. \tag{6.50}$$

We revisit the previous example, designing an FIR filter for compensation of chromatic dispersion for a 35 GBd signal (with root raised cosine shape with $\beta = 0.01$), while using the same number of taps $N = \lfloor 2\pi\beta_2 z/T_{\mathrm{s}}^2 \rfloor$ as for the truncated impulse response. As can be seen in Fig. 6.6, for the distance of 250 km, the penalty for 64 QAM reduces significantly from 2 dB using the truncated impulse response to less than 0.4 dB using the least-squares formulation.

Nevertheless, for highly spectrally efficient formats such as 64 QAM, the penalty can be as high as 3 dB for short distances (≈ 30 km). The penalty may, however, be mitigated by allowing the number of taps in the least-squares design to increase beyond $N = \lfloor 2\pi\beta_2 z/T_{\mathrm{s}}^2 \rfloor$. To illustrate this, we again consider a 35 GBd signal with a near rectangular Nyquist-shaped spectrum (root

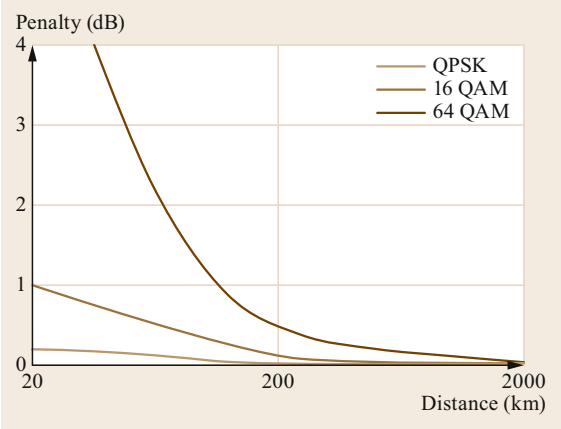


Fig. 6.6 Penalty at BER = 2% for a 35 GBd signal ($\beta = 0.01$) with a least-squares impulse response design. The single-mode fiber has a chromatic dispersion of $D = 17\,\mathrm{ps/(nm\,km)}$

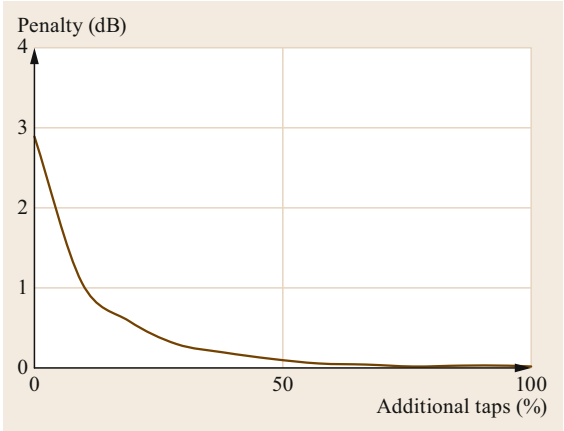


Fig. 6.7 Effect of increasing number of taps on the penalty at BER = 2% for a 35 GBd DP-64 QAM signal transmitted over 31.25 km of standard single-mode fiber ($D = 17\,\mathrm{ps/(nm\,km)}$) ►

raised cosine shape with $\beta = 0.01$), transmitting DP-64 QAM over a distance of 31.25 km of single-mode fiber with $D = 17\,\mathrm{ps/(nm\,km)}$. While this gives a minimum number of taps required of 20, as can be seen in Fig. 6.7, if the number of taps are increased by 50%, the penalty can be made negligible (< 0.1 dB) from the initial value of 2.9 dB with the minimum number of taps. We note that if FIR filtering is implemented via the frequency domain, the increase in the number of complex multiplications to achieve this is less than 10%.

6.6 Adaptive Equalization

Within coherent receivers, adaptive equalization is employed in order to track time-varying phenomena. In general, the cause of temporal variations is due to polarization effects, such as random polarization rotations or polarization mode dispersion (PMD).

6.6.1 Equalization of PMD

Polarization mode dispersion (PMD) arises due to variations in the circular symmetry of the optical fiber, resulting in localized birefringence [6.13]. While the PMD is a unitary operation, generally, compensation of PMD is included within a subsystem that includes mitigation of PDL, relaxing the unitary requirement. By removing the unitary condition polarization, independent effects such as nonideal matched filtering can also be mitigated. Polarization mode dispersion results in information being coupled from one polarization to another, such that the information in the x and y polarization at the output $\boldsymbol{E}_{\mathrm{out}}$ is related to the input states $\boldsymbol{E}_{\mathrm{in}}(\omega)$ via [6.16]

$$\boldsymbol{E}_{\mathrm{out}} = \mathbf{U}(\omega)\boldsymbol{E}_{\mathrm{in}}\,, \tag{6.51}$$

where $\mathbf{U}(\omega)$ is a 2×2 unitary matrix. The simplest manifestation of PMD is as a differential group delay (DGD) of $\Delta\tau$, such that

$$\mathbf{U}(\omega) = \mathbf{R}(-\theta)\begin{pmatrix} \mathrm{e}^{\frac{\mathrm{j}\omega\tau}{2}} & 0 \\ 0 & \mathrm{e}^{-\frac{\mathrm{j}\omega\tau}{2}} \end{pmatrix}\mathbf{R}(\theta)\,, \tag{6.52}$$

where

$$\mathbf{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\,. \end{pmatrix} \tag{6.53}$$

From this, we note that the inverse Jones matrix in the time domain is given by

$$\mathbf{u}(t) = \mathbf{R}(-\theta)\begin{pmatrix} \delta\left(t-\frac{\tau}{2}\right) & 0 \\ 0 & \delta\left(t+\frac{\tau}{2}\right) \end{pmatrix}\mathbf{R}(\theta)\,. \tag{6.54}$$

This reveals that if the equalizer is to correct for a DGD of τ, then the temporal span $(N-1)T_{\mathrm{s}}$ of an N tap FIR filter must exceed τ, requiring $N \geq \lceil \tau/T_{\mathrm{s}} \rceil + 1$. As we will discuss in the subsequent sections, this is very much a lower bound on the number of taps, and in practice, when the convergence time of the equalizer is of concern, more taps may be required. Since PMD is due to random coupling between the polarization modes, DGD has a statistical variation with a Maxwellian distribution [6.17] with a probability density function given by

$$f(\tau) = \frac{32}{\langle\tau\rangle\,\pi^2}\left(\frac{\tau}{\langle\tau\rangle}\right)^2 \exp\left[-\frac{4}{\pi}\left(\frac{\tau}{\langle\tau\rangle}\right)^2\right], \tag{6.55}$$

where $\langle\tau\rangle$ is the mean DGD, and the resulting outage probability given by the corresponding tail distribution $F_{\mathrm{c}}(\tau)$ is given by

$$\begin{aligned} F_{\mathrm{c}}(\tau) &= \int_{\tau}^{\infty} f(x)\mathrm{d}x \\ &= \frac{4}{\pi}\left(\frac{\tau}{\langle\tau\rangle}\right)\exp\left[-\frac{4}{\pi}\left(\frac{\tau}{\langle\tau\rangle}\right)^2\right] \\ &\quad + \mathrm{erfc}\left[\frac{2}{\sqrt{\pi}}\left(\frac{\tau}{\langle\tau\rangle}\right)\right]. \end{aligned} \tag{6.56}$$

For the range that is typically of interest, namely $\tau \in (\langle\tau\rangle, 6\langle\tau\rangle)$, we can approximate the outage probability

by

$$\log_{10} F_c(\tau) \approx 0.25 \left(\frac{\tau}{\langle\tau\rangle}\right) - 0.57 \left(\frac{\tau}{\langle\tau\rangle}\right)^2 . \quad (6.57)$$

Consider data from a field trial of 100 Gb/s with 31.25 GBd DP-QPSK with $\langle\tau\rangle = 36.6$ ps with an adaptive equalizer with 13 taps [6.18]. This gives $\tau/\langle\tau\rangle = 5.2$, and hence an negligible outage probability of 4×10^{-15}. While it was also noted that no penalty was observed with $\langle\tau\rangle$ up to 55 ps, the corresponding outage probability is 8×10^{-7}, and as such a penalty is not expected to be observed.

6.6.2 Obtaining the Inverse Jones Matrix of the Channel

The impact of polarization-dependent effects on the propagation may be modeled by a Jones matrix. In general, this matrix is not unitary due to polarization-dependent loss (PDL), and furthermore, it will be frequency dependent due to polarization mode dispersion (PMD). The task is, therefore, to estimate the Jones matrix and obtain the inverse to compensate for the impairments incurred. In contrast to the chromatic dispersion that may be considered relatively constant, the Jones matrix may evolve in time due to effects such as rapid variations in the polarization state, and, therefore, the compensation scheme must be adaptive. The problem of compensating polarization rotations digitally was first considered by *Betti* et al. [6.19] and later demonstrated utilizing the formalism of multiple-input-multiple-output (MIMO) systems [6.20]. For inputs

$$\boldsymbol{x} = (x[k], x[k-1], \ldots x[k-(N-1)])^{\mathrm{T}}$$

and

$$\boldsymbol{y} = (y[k], y[k-1], \ldots y[k-(N-1)])^{\mathrm{T}} ,$$

the outputs $x_o[k]$ and $y_o[k]$ are given by

$$x_o[k] = \boldsymbol{h}_{xx}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{h}_{xy}^{\mathrm{T}}\boldsymbol{y} \quad (6.58)$$

and

$$y_o[k] = \boldsymbol{h}_{yx}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{h}_{yy}^{\mathrm{T}}\boldsymbol{y} , \quad (6.59)$$

where $\boldsymbol{h}_{xx}$, $\boldsymbol{h}_{xy}$, $\boldsymbol{h}_{yx}$ and $\boldsymbol{h}_{yy}$ are adaptive filters each of which have length N taps. While there are a number of methods for adapting the equalizer in MIMO systems, we shall restrict ourselves to a specific example that exploits properties of the data, namely that for dual polarization QPSK (DP-QPSK) the signal for each polarization should have a constant modulus. This constant modulus algorithm (CMA) has also been shown to be effective even when the modulus is not constant, such as higher-order quadrature amplitude modulation.

6.6.3 Constant Modulus Algorithm

For signals of unit amplitude, the equalizer will attempt to minimize, in a mean squares sense, the magnitude of $\epsilon_x = 1 - |x_o|^2$ and $\epsilon_y = 1 - |y_o|^2$. Hence, to obtain the optimal tap weights, a set of stochastic gradient algorithms with convergence parameter μ are used, such that

$$\boldsymbol{h}_{xx} \leftarrow \boldsymbol{h}_{xx} - \mu \frac{\partial \epsilon_x^2}{\partial \boldsymbol{h}_{xx}^*} = \boldsymbol{h}_{xx} + 2\mu\epsilon_x x_o \boldsymbol{x}^* . \quad (6.60)$$

Similarly

$$\boldsymbol{h}_{xy} \leftarrow \boldsymbol{h}_{xy} + 2\mu\epsilon_x x_o \boldsymbol{y}^* , \quad (6.61)$$

$$\boldsymbol{h}_{yx} \leftarrow \boldsymbol{h}_{yx} + 2\mu\epsilon_y y_o \boldsymbol{x}^* , \quad (6.62)$$

$$\boldsymbol{h}_{yy} \leftarrow \boldsymbol{h}_{yy} + 2\mu\epsilon_y y_o \boldsymbol{y}^* , \quad (6.63)$$

where $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ denote the complex conjugate of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. In order to initialize the algorithm, all tap weights are set to zero with the exception of the central tap of $\boldsymbol{h}_{xx}$ and $\boldsymbol{h}_{yy}$, which are set to unity. Given that the equalizer is unconstrained with respect to its outputs, it is possible for both outputs of the equalizer to converge on the same output, corresponding to the Jones matrix becoming singular, albeit, there are numerous means of overcoming this limitation [6.21–23].

6.6.4 Decision-Directed Equalizer

Once the equalizer has converged, the equalizer may move into a decision-directed mode, such that if $D(x)$ is the symbol closest to x, then the decision-directed least mean squared (DD-LMS) algorithm minimizes $\epsilon_x = D(x_0) - x_o$ and $\epsilon_y = D(y_0) - y_o$, giving the update algorithm as [6.24]

$$\boldsymbol{h}_{xx} \leftarrow \boldsymbol{h}_{xx} - \mu \frac{\partial |\epsilon_x|^2}{\partial \boldsymbol{h}_{xx}^*} = \boldsymbol{h}_{xx} + \mu\epsilon_x \boldsymbol{x}^* \quad (6.64)$$

and likewise

$$\boldsymbol{h}_{xy} \leftarrow \boldsymbol{h}_{xy} + \mu\epsilon_x \boldsymbol{y}^* , \quad (6.65)$$

$$\boldsymbol{h}_{yx} \leftarrow \boldsymbol{h}_{yx} + \mu\epsilon_y \boldsymbol{x}^* , \quad (6.66)$$

$$\boldsymbol{h}_{yy} \leftarrow \boldsymbol{h}_{yy} + \mu\epsilon_y \boldsymbol{y}^* . \quad (6.67)$$

One of the challenges of the decision-directed equalizer is needing to combine the carrier recovery with the equalization, since the decisions are made on the phase corrected signal. The resulting feedback path can, therefore, be more challenging for CMOS ASIC implementation than the blind equalizer, which partitions the equalization from the carrier recovery.

6.6.5 Radially-Directed Equalizer Update Algorithm

While the CMA is well suited to constant modulus formats such at QPSK, many of the formats considered for future optical networks are not constant modulus, such as DP-16 QAM, such that the CMA will never converge to zero error [6.24]. Nevertheless, the CMA can be adapted to a radially directed equalizer. In this case,

$$\epsilon_x = Q_\mathrm{r}\left(|x_\mathrm{o}|^2\right) - |x_\mathrm{o}|^2 \tag{6.68}$$

and

$$\epsilon_y = Q_\mathrm{r}\left(|y_\mathrm{o}|^2\right) - |y_\mathrm{o}|^2 \,, \tag{6.69}$$

where $Q_\mathrm{r}(r^2)$ is a function that quantizes the radius according to the number of possible points. Once the notional radius has been determined, the CMA algorithm is used to update the tap weights accordingly. One of the key benefits of using a radially-directed equalizer is that it is invariant to the phase of the incoming signal, and hence allows the equalization and the carrier recovery to be partitioned. However, for more dense modulation formats, often it is preferable to use the CMA initially and then switch to a decision-directed equalizer.

6.6.6 Parallel Realization of the FIR Filter

One of the key benefits of the FIR filter is that it can readily be implemented in CMOS. Thus far, all of the implementations discussed have operated on a symbol-by-symbol basis, rather than based on a lower speed CMOS bus. In this case, the DSP is similar to before, but now we write [6.25]

$$\boldsymbol{x}_\mathrm{o} = \boldsymbol{h}_{xx}^\mathrm{T}\mathbf{X} + \boldsymbol{h}_{xy}^\mathrm{T}\mathbf{Y} \tag{6.70}$$

and

$$\boldsymbol{y}_\mathrm{o} = \boldsymbol{h}_{yx}^\mathrm{T}\mathbf{X} + \boldsymbol{h}_{yy}^\mathrm{T}\mathbf{Y} \,, \tag{6.71}$$

where $\boldsymbol{h}_{xx}$ etc. are vectors of length N containing the tap weights, $\boldsymbol{x}_\mathrm{o}$ and $\boldsymbol{y}_\mathrm{o}$ are vectors of length M containing the outputs, and $\mathbf{X}$ and $\mathbf{Y}$ are $N \times M$ matrices whose columns represent the parallel inputs. In this case, the error term for the CMA becomes $\boldsymbol{\epsilon_x} = \mathbf{1} - \boldsymbol{x}_\mathrm{o} \circ \boldsymbol{x}_\mathrm{o}^*$, and similarly $\boldsymbol{\epsilon_y} = \mathbf{1} - \boldsymbol{y}_\mathrm{o} \circ \boldsymbol{y}_\mathrm{o}^*$, where $\circ$ denotes the Hadamard product (being element-by-element multiplication), with the resulting update algorithm for the CMA becoming

$$\boldsymbol{h}_{xx} \leftarrow \boldsymbol{h}_{xx} + 2\mu\left(\boldsymbol{\epsilon}_x \circ \boldsymbol{x}_\mathrm{o}\right)^\mathrm{T}\mathbf{X}^* \,, \tag{6.72}$$

$$\boldsymbol{h}_{xy} \leftarrow \boldsymbol{h}_{xy} + 2\mu\left(\boldsymbol{\epsilon}_x \circ \boldsymbol{x}_\mathrm{o}\right)^\mathrm{T}\mathbf{Y}^* \,, \tag{6.73}$$

$$\boldsymbol{h}_{yx} \leftarrow \boldsymbol{h}_{yx} + 2\mu\left(\boldsymbol{\epsilon}_y \circ \boldsymbol{y}_\mathrm{o}\right)^\mathrm{T}\mathbf{X}^* \,, \tag{6.74}$$

$$\boldsymbol{h}_{yy} \leftarrow \boldsymbol{h}_{yy} + 2\mu\left(\boldsymbol{\epsilon}_y \circ \boldsymbol{y}_\mathrm{o}\right)^\mathrm{T}\mathbf{Y}^* \,. \tag{6.75}$$

Once the CMA has converged, a decision directed equalizer is employed with

$$\boldsymbol{h}_{xx} \leftarrow \boldsymbol{h}_{xx} + \mu\boldsymbol{\epsilon}_x^\mathrm{T}\mathbf{X}^* \,, \tag{6.76}$$

$$\boldsymbol{h}_{xy} \leftarrow \boldsymbol{h}_{xy} + \mu\boldsymbol{\epsilon}_x^\mathrm{T}\mathbf{Y}^* \,, \tag{6.77}$$

$$\boldsymbol{h}_{yx} \leftarrow \boldsymbol{h}_{yx} + \mu\boldsymbol{\epsilon}_y^\mathrm{T}\mathbf{X}^* \,, \tag{6.78}$$

$$\boldsymbol{h}_{yy} \leftarrow \boldsymbol{h}_{yy} + \mu\boldsymbol{\epsilon}_y^\mathrm{T}\mathbf{Y}^* \,. \tag{6.79}$$

6.6.7 Generalized 4 × 4 Equalizer

Provided that the equalizer is not constrained to be unitary, all of the equalizers discussed in this section can also be used to mitigate the impact of frequency-dependent loss or polarization-dependent loss. The equalizer can, however, be generalized further to the 4×4 equalizer by relaxing the assumptions in the 2×2 complex equalizer that the real and imaginary signals are sampled synchronously and are orthonormal. In this case, the equalizer structure becomes [6.11, 26]

$$x_{\mathrm{o_r}} = \boldsymbol{h}_{x_\mathrm{r}x_\mathrm{r}}^\mathrm{T}\boldsymbol{x}_\mathrm{r} + \boldsymbol{h}_{x_\mathrm{r}x_\mathrm{i}}^\mathrm{T}\boldsymbol{x}_\mathrm{i} + \boldsymbol{h}_{x_\mathrm{r}y_\mathrm{r}}^\mathrm{T}\boldsymbol{y}_\mathrm{r} + \boldsymbol{h}_{x_\mathrm{r}y_\mathrm{i}}^\mathrm{T}\boldsymbol{y}_\mathrm{i} \,, \tag{6.80}$$

$$x_{\mathrm{o_i}} = \boldsymbol{h}_{x_\mathrm{i}x_\mathrm{r}}^\mathrm{T}\boldsymbol{x}_\mathrm{r} + \boldsymbol{h}_{x_\mathrm{i}x_\mathrm{i}}^\mathrm{T}\boldsymbol{x}_\mathrm{i} + \boldsymbol{h}_{x_\mathrm{i}y_\mathrm{r}}^\mathrm{T}\boldsymbol{y}_\mathrm{r} + \boldsymbol{h}_{x_\mathrm{i}y_\mathrm{i}}^\mathrm{T}\boldsymbol{y}_\mathrm{i} \,, \tag{6.81}$$

$$y_{\mathrm{o_r}} = \boldsymbol{h}_{y_\mathrm{r}x_\mathrm{r}}^\mathrm{T}\boldsymbol{x}_\mathrm{r} + \boldsymbol{h}_{y_\mathrm{r}x_\mathrm{i}}^\mathrm{T}\boldsymbol{x}_\mathrm{i} + \boldsymbol{h}_{y_\mathrm{r}y_\mathrm{r}}^\mathrm{T}\boldsymbol{y}_\mathrm{r} + \boldsymbol{h}_{y_\mathrm{r}y_\mathrm{i}}^\mathrm{T}\boldsymbol{y}_\mathrm{i} \,, \tag{6.82}$$

$$y_{\mathrm{o_i}} = \boldsymbol{h}_{y_\mathrm{i}x_\mathrm{r}}^\mathrm{T}\boldsymbol{x}_\mathrm{r} + \boldsymbol{h}_{y_\mathrm{i}x_\mathrm{i}}^\mathrm{T}\boldsymbol{x}_\mathrm{i} + \boldsymbol{h}_{y_\mathrm{i}y_\mathrm{r}}^\mathrm{T}\boldsymbol{y}_\mathrm{r} + \boldsymbol{h}_{y_\mathrm{i}y_\mathrm{i}}^\mathrm{T}\boldsymbol{y}_\mathrm{i} \,, \tag{6.83}$$

where $x_{\mathrm{o_r}} = \mathrm{Re}\,(x_\mathrm{o})$, $x_{\mathrm{o_i}} = \mathrm{Im}\,(x_\mathrm{o})$, $\boldsymbol{x}_\mathrm{r} = \mathrm{Re}(\boldsymbol{x})$, $\boldsymbol{x}_\mathrm{i} = \mathrm{Im}(\boldsymbol{x})$ etc. For the CMA, the update algorithms are of the form

$$\boldsymbol{h}_{x_\mathrm{r}k} \leftarrow \boldsymbol{h}_{x_\mathrm{r}k} + 4\mu\epsilon_x x_{\mathrm{o_r}}\boldsymbol{k} \,, \tag{6.84}$$

$$\boldsymbol{h}_{x_\mathrm{i}k} \leftarrow \boldsymbol{h}_{x_\mathrm{i}k} + 4\mu\epsilon_x x_{\mathrm{o_i}}\boldsymbol{k} \,, \tag{6.85}$$

$$\boldsymbol{h}_{y_\mathrm{r}k} \leftarrow \boldsymbol{h}_{y_\mathrm{r}k} + 4\mu\epsilon_y y_{\mathrm{o_r}}\boldsymbol{k} \,, \tag{6.86}$$

$$\boldsymbol{h}_{y_\mathrm{i}k} \leftarrow \boldsymbol{h}_{y_\mathrm{i}k} + 4\mu\epsilon_y y_{\mathrm{o_i}}\boldsymbol{k} \,, \tag{6.87}$$

where $k \in \{x_\mathrm{r}, x_\mathrm{i}, y_\mathrm{r}, y_\mathrm{i}\}$. The resulting equalizer can not only compensate for frequency and polarization-dependent effects but also receiver skew.

6.7 Carrier and Timing Synchronization Algorithms

The coherent optical transceivers developed for use in optical transport networks (OTN) are asynchronous and use free-running transmitter and receiver lasers. Therefore, the symbol timing frequency and phase must be recovered by the receiver, and the frequency and phase difference between the lasers at the transmitter and receiver must also be estimated at the receiver for complex modulation formats such as QAM.

6.7.1 Symbol Timing Recovery

Since the clocks that determine the symbol rate in the transmitter and receiver of a transmission system are independent and free-running, some control algorithm is required for phase-locking at the receiver [6.27]. One scheme that is often used for phase-locking involves the coarse adjustment of a voltage controlled oscillator in the analog domain, combined with a fine adjustment of a digital retiming circuit (normally an interpolating polyphase filter such as a Farrow filter [6.28]). Both the analog and digital components of such a system can be driven by the same timing-error detector (TED). Due to the lack of an observable *eye* in coherent optical transmission systems before sampling and processing, both the analog and digital parts of the hybrid timing recovery circuit may be driven by a digital-timing error estimator. The hybrid scheme is shown in Fig. 6.8.

Gardner Algorithm

The most commonly used algorithm for the symbol timing error detector (TED) in the coherent optical transmission literature is the Gardner algorithm [6.29]. This algorithm determines an error to be minimized with two samples per symbol input and assumes a constant symbol energy. The TED output ϵ is defined as

$$\epsilon = \mathrm{Re}\left[\sum_{n=0}^{N/2-1}(x_{2n}-x_{2n+2})x_{2n+1}^{*}\right], \tag{6.88}$$

where N is the length of the averaging window in samples, x_n is the input sample x at instant n, * is the complex conjugate, and Re denotes the real part of a complex number.

This error function allows for good performance in terms of locking, tracking, and residual jitter. However, distortions such as polarization mode dispersion (PMD), or residual chromatic dispersion (CD) which was not compensated by the proceeding static equalizer may cause failure. Equally utilizing the Gardner algorithm in systems with multilevel modulation or Nyquist signaling can lead to a reduction in performance or increased frequency of failure. For these reasons, a great many extensions have been proposed to improve the performance of the Gardner and related algorithms in the context of high-capacity optical transmission.

One such extension is to implement the classic Gardner algorithm in the frequency domain [6.30]. By calculating the error function in the frequency domain, it is possible to estimate the distortion due to residual CD and first-order PMD with a simple model that reflects the physical nature of the distortions.

Two extensions were proposed simultaneously in [6.31] and [6.32], which enable performance of a Gardner-derived algorithm for signals with near Nyquist-limited bandwidth. While the architecture of the algorithm is not changed, a correction in the timing error detector calculation as proposed in either of these algorithms eliminates the sensitivity of the timing recovery system on the transmitted pulse shape, thereby permitting the use of signals with minimum support in the frequency domain.

6.7.2 Carrier Phase and Frequency Estimation

The carrier phase and frequency estimation problems may be formulated as follows for a signal after proper equalization

$$y(t) = x(t)\mathrm{e}^{\mathrm{j}[\omega_{\mathrm{IF}}t+\theta(t)]} + N\,, \tag{6.89}$$

where t is time, x is the transmitted symbol sequence, y is the received symbol sequence including all distortions and noise, ω_{IF} is the angular intradyne frequency –

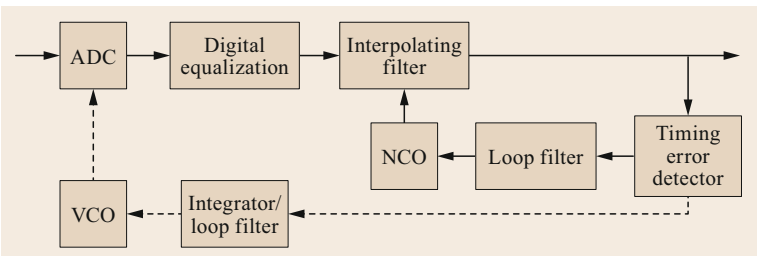


Fig. 6.8 A block diagram showing a hybrid timing-recovery circuit with coarse frequency correction of the clock driving the ADCs using a voltage-controlled oscillator (VCO) and fine phase-locking entirely in the digital domain using a circuit driven by a number-controlled oscillator (NCO)

that is, the difference in angular frequency between transmitter and receiver lasers, θ is the time varying phase noise, and N is the additive noise term.

Carrier phase and frequency estimation algorithms may be considered to fall under two broad categories: pilot-aided and blind [6.12]. Pilot-aided algorithms make use of components of the signal,of which the receiver algorithm has a-priori knowledge. These components may be predetermined symbols, interspersed with unknown data symbols, or sinusoids which have a fixed relation to the optical carrier frequency and phase. Blind algorithms use a nonlinear function operating on the symbols received to remove the influence of modulation on the phase received.

While there is a well-developed literature dedicated to intradyne frequency estimation algorithms [6.33, 34], research in this area has slowed significantly in recent years. Due to the slow variation of frequency in lasers, extremely accurate estimates may be performed with very large observation windows to minimize the effects of noise and distortion. Since these estimates must be obtained infrequently, computational cost is low, and performance good for all practical systems. We, therefore, focus on the problem of carrier phase estimation, which is a much faster varying estimation problem and has continued to prove challenging as modulation and coding techniques have evolved.

Algorithms for CPE

Viterbi and Viterbi. The Viterbi and Viterbi algorithm uses an M-th power instantaneous nonlinearity to eliminate the impact of modulation on phase received. The simple nature of the algorithm, combined with its use of circular symmetry in the constellation mean that it is an algorithm particularly well suited to phase-shift keying (PSK)-based modulation. The most general form, given in [6.35], is

$$\hat{\phi}_k = \frac{\arg\left(\sum_{i=k-L}^{k+L} F(|y_i|) y_i^M\right)}{M}\,, \tag{6.90}$$

where $\hat{\phi}_k$ is the phase estimate at symbol index k, F is an arbitrary nonlinear function, L is the half-length of an averaging filter, M is the nonlinear power, and arg(.) is the complex argument function.

This results in a phase estimate that is bounded on $\pm\pi/M$, which must then be unwrapped. This is done by calculating a per-symbol unwrapping phase as

$$a_k = a_{k-1} + \left\lfloor \frac{1}{2} + \frac{M}{2\pi}(\hat{\phi}_k - \hat{\phi}_{k-1}) \right\rfloor\,. \tag{6.91}$$

The unwrapped phase can then be calculated as

$$\hat{\theta}_k = \hat{\phi}_k + a_k\,. \tag{6.92}$$

Blind Phase Search. Due to the sensitivity to phase noise of high spectral efficiency modulation formats such as QAM, accurate carrier phase estimation is essential. The recent interest in advanced modulation formats and modulation format flexible transceivers also provides motivation for algorithms that are independent of the modulation format being used.

The blind phase search (BPS) algorithm [6.36] uses a search over a set of test phases $\boldsymbol{\Phi}$ with B elements over a range of $\pi/2$, given by

$$\boldsymbol{\Phi} = \frac{b\pi}{2B}\,; \quad b \in \{0, 1, \ldots, B-1\}\,, \tag{6.93}$$

the test vector is used to determine the most likely test phase according to the minimum mean squared error. A wrapped phase estimate $\hat{\phi}_k$ is, therefore, calculated with

$$\hat{\phi}_k = \underset{\Phi}{\operatorname{argmin}} \left[\sum_{i=k-L}^{k+L} \left| \mathrm{e}^{-\mathrm{j}\Phi} y_i - D\left(\mathrm{e}^{-\mathrm{j}\Phi} y_i\right) \right|^2 \right], \tag{6.94}$$

where D is the minimum Euclidean distance symbol decision function, for the modulation format under consideration. As with the Viterbi and Viterbi algorithm, the phase estimate must be unwrapped before being applied to the signal.

As the calculations may be done in parallel for the B elements in the vector $\boldsymbol{\Phi}$, and in parallel for different samples in time, this algorithm may be suitable for implementation in hardware. However, the complexity of the BPS algorithm is linearly dependent on the number of test phases. There has, therefore, been a great deal of interest in developing reduced-complexity alternatives to the BPS algorithm.

Advanced Algorithms for Future Implementation

While advanced modulation formats have enabled the implementation of 200 and 400 Gb/s single carrier systems, the desire of equipment manufacturers to improve performance and reduce complexity also drives research into new carrier recovery algorithms. As system implementations approach the Shannon bound [6.37], the accuracy of purely blind carrier systems is diminished, as signal and noise become harder to distinguish. As a result, research into carrier phase estimation algorithms for capacity approaching systems have focused on two areas: dual-stage algorithms and statistical signal processing approaches.

Dual-stage algorithms have been investigated [6.38–40] in detail for their ability to determine a somewhat accurate initial baseline phase without ambiguity for minimal computation, possibly with the aid of pilot symbols at the cost of a small overhead

in transmission rate. A second stage can then provide more accurate estimation of the carrier phase given the initial estimate is closest to the true phase.

Statistical signal processing algorithms such as Bayesian filtering [6.41] and Kalman filtering [6.40, 42] have been used to accurately recover carrier phase for both non-Lorentzian phase noise, and for Lorentzian systems with very low SNR. Kalman smoothing has also been used to enhance the initial estimate of phase given by a pilot-aided algorithms. While these algorithms are currently prohibitive in their complexity, they indicate the manner of improvement that may be possible in the future as complexity is reduced and computational resources increase.

6.8 Forward Error Correction Coding

The origins of forward error correction coding lie in the renowned paper by *Shannon* [6.37], which provided the genesis for the fields of communication theory, information theory, and coding theory. Shannon's *noisy channel coding theorem* is given by

$$C = B \log_2 \left(1 + \frac{P_{\text{sig}}}{P_{\text{noise}}} \right) . \tag{6.95}$$

This theorem gives us the maximal rate C in bits per second that can be transmitted with good reliability over an AWGN channel with bandwidth B, signal power P_{sig}, and noise power P_{noise}. A corollary to this is that reliable communication may be achieved over any channel with an appropriate channel code or forward error correction (FEC) code.

The basic principle of FEC coding is that when a random sequence is injectively mapped onto a larger space, the set of all possible sequences in this larger space can become sufficiently sparse that different sequences in this larger space may be estimated in the presence of noise with an arbitrarily low probability of error. The increase in dimension between the initial sequence and the sequence after injective mapping may be considered as the overhead required for forward error correction, which must increase as the signal-to-noise ratio decreases in order to maintain reliable communication.

In order to characterize the amount of overhead required for FEC that allows reliable communication at a fixed signal-to-noise ratio, we must account for several factors: the modulation format under consideration, and the nature of the information considered by the decoder (i.e., hard decision or soft decision), the cardinality of the coding scheme under consideration (i.e., binary or over some higher order field). This calculation can, therefore, provide us with an achievable information rate for a given channel, modulation format, and coding strategy.

We note that the FEC strategies used in optical communications systems are generally designed to be optimal for the additive white Gaussian noise (AWGN) channel. While the optical fiber channel is not, in fact, an AWGN channel, it has been shown to be well approximated by AWGN [6.43], and such algorithms are often well suited to implementation in hardware.

The FEC schemes used in optical communications are currently exclusively block coding-based schemes (including the somewhat misleadingly named LDPC convolutional code). We may describe such a linear block coding scheme in the following way

$$\boldsymbol{w} = \boldsymbol{u}\mathbf{G} , \tag{6.96}$$

where $\boldsymbol{u}$ is the uncoded input word, $\mathbf{G}$ is the generator matrix, and $\boldsymbol{w}$ is the resulting codeword. For an input word $\boldsymbol{u}$ of length k and generator matrix $\mathbf{G}$ of dimension $n \times k$, $n-k$ redundant dimensions are introduced to the codeword $\boldsymbol{w}$, resulting in a code rate of

$$R = \frac{k}{n} . \tag{6.97}$$

The parity constraints imposed by the code are described by the parity check matrix $\mathbf{G}$, which is defined by

$$\mathbf{I} = \mathbf{G}\mathbf{H} , \tag{6.98}$$

where $\mathbf{I}$ is the identity matrix. For an estimated codeword $\tilde{\boldsymbol{w}}$, we can calculate the syndrome $\boldsymbol{s}$ as

$$\boldsymbol{s} = \tilde{\boldsymbol{w}}\mathbf{H} . \tag{6.99}$$

Each of the $n-k$ elements in the vector $\boldsymbol{s}$ determines the sum of inputs to each parity equation described by the generator matrix. In the case when the syndrome vector $\boldsymbol{s}$ consists only of zeros, therefore, $\tilde{\boldsymbol{w}}$ represents an estimated codeword that satisfies all parity constraints determined by $\mathbf{H}$, and is, therefore, valid (although not necessarily equal to the transmitted codeword).

6.8.1 Coding Schemes

While early coherent transmission systems employed hard-decision FEC in order to minimize complex-

ity [6.44], subsequent systems have focused on the use of soft-decision coding schemes for increased gain [6.45]. By decoding using not only the bit received, but their relative likelihoods, an increase in coding gain of 2 dB or more can be achieved. As modulation density has increased for advanced transceiver designs, using 8, 16, 32, and 64 QAM, bit-interleaved coded modulation (BICM) [6.46] has been utilized to minimize the complexity of the coding system in hardware. It uses a binary code with an interleaver that assigns equal proportions of each bit-position in the constellation (i.e., from most significant to least significant) to the codeword. This enables the decoder to have an equal proportion of the most reliable bits (that is, the most significant) and least reliable (that is, the least significant). While this is known not to be the optimal strategy, it offers only a small penalty from the optimum in many cases, and has very low complexity.

Low-Density Parity Check Coding

The most widely adopted soft-decision coding scheme for coherent optical transmission systems has been the combination of low-density parity check (LDPC) codes with BICM. LDPC codes were first discovered in the early 1960s by *Gallager* [6.47] as a PhD student and were widely considered impractical until their rediscovery in the 1990s [6.48]. One of the primary drivers for the adoption of LDPC codes recently has been the development of belief propagation (BP)-based approximate decoding algorithms such as the sum-product algorithm (SPA) [6.49]. BP decoding algorithms pass many messages in parallel between two sets of parallel decoders, which perform simple operations. As the decoding algorithm is inherently parallel, the long codewords required by LDPC (on the order of 10 000 bits) are not prohibitive in either latency or complexity.

Advanced Coding Techniques

While coding schemes that approach the Shannon bound for a binary input AWGN channel have been described for more than 50 years, much research is ongoing in the fields of coding and modulation.

A significant development in recent years has been the development of polar codes, which were discovered by *Stolte* [6.50] in 2002 and, later – independently – by *Arikan* [6.51]. Polar codes can achieve capacity approaching performance at far shorter codeword lengths than those of traditional methods such as LDPC [6.52]. While the structure of polar codes does not favor belief-propagation type decoding, much progress has been made recently in decoder algorithms. Particularly influential has been the work of *Tal* and *Vardy*, whose successive cancelation list decoder [6.53] may be realizable in hardware.

Other topics of research have been focused on reducing the loss in capacity due to the use of binary codes with interleaving. Nonbinary coding schemes such as nonbinary LDPC (NB-LDPC) [6.54–56] can achieve better performance than binary codes and are optimal when the cardinality of modulation is equal to the order of the field over which the code is constructed. While nonbinary coding can achieve nontrivial gains in some cases, implementation has proven difficult due to the required increase in complexity. For BP type decoders, a modulation of code constructed in GF(M) requires $M-1$ messages to be passed over each edge in the Tanner graph per decoder iteration.

BICM with iterative demodulation utilizes feedback from the FEC decoder to update the demodulator [6.46]. This technique may also approach the optimal performance bounded by NB-LDPC, albeit with significant complexity. Iterative demodulation may be done with reduced frequency – that is, not at every iteration of the decoder – although this is associated with a reduction in performance [6.57].

6.9 Current Research Trends

6.9.1 Constellation Shaping

Constellation shaping provides a methodology for digital communications systems to approach the Shannon bound for the AWGN channel. While the Shannon bound assumes a source with a continuous Gaussian distribution, a digital communication system with finite code rate must have a discrete source distribution. There are two types of source distribution that are of particular interest in constellation shaping: distributions with equiprobable symbols, which are unevenly distributed in the signal space, known as geometric shaping; and distributions with unequally probable symbols, which are evenly distributed in the signal space, known as probabilistic shaping.

Geometric Shaping

Geometric shaping systems are in their most general form designed to be a throughput maximizing arrangement of equally probable constellation points for a given channel [6.58] (for simplicity, normally assumed to be the AWGN channel). For modulation of

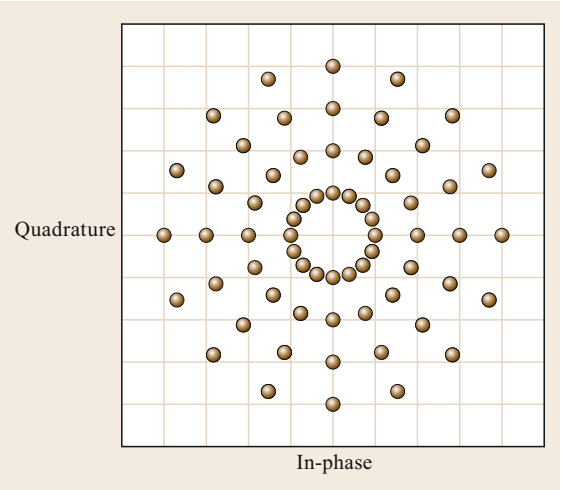


Fig. 6.9 A geometrically shaped constellation with 64 points, able to represent 6 bits of information. There are 16 equally-spaced phase levels, representing 4 bits of information, and 8 amplitude levels, representing 3 bits of information

cardinality M, this is an optimization with $M-2$ dimensions, if we assume that the constellation has zero mean and unit power. As optimization in this number of dimensions quickly becomes impossible for the scale of M which is of interest (for example, 16–1024), some symmetry is often assumed in the constellation such that the number of dimensions can be reduced to the level whereby numerical optimization can be employed.

A common simplification to the constellation is to assume that phase is modulated with equally spaced levels on a number of rings [6.59, 60], and that the amplitude levels of the rings are optimized in order to maximize throughput at a given SNR. This configuration is often denoted amplitude and phase-shift keying (APSK). An example of 64-APSK with 16 amplitude levels and 4 phase levels is shown in Fig. 6.9.

While the performance of APSK is demonstrably superior to that of uniformly distributed square QAM, it is known that APSK approaches the Shannon bound only when symbol metric decoding is used [6.59], that is, in cases where nonbinary coding or BICM with iterative demodulation are used. As has been previously noted, these coding schemes are considered to be prohibitive in terms of complexity, and, therefore, alternative shaping methods are of considerable interest.

Probabilistic Shaping

Probabilistic shaping systems are also in their most general form designed to be a throughput maximizing probability distribution of arbitrarily arranged constellation points for a given channel (for simplicity, normally assumed to be the AWGN channel). In the contemporary literature for optical communications, several other characteristics are near universal. The constellation points are normally considered to be square QAM; constellations are assumed to be shaped in the amplitude domain only (on a per quadrature basis), and, therefore, have reflective symmetry about the real and imaginary axes. This allows for the FEC code to be applied to the stream of shaped bits, and the uniformly distributed parity bits assigned to the sign bits of the constellation – this configuration is also known as *reverse concatenation* of FEC and distribution matching. A further common feature is that each shaped symbol sequence that corresponds to a block of uniformly distributed bits has the desired amplitude probability mass function (PMF) – this is known as constant composition distribution matching (CCDM) [6.61].

The reverse concatenation approach to probabilistic amplitude shaping (PAS) is critical to realizable implementations and is described in the Fig. 6.10. Uniformly distributed bits from a source are mapped by a distribution matcher to symbol sequences, which have the desired distribution of amplitudes. The bits representing the shaped amplitudes are then encoded with an FEC encoder, and the parity bits (which have a uniform distribution) are assigned to the sign bits of the constellation. At the receiver side, the FEC decoder operates on the noisy received symbols in the conventional manner, and the noise-free amplitude sequences are passed to a distribution dematcher, which outputs uniformly distributed data bits. This architecture provides several important benefits: firstly, the Shannon bound may be approached with a BICM coding scheme, which is an important condition for practicality; and secondly, the distribution dematcher can operate on noise-free symbol sequence inputs. This reduces the complexity of the dematching algorithm requirements. By varying the entropy of the target distribution and the rate of the distribution matcher, this also enables the transmission rate to be varied while maintaining a constant code rate and symbol rate [6.62].

While these advances have pointed to possible implementations of probabilistic shaping in the near future, significant hurdles remain. While CCDM remains the most widely studied technique at the time of writing, the high degree of serialism required in the distribution matcher and dematcher, combined with long block lengths required for good performance mean that it is highly challenging to implement CCDM in hardware. There is, therefore, significant impetus in the development of alternative architectures for PAS [6.63–65] that do not suffer the problems of CCDM.

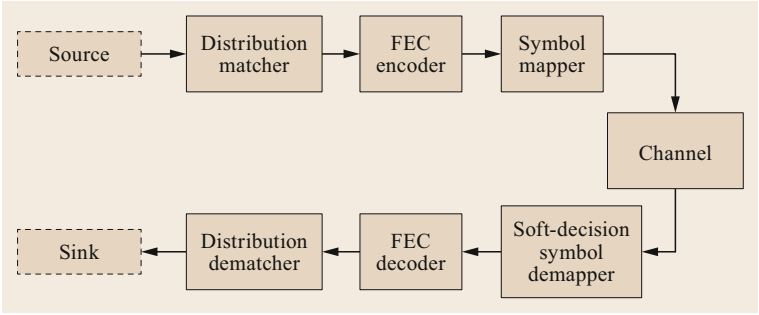


Fig. 6.10 A schematic of a communication system employing probabilistic amplitude shaping. The information bits are first shaped by a distribution matcher, then encoded by an FEC encoder, before being mapped to symbols and transmitted over the channel. The parity bits have uniform distribution and are mapped to the sign bits in the desired constellation. At the receiver, the symbols are demapped before the FEC code is decoded. The noise-free shaped bits are then passed to a distribution dematcher to recover the original information bits

6.9.2 Power Efficient Transceiver Design

As coherent technology has moved into shorter reach transmission systems, the impact of transceiver power consumption on the overall system power consumption has become increasingly important. While steady improvements in CMOS technology have provided a reduction in power consumption per computational operation, reducing the power consumption of the transceiver ASIC is also an active topic for research.

By reducing the sampling rates at the signal converters and throughout the DSP chain, it is possible to reduce the number of samples processed, and, therefore, the power consumed. While these techniques can be employed without fundamental penalty, they require new algorithms for equalization and symbol timing recovery.

Algorithmic improvements in important DSP blocks such as the adaptive equalizer [6.66, 67] and CPE [6.68] can demonstrate significant reductions in the number and complexity of computational operations required to achieve the desired functionality. Furthermore, eliminating the DSP entirely (or almost entirely) has been proposed [6.69].

Computational complexity has also been examined as a primary constraint in the design of FEC codes [6.70]. The ability to trade gain for complexity during the design process is a direct result of the desire to limit complexity for a desired level of performance.

We must also note that recent research has illuminated the importance of hardware emulation as a tool for estimating power consumption improvements. While the broadest effects may be considered by the counting of multipliers and adders, the frequency with which gates are flipped, wiring losses, and the characteristics of different CMOS processes are ignored without a fuller examination of the ASIC design process [6.71].

6.9.3 Nonlinearity Mitigation

Digital nonlinearity mitigation has been popularized over the last several years with two algorithmic designs: split-step type algorithms and perturbation type algorithms [6.72]. While both of these methods have the same performance under ideal conditions, perturbation type algorithms are generally considered to be capable of better performance at the low complexity required for hardware implementation. Despite this popularity, the gain available from digital nonlinearity compensation seems to be limited to the region of 1 dB or less for realistic system configurations [6.73].

An alternative approach is nonlinear frequency division multiplexing. The nonlinear Fourier transform (NLFT) has been known as an application of the inverse scattering transform to the nonlinear Schrödinger equation (NLSE) since the 1970s [6.74]. More recently, the elements of the discrete NLFT were proposed as a basis for transmitting information in parallel nonlinear channels [6.75]. This basis – which may be considered as a set of orthogonal solitons [6.76] – can be used to transmit information over the nonlinear optical channel without accumulating distortion due to nonlinearity during transmission. This technique has been referred to as nonlinear frequency-division multiplexing or NFDM [6.77].

More recently, an algorithm with reduced complexity – the fast nonlinear Fourier transform [6.78] – was developed to enable efficient modulation of an optical carrier in the domain of this nonlinear basis, with $\mathcal{O}(n \log n)$ complexity.

A great deal of research effort has been expended in developing transmission systems based on these ideas [6.79–81]. However, limitations on the amount of bandwidth that is possible to jointly process, along with nonlinear distortions induced by copropagating noise have limited the gains achieved thus far to approximately those already available with other methods of digital nonlinearity mitigation [6.82].

6.10 Future Opportunities

6.10.1 High-Capacity Core Transport

Commercial systems operating at 1 Tb/s on a single wavelength may be a reality in only a few years with systems transmitting 600 Gb/s per wavelength already under development. Increasingly, transceivers are using all of the available options, being bit-rate adaptive, using constellation shaping to enable a single transceiver to flex from QPSK up to 128 QAM for each polarization. Such transceivers are bit-rate adaptive, enabling 400 Gb/s to be transmitted over 1200 km [6.72]. Most demonstrations of 1 Tb/s have used in the region of 100 GHz optical bandwidth, which will require RF and mixed signal circuits to operate in the region of 60 GHz or more RF bandwidth. While test equipment is available at these levels, new optical devices will be required to take advantage of such high bandwidths. In terms of aggregate capacity, space division multiplexed systems can offer significant increases in capacity. This again offers significant opportunities for DSP, in particular for systems with mode mixing.

6.10.2 Metro-Access

While DSP has become a key technology for long-haul systems, as power consumption reduces it can be expected to be applied to shorter-reach systems. Efforts are underway at the time of writing to provide standards for 400 Gb/s transmission systems, suitable for distances of 10–100 km. These systems are particularly important in the growing market for data center interconnect products. In metro-access systems, the increase in the demands of wireless networks indicate a growing potential space for coherent systems in wireless backhaul, and potential in midhaul and even fronthaul.

6.10.3 Short-Reach Systems

At present, a very promising avenue for research relates to DSP for direct-detection systems. While for the shortest of links, the power consumption and footprint of full digital coherent systems appears prohibitive, DSP-enabled direct-detection systems may be capable of bridging this gap and providing gains in performance and functionality for short-reach and low-cost applications. One such proposal is the introduction of equalization to the receiver DSP chain for Ethernet systems operating at 100–400 Gb/s using 4-ary pulse amplitude modulation (4-PAM). A short feed-forward equalizer (FFE) and decision-feedback equalizer (DFE) pair can mitigate the bandwidth constraints of the optical and electronic components, while having only small penalties in complexity and latency [6.83]. Single-sideband modulation, combined with the Kramers–Kronig receiver, which utilizes a minimum phase criterion to link the amplitude and phase via a Hilbert transform [6.84], offers the possibility of compensating the effects of chromatic dispersion, while requiring only an increase in DSP complexity at the receiver, and the use of an I/Q modulator. Alternately, by utilizing a dual-polarization I/Q modulator and a more complex receiver, Stokes-vector direct detection may be implemented in order to increase spectral efficiency and compensate for transmission impairments [6.85].

6.11 Closing Remarks

DSP has emerged as a key technology for modern optical fiber communication systems. Enabling, for example, 1 Tb/s to be transmitted over fiber which could not previously have been used for 10 Gb/s due to PMD. It has had a significant impact also on the design of systems, optical chromatic dispersion compensation is now all but obsolete, but also moving to a digital transceiver has resulted in a better quality of transceiver, requiring less system margin to account for variability. Improvements in power consumption have been predicated on Moore's law, with modern coherent transceivers utilizing over 1 billion transistors, compared to just 20 million in the first coherent ASIC deployed just a decade ago [6.2]. Going forward we, can expect DSP to impact all optical fiber communication systems from short-reach systems within a data center to transoceanic submarine links.

References

6.1 K. Poulton, R. Neff, B. Setterberg, B. Wuppermann, T. Kopley, R. Jewett, J. Pernillo, C. Tan, A. Montijo: A 20 GS/s 8 b ADC with a 1 MB memory in 0.18 μm CMOS. In: *IEEE Int. Solid-State Circuits Conf. (ISSCC)* (2003), https://doi.org/10.1109/ISSCC.2003.1234315

6.2 H. Sun, K.T. Wu, K. Roberts: Real-time measurements of a 40 Gb/s coherent system, Opt. Express **16**(2), 873–879 (2008)

6.3 K. Roberts, M. O'Sullivan, K.T. Wu, H. Sun, A. Awadalla, D.J. Krause, C. Laperle: Performance of dual-polarization QPSK for optical transport systems, J. Lightwave Technol. **27**(16), 3546–3559 (2009)

6.4 M. O'Sullivan: Progress in real time, DSP assisted, coherent optical transmission. In: *Opto-Electron. Commun. Conf.* (2009), https://doi.org/10.1109/OECC.2009.5219777

6.5 P.J. Winzer: Beyond 100G ethernet, IEEE Commun. Mag. **48**(7), 26–30 (2010)

6.6 S.S. Kidambi, R.P. Ramachandran: Complex coefficient nonrecursive digital filter design using a least-squares method, IEEE Trans. Signal Process. **44**(3), 710–713 (1996)

6.7 D. Godard: Self-recovering equalization and carrier tracking in two-dimensional data communication systems, IEEE Trans. Commun. **28**(11), 1867–1875 (1980)

6.8 A.V. Oppenheim, R.W. Schafer: *Discrete-Time Signal Processing*, 3rd edn. (Prentice Hall, Upper Saddle River 2009)

6.9 I. Fatadin, S.J. Savory, D. Ives: Compensation of quadrature imbalance in an optical QPSK coherent receiver, IEEE Photonics Technol. Lett. **20**(20), 1733–1735 (2008)

6.10 N. Stojanovic, X. Changsong: An efficient method for skew estimation and compensation in coherent receivers, IEEE Photonics Technol. Lett. **28**(4), 489–492 (2016)

6.11 M. Paskov, D. Lavery, S.J. Savory: Blind equalization of receiver in-phase/quadrature skew in the presence of Nyquist filtering, IEEE Photonics Technol. Lett. **25**(24), 2446–2449 (2013)

6.12 J.G. Proakis, M. Salehi: *Digital Communications* (McGraw-Hill, New York 2008)

6.13 G.P. Agrawal: *Fiber-Optic Communication Systems*, Wiley Series in Microwave and Optical Engineering (Wiley, Hoboken 2012)

6.14 S.J. Savory: Digital filters for coherent optical receivers, Opt. Express **16**(2), 804–817 (2008)

6.15 A. Eghbali, H. Johansson, O. Gustafsson, S.J. Savory: Optimal least-squares FIR digital filters for compensation of chromatic dispersion in digital coherent optical receivers, J. Lightwave Technol. **32**(8), 1449–1456 (2014)

6.16 C.D. Poole, R.E. Wagner: Phenomenological approach to polarisation dispersion in long single-mode fibres, Electron. Lett. **22**(19), 1029–1030 (1986)

6.17 G.J. Foschini, C.D. Poole: Statistical theory of polarization dispersion in single mode fibers, J. Lightwave Technol. **9**(11), 1439–1456 (1991)

6.18 S. Yamamoto, S. Yamanaka, A. Matsuura, T. Kobayashi, A. Iwaki, M. Suzuki, T. Inui, T. Sakano, M. Tomizawa, Y. Miyamoto, T. Kotanigawa, A. Maeda: PMD tolerance of 100-Gbit/s digital coherent PDM-QPSK in DSF-installed field testbed. In: *16th Opto-Electron. Commun. Conf.* (2011) pp. 212–213

6.19 S. Betti, F. Curti, G. De Marchis, E. Iannone: A novel multilevel coherent optical system: 4-quadrature signaling, J. Lightwave Technol. **9**(4), 514–523 (1991)

6.20 Y. Han, G. Li: Coherent optical communication using polarization multiple-input-multiple-output, Opt. Express **13**(19), 7527–7534 (2005)

6.21 C.B. Papadias, A.J. Paulraj: A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays, IEEE Signal Process. Lett. **4**(6), 178–181 (1997)

6.22 A. Vgenis, C.S. Petrou, C.B. Papadias, I. Roudas, L. Raptis: Nonsingular constant modulus equalizer for PDM-QPSK coherent optical receivers, IEEE Photonics Technol. Lett. **22**(1), 45–47 (2010)

6.23 J. Zhou, G. Zheng, J. Wu: Constant modulus algorithm with reduced probability of singularity enabled by PDL mitigation, J. Lightwave Technol. **35**(13), 2685–2694 (2017)

6.24 S.J. Savory: Digital coherent optical receivers: Algorithms and subsystems, IEEE J. Sel. Top. Quantum Electron. **16**(5), 1164–1179 (2010)

6.25 B.C. Thomsen, R. Maher, D.S. Millar, S.J. Savory: Burst mode receiver for 112 Gb/s DP-QPSK with parallel DSP, Opt. Express **19**(26), B770–B776 (2011)

6.26 M.S. Faruk, K. Kikuchi: Compensation for in-phase/quadrature imbalance in coherent-receiver front end for optical quadrature amplitude modulation, IEEE Photonics J. **5**(2), 7800110 (2013)

6.27 F.M. Gardner: *Phaselock Techniques* (Wiley, Hoboken 2005)

6.28 L. Erup, F.M. Gardner, R.A. Harris: Interpolation in digital modems. II. Implementation and performance, IEEE Trans. Commun. **41**(6), 998–1008 (1993)

6.29 F.M. Gardner: A BPSK/QPSK timing-error detector for sampled receivers, IEEE Trans. Commun. **34**(5), 423–429 (1986)

6.30 H. Sun, K.T. Wu: A novel dispersion and PMD tolerant clock phase detector for coherent transmission systems. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf.* (2011), Paper OMJ4

6.31 M. Yan, Z. Tao, L. Dou, L. Li, Y. Zhao, T. Hoshida, J.C. Rasmussen: Digital clock recovery algorithm for Nyquist signal. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf.* (2013), Paper OTu2I.7

6.32 N. Stojanovic, C. Xie, Y. Zhao, B. Mao, N.G. Gonzalez, J. Qi, N. Binh: Modified Gardner phase detector for Nyquist coherent optical transmission systems. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf.* (2013), Paper JTh2A.50

6.33 A. Leven, N. Kaneda, U. Koc, Y. Chen: Frequency estimation in intradyne reception, IEEE Photonics Technol. Lett. **19**(5/8), 366 (2007)

6.34 I. Fatadin, S.J. Savory: Compensation of frequency offset for 16-QAM optical coherent systems us-

ing QPSK partitioning, IEEE Photonics Technol. Lett. **23**(17), 1246–1248 (2011)
6.35 A. Viterbi: Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission, IEEE Trans. Inf. Theor. **29**(4), 543–551 (1983)
6.36 T. Pfau, S. Hoffmann, R. Noe: Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for *M*-QAM constellations, J. Lightwave Technol. **27**(8), 989–999 (2009)
6.37 C.E. Shannon: A mathematical theory of communication, The Bell Syst. Tech. J. **27**(3), 379–423 (1948)
6.38 S.M. Bilal, C.R.S. Fludger, V. Curri, G. Bosco: Multi-stage carrier phase estimation algorithms for phase noise mitigation in 64-quadrature amplitude modulation optical systems, J. Lightwave Technol. **32**(17), 2973–2980 (2014)
6.39 M. Magarini, L. Barletta, A. Spalvieri, F. Vacondio, T. Pfau, M. Pepe, M. Bertolini, G. Gavioli: Pilot-symbols-aided carrier-phase recovery for 100-G PM-QPSK digital coherent receivers, IEEE Photonics Technol. Lett. **24**(9), 739–741 (2012)
6.40 M. Pajovic, D.S. Millar, T. Koike-Akino, R. Maher, D. Lavery, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B.C. Thomsen, S.J. Savory, P. Bayvel: Experimental demonstration of multi-pilot aided carrier phase estimation for DP-64QAM and DP-256QAM. In: *Eur. Conf. Opt. Commun. (ECOC)* (2015), https://doi.org/10.1109/ECOC.2015.7341655
6.41 D. Zibar, L. Carvalho, M. Piels, A. Doberstein, J. Diniz, B. Nebendahl, C. Franciscangelis, J. Estaran, H. Haisch, N.G. Gonzalez, J.R.F. de Oliveira, I.T. Monroy: Bayesian filtering for phase noise characterization and carrier synchronization of up to 192 Gb/s PDM 64-QAM. In: *Eur. Conf. Opt. Commun. (ECOC)* (2014), https://doi.org/10.1109/ECOC.2014.6963844
6.42 L. Pakala, B. Schmauss: Extended Kalman filtering for joint mitigation of phase and amplitude noise in coherent QAM systems, Opt. Express **24**(6), 6391–6401 (2016)
6.43 P. Poggiolini: The GN model of non-linear propagation in uncompensated coherent optical systems, J. Lightwave Technol. **30**(24), 3857–3879 (2012)
6.44 M. Scholten, T. Coe, J. Dillard: Continuously-interleaved BCH (CI-BCH) FEC delivers best-in-class NECG for 40G and 100G metro applications. In: *Nat. Fiber Opt. Eng. Conf.* (2010), https://doi.org/10.1364/NFOEC.2010.NTuB3
6.45 F. Chang, K. Onohara, T. Mizuochi: Forward error correction for 100 G transport networks, IEEE Commun. Mag. **48**(3), S48–S55 (2010)
6.46 L. Szczecinski, A. Alvarado: *Bit-Interleaved Coded Modulation: Fundamentals, Analysis and Design* (Wiley, Hoboken 2015)
6.47 R. Gallager: Low-density parity-check codes, IRE Trans. Inf. Theor. **8**(1), 21–28 (1962)
6.48 D.J.C. MacKay: Good error-correcting codes based on very sparse matrices, IEEE Trans. Inf. Theor. **45**(2), 399–431 (1999)
6.49 T.J. Richardson, R.L. Urbanke: The capacity of low-density parity-check codes under message-passing decoding, IEEE Trans. Inf. Theor. **47**(2), 599–618 (2001)
6.50 N. Stolte: *Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung*, Ph.D. Thesis (TU Darmstadt, Darmstadt 2002)
6.51 E. Arikan: Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, IEEE Trans. Inf. Theor. **55**(7), 3051–3073 (2009)
6.52 T. Koike-Akino, C. Cao, Y. Wang, S.C. Draper, D.S. Millar, K. Kojima, K. Parsons, L. Galdino, D.J. Elson, D. Lavery, P. Bayvel: Irregular polar coding for complexity-constrained lightwave systems, J. Lightwave Technol. **36**(11), 2248–2258 (2018)
6.53 I. Tal, A. Vardy: List decoding of polar codes, IEEE Trans. Inf. Theor. **61**(5), 2213–2226 (2015)
6.54 M. Arabaci, I.B. Djordjevic, R. Saunders, R.M. Marcoccia: High-rate nonbinary regular quasi-cyclic LDPC codes for optical communications, J. Lightwave Technol. **27**(23), 5261–5267 (2009)
6.55 L. Schmalen, A. Alvarado, R. Rios-Müller: Performance prediction of nonbinary forward error correction in optical transmission experiments, J. Lightwave Technol. **35**(4), 1015–1027 (2017)
6.56 T. Koike-Akino, K. Sugihara, D.S. Millar, M. Pajovic, W. Matsumoto, A. Alvarado, R. Maher, D. Lavery, M. Paskov, K. Kojima, K. Parsons, B.C. Thomsen, S.J. Savory, P. Bayvel: Experimental demonstration of nonbinary LDPC convolutional codes for DP-64QAM/256QAM. In: *42nd Eur. Conf. Opt. Commun. (ECOC)* (2016)
6.57 H. Buelow, X. Lu, L. Schmalen, A. Klekamp, F. Buchali: Experimental performance of 4D optimized constellation alternatives for PM-8QAM and PM-16QAM. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.M2A.6
6.58 T.H. Lotz, X. Liu, S. Chandrasekhar, P.J. Winzer, H. Haunstein, S. Randel, S. Corteselli, B. Zhu, D.W. Peckham: Coded PDM-OFDM transmission with shaped 256-iterative-polar-modulation achieving 11.15-b/s/Hz intrachannel spectral efficiency and 800-km reach, J. Lightwave Technol. **31**(4), 538–545 (2013)
6.59 D.S. Millar, T. Fehenberger, T. Koike-Akino, K. Kojima, K. Parsons: Coded modulation for next-generation optical communications. In: *Opt. Fiber Commun. Conf.* (2018), https://doi.org/10.1364/OFC.2018.Tu3C.3
6.60 B. Chen, C. Okonkwo, H. Hafermann, A. Alvarado: Increasing achievable information rates via geometric shaping, arXiv:1804.08850 [cs.IT] (2018)
6.61 P. Schulte, G. Böcherer: Constant composition distribution matching, IEEE Trans. Inf. Theor. **62**(1), 430–434 (2016)
6.62 F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, W. Idler: Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration, J. Lightwave Technol. **34**(7), 1599–1609 (2016)
6.63 J. Cho, S. Chandrasekhar, R. Dar, P.J. Winzer: Low-complexity shaping for enhanced nonlinearity tolerance. In: *42nd Eur. Conf. Opt. Commun. (ECOC)* (2016)
6.64 T. Fehenberger, D.S. Millar, T. Koike-Akino, K. Kojima, K. Parsons: Multiset-partition distribution matching, arXiv:1801.08445 [eess.SP] (2018)

6.65 T. Yoshida, M. Karlsson, E. Agrell: Short-block-length shaping by simple mark ratio controllers for granular and wide-range spectral efficiencies. In: *Eur. Conf. Opt. Commun. (ECOC)* (2017), https://doi.org/10.1109/ECOC.2017.8346146

6.66 D. Cardenas, D. Lavery, P. Watts, S.J. Savory: Reducing the power consumption of the CMA equalizer update for a digital coherent receiver. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.Th4D.5

6.67 M.S. Faruk, D. Lavery, R. Maher, S.J. Savory: A low complexity hybrid time-frequency domain adaptive equalizer for coherent optical receivers. In: *Opt. Fiber Commun. Conf.* (2016), https://doi.org/10.1364/OFC.2016.Th2A.39

6.68 K.P. Zhong, J.H. Ke, Y. Gao, J.C. Cartledge, A.P.T. Lau, C. Lu: Carrier phase estimation for DP-16QAM using QPSK partitioning and quasi-multiplier-free algorithms. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.W4K.2

6.69 J. Krause Perin, A. Shastri, J.M. Kahn: Design of low-power DSP-free coherent receivers for data center links, J. Lightwave Technol. **35**(21), 4650–4662 (2017)

6.70 T. Koike-Akino, K. Kojima, D.S. Millar, K. Parsons, T. Yoshida, T. Sugihara: Pareto optimization of adaptive modulation and coding set in nonlinear fiber-optic systems, J. Lightwave Technol. **35**(4), 1041–1049 (2017)

6.71 C. Fougstedt, K. Szczerba, P. Larsson-Edefors: Low-power low-latency BCH decoders for energy-efficient optical interconnects, J. Lightwave Technol. **35**(23), 5201–5207 (2017)

6.72 K. Roberts, Q. Zhuge, I. Monga, S. Gareau, C. Laperle: Beyond 100 Gb/s: Capacity, flexibility, and network optimization, J. Opt. Commun. Netw. **9**(4), C12–C24 (2017)

6.73 R. Dar, P.J. Winzer: On the limits of digital back-propagation in fully loaded WDM systems, IEEE Photonics Technol. Lett. **28**(11), 1253–1256 (2016)

6.74 M.J. Ablowitz, J.F. Ladik: A nonlinear difference scheme and inverse scattering, Stud. Appl. Math. **55**(3), 213–229 (1976)

6.75 M.I. Yousefi, F.R. Kschischang: Information transmission using the nonlinear Fourier transform, Part I: Mathematical tools, IEEE Trans. Inf. Theor. **60**(7), 4312–4328 (2014)

6.76 J.D. Gibbon, P.J. Caudrey, R.K. Bullough, J.C. Eilbeck: An N-soliton solution of a nonlinear optics equation derived by a general inverse method, Lett. Nuovo Cim. **8**(13), 775–779 (1973)

6.77 M.I. Yousefi, X. Yangzhang: Linear and nonlinear frequency-division multiplexing. In: *42nd Eur. Conf. Opt. Commun.* (2016)

6.78 S. Wahls, H.V. Poor: Fast numerical nonlinear Fourier transforms, IEEE Trans. Inf. Theor. **61**(12), 6957–6974 (2015)

6.79 S.T. Le, J.E. Prilepsky, S.K. Turitsyn: Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers, Opt. Express **22**(22), 26720–26741 (2014)

6.80 V. Aref, H. Bülow, K. Schuh, W. Idler: Experimental demonstration of nonlinear frequency division multiplexed transmission. In: *Eur. Conf. Opt. Commun. (ECOC)* (2015), https://doi.org/10.1109/ECOC.2015.7341903

6.81 Z. Dong, S. Hari, T. Gui, K. Zhong, M.I. Yousefi, C. Lu, P.K.A. Wai, F.R. Kschischang, A.P.T. Lau: Nonlinear frequency division multiplexed transmissions based on NFT, IEEE Photonics Technol. Lett. **27**(15), 1621–1623 (2015)

6.82 S.K. Turitsyn, J.E. Prilepsky, S.T. Le, S. Wahls, L.L. Frumin, M. Kamalian, S.A. Derevyanko: Nonlinear Fourier transform for optical data processing and transmission: Advances and perspectives, Optica **4**(3), 307–322 (2017)

6.83 N. Eiselt, J. Wei, H. Griesser, A. Dochhan, M.H. Eiselt, J.-P. Elbers, J.J. Vegas Olmos, I. Tafur Monroy: Evaluation of real-time 8 x 56.25 Gb/s (400G) PAM-4 for inter-data center application over 80 km of SSMF at 1550 nm, J. Lightwave Technol. **35**(4), 955–962 (2017)

6.84 A. Mecozzi, C. Antonelli, M. Shtaif: Kramers–Kronig coherent receiver, Optica **3**(11), 1220–1227 (2016)

6.85 D. Che, A. Li, X. Chen, Q. Hu, Y. Wang, W. Shieh: Stokes vector direct detection for linear complex optical channels, J. Lightwave Technol. **33**(3), 678–684 (2015)

Seb J. Savory

Dept. of Engineering
University of Cambridge
Cambridge, UK
sjs1001@cam.ac.uk

Seb J. Savory received his PhD in Engineering from the University of Cambridge and then worked for Nortel in Harlow. He moved from Nortel to UCL in 2005, departing in 2015 to take up a position at the University of Cambridge where he is currently a Professor. He is a Chartered Engineer and Fellow of the IEEE, IET, OSA, and HEA.

David S. Millar

Mitsubishi Electric Research Laboratories
Cambridge, MA, USA
millar@merl.com

David S. Millar received his PhD from University College London (UCL) in 2011. Since 2012, he has been a researcher at Mitsubishi Electric Research Laboratories (MERL) in Cambridge, MA, USA. His research is primarily focused on signal processing algorithms and coded modulation for coherent optical communications.

7. Forward Error Correction for Optical Transponders

Alexandre Graell i Amat, Laurent Schmalen

Forward error correction (FEC) is an essential technique required in almost all communication systems to guarantee reliable data transmission close to the theoretical limits. In this chapter, we discuss the state-of-the-art FEC schemes for fiber-optic communications. Following a historical overview of the evolution of FEC schemes, we first introduce the fundamental theoretical limits of common communication channel models and show how to compute them. These limits provide the reader with guidelines for comparing different FEC codes under various assumptions. We then provide a brief introduction to the general basic concepts of FEC, followed by an in-depth introduction to the main classes of codes for soft-decision decoding and hard-decision decoding. We include a wide range of performance curves, compare the different schemes, and give the reader guidelines on which FEC scheme to use. We also introduce the main techniques to combine coding and higher-order modulation (coded modulation), including constellation shaping. Finally, we include a guide on how to evaluate the performance of FEC schemes in transmission experiments. We conclude the chapter with an overview of the properties of some state-of-the-art FEC schemes used in optical communications, and an outlook.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_7

The fundamental problem of communication is that of sending information from one point to another (either in space or time) efficiently and reliably. Examples of communication are the data transmission between a ground station and a space probe through the atmosphere and free space, the write and read of information on flash memory, and communication between servers in a data center through optical interconnects. The transmission over the physical medium, e.g., the optical fiber, is prone to noise due to amplification, distortion, and other impairments. As a consequence, the transmitted (digital) message may be received with errors.

To reduce the probability of error, one could (if possible) increase the signal-to-noise ratio (SNR), i.e., transmit at a higher power, or transmit using a larger bandwidth. In his landmark 1948 paper [7.1], *Claude E. Shannon* showed that there is a third parameter one can play with: the system complexity. Shannon proved that, for any given channel (e.g., a given SNR and bandwidth), there is a fundamental limit, the so-called *channel capacity*, at which information can be transmitted reliably, i.e., with diminishing error probability. Aside from determining the channel capacity, he also showed how this limit can be achieved: by the use of coding.

Shannon's contribution marks the birth of the fields of information theory and coding theory. While information theory aims at determining the fundamental limits of communication, coding theory aims at finding *practical* codes that achieve (or approach) these limits. The principle of coding is very simple: to introduce redundancy in the transmitted sequence in a controlled manner, such that it can be exploited by the receiver to correct the errors introduced by the channel. However, while Shannon proved the *existence* of capacity-achieving codes, he did not provide any insight on how to construct *practical codes*. His proof was based on a random-coding argument. Informally, the argument says that a random code will achieve capacity with high probability. Unfortunately, random codes are not practical, as both encoding and decoding require computational complexity and storage that grow exponentially with the block length (and Shannon showed that to achieve capacity, large block lengths—in truth infinitely large—are required!). Therefore, since the late 1940s, finding practical codes that are able to perform close to capacity with reasonable decoding complexity has been the holy grail of coding theory. For decades, researchers worked to develop powerful codes with a rich algebraic structure, among them the celebrated Bose–Chaudhuri–Hocquenghem (BCH) codes and Reed–Solomon codes, which unfortunately still performed far from the channel capacity with feasible decoding algorithms. Two breakthroughs came in the 1990s, the first when two researchers, *Claude Berrou* and *Alain Glavieux*, introduced turbo codes, the first class of codes that was shown to approach capacity with reasonable decoding complexity, and the second with the rediscovery of low-density parity-check (LDPC) codes (originally introduced in 1960 by *Robert Gallager*) by *David MacKay*. The advent of turbo codes and the rediscovery of LDPC codes constituted a cornerstone in coding theory and gave birth to what nowadays is referred to as the field of *modern coding theory*, which is also intimately related to the iterative decoding of these codes. LDPC codes and turbo codes have now been adopted in a plethora of communication standards.

The application of coding goes far beyond the classical communication and storage problem. Coding is a very powerful, fundamental tool that plays a key role in many other applications and research fields, including distributed storage and caching [7.2, 3], uncoordinated multiple access [7.4], to speed up distributed computing tasks [7.5, 6], for quantum key distribution [7.7], and post-quantum cryptography [7.8].

Coding theory is a vast field that is impossible to cover in a single book chapter. This is certainly not our aspiration. The aim of this chapter is simply to provide a brief and accessible introduction to the main concepts underlying the art of coding and a simple but rather complete overview of the main coding schemes that will arguably be used in future optical communications systems. It is intended to be an entrance door to the exciting and important realm of coding for researchers in the field of optical communications that, while perhaps not ambitioning to become experts on coding, would like to become familiar with the field or are simply interested in widening their knowledge. It is also a handy source intended for engineers and practitioners of optical communication systems who are faced with the choice of a coding scheme.

One of the purposes of this chapter is to review the principal coding schemes for optical communications.

Because of the limited space, while we briefly describe classical codes such as BCH and Reed–Solomon codes, the main focus of the chapter is on *modern codes*, in particular LDPC codes and product-like codes such as staircase codes. Modern codes are also commonly referred to as *graph-based* codes or *codes on graphs*, since they can be conveniently described by means of a graph. LDPC and product-like codes are the two main code classes currently being used in fiber-optic communications with soft-decision and hard-decision decoding, and will likely continue being adopted in future systems. We will briefly describe the main building blocks of these codes, which are *rooted* in classical coding theory.

The remainder of the chapter is organized as follows. The chapter starts with a brief overview of the history of coding, with a special focus on the field of optical communications. In Sect. 7.2, an introduction to the fundamental performance limits, to which the performance of practical coding schemes can be compared, is provided. We then introduce the basic definitions and concepts of coding in Sect. 7.3. Sections 7.4 and 7.5 are devoted to soft-decision forward error correction (FEC) and hard-decision FEC, respectively. In Sect. 7.6, the combination of coding and higher-order modulation, dubbed *coded modulation*, is discussed, and some basics of constellation shaping are given. In Sect. 7.7, we discuss how to evaluate the performance of FEC schemes in transmission experiments. Finally, in Sect. 7.8 we draw some conclusions and provide an outlook on open research problems.

Notation: The following notation is used throughout the chapter. We use boldface letters to denote vectors and matrices, e.g., $\boldsymbol{x}$ and $\mathbf{X}$, respectively, and calligraphic letters to denote sets, e.g., $\mathcal{X}$. The cardinality of a set $\mathcal{X}$ is given by $|\mathcal{X}|$. The real and imaginary parts of a complex number are denoted by $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$, respectively. The probability mass function (PMF) and the probability density function (PDF) of a random variable (RV) X are denoted by $P_X(x)$ and $p_X(x)$, respectively, as shorthand for $P_X(X=x)$ and $p_X(X=x)$. Sometimes we will simply write P_X and p_X. Also, we write $p_{XY}(x,y)$ to denote the joint PDF of two RVs X and Y, and $p_{Y|X}(y|x)$ to denote the conditional PDF of Y conditioned on $X=x$. We use similar notation for the joint and conditional PMFs. Probability is denoted by $\Pr(\cdot)$ and expectation by $\mathbb{E}$. More precisely, the expectation of a function $g(X)$ with respect to p_X is denoted as $\mathbb{E}_{p_X}[g(x)]$, or, in short, $\mathbb{E}_X[g(x)]$, i.e., $\mathbb{E}_X[g(x)] = \int g(x)p_X(x)\mathrm{d}x$. Likewise, the expectation of a function $g(x)$ with respect to a (discrete) PMF X is given by $\mathbb{E}_{P_X}[g(x)] = \sum_x g(x)P_X(x)$. Vector and matrix transpose are denoted by $(\cdot)^\top$, and $\|x\|$ denotes the ℓ_2-norm of x; $\mathbb{1}_{\{s\}}$ denotes the indicator function returning 1 if the statement s is true and 0 otherwise.

7.1 History of Forward Error Correction in Optical Communications

Forward error correction (FEC) in optical communications was first demonstrated in 1988 [7.9]. Since then, it has evolved significantly. This pertains not only to the techniques themselves but also to the encoder and decoder architectures. The history of FEC is, however, much older and has found more widespread application than only optical communications. In this section, we give a historical overview of FEC and link the historical developments to the developments in fiber-optic communications.

The first error-correcting codes, introduced by *Richard Hamming* in 1950 [7.10], were binary linear block codes that were designed to improve the reliability of electromechanical computing machines [7.11, 12]. Hamming codes are based on simple algebraic descriptions and have a straightforward hard-decision decoding (HDD) algorithm to correct single errors. Hamming codes were one of the first codes applied in optical communications. In 1988, *Wayne Grover* [7.9] reported a performance improvement not only with respect to the required received power and the observed residual error floor, but also with respect to chromatic dispersion tolerance and resilience to laser mode hopping.

Following the introduction of Hamming codes and the emergence of the field of coding theory, many new classes of codes were introduced, some with important properties and applications. One very powerful early code was the binary triple-error-correcting Golay code of length 23 bit and its ternary double-error-correcting counterpart (length 11 ternary digits) that were constructed by the Swiss engineer *Marcel Golay* in 1949 [7.13]. We should note that, while Hamming only published his work in 1950 after Golay's code, his work was already known to Shannon in 1948 and highlighted in [7.1], and hence several authors consider Hamming codes to be the first error-correcting codes. It is surprising that the ternary Golay code was in fact first discovered by the Finnish football fan *Juhani Virtakallio*, who published it in the 1947 issue of the Finnish football magazine *Veikkaaja* as a betting system [7.14]. The paper [7.14] gives an excellent overview of the early days of coding theory and its connections with neighboring scientific fields. The Golay

code is an exceptional code, as it is the *only* binary *perfect code* that can correct more than a single error ([7.15, Chap. 6, Sect. 10]). This means that there is no other binary code of smaller size that can correct up to three errors with HDD. The Golay code has found widespread application, and it was famously used in the two 1977 Voyager space missions [7.11]. The introduction of Hamming codes and the Golay code is a landmark in coding theory. For this reason, together with Shannon, Hamming and Golay are considered the fathers of the mathematical discipline of coding theory [7.15].

The 1950s saw the development of many new coding schemes that could correct multiple errors, for example the important Reed–Muller codes, with a construction first presented by *David Muller* in 1954 [7.16], and the decoding algorithm given by *Irving Reed* in the same year [7.17]. Also in 1954, *Peter Elias* introduced product codes (PCs) [7.18], a very relevant class of codes and the first example of codes built from smaller component codes. PCs are adopted nowadays for data storage on magnetic media and flash memory [7.19, 20], in the WiMAX (Worldwide Interoperability for Microwave Access) wireless communication standard [7.21], and for transmission over fiber-optic links [7.22]. Their excellent performance is due to a decoding algorithm that iterates between the decoding of the component codes. This iterative decoding algorithm was not originally envisaged by Elias, but only a few decades later, which explains why PCs were not initially adopted after their invention. A year later, *Elias* himself introduced the concept of convolutional codes [7.23] as an alternative to the then dominant block codes. In contrast to block codes, convolutional codes contain memory, and the output of the decoder depends not only on the input at the current time instant, but also on previous inputs. Because of their excellent performance with relatively low decoding latency, they have found their way into numerous communication standards. They are used in satellite and spacecraft links [7.11], for cellular radio (e.g., Global System for Mobile Communications (GSM)) [7.24], wireless local area networks [7.25], and Ethernet cables [7.26]. A particularly popular class of convolutional codes is based on simple binary linear shift registers. In this case, the encoder is a simple shift register with only a few memory elements. The key enabler for the success of convolutional codes was the Viterbi algorithm, introduced in 1967 [7.27, 28], which provides an efficient means of implementing optimal maximum likelihood (ML) decoding. The Viterbi algorithm enabled—for the first time—low-complexity soft-decision decoding (SDD) exploiting the full output of the channel. This represented a big step forward, as we will later see in Sect. 7.4. Despite their widespread use in many communication standards, convolutional codes have not yet found widespread use in fiber-optic communications. This is mostly due to the fact that parallelization of the decoder is not trivial, and to the availability of stronger codes at the time coding was adopted in fiber-optic communications. However, some notable exceptions are an early study of on-off-keying (OOK) modulation with different types of error-correcting codes, including convolutional codes [7.29], a paper analyzing the combination of convolutional codes and bandwidth-efficient modulation formats for fiber-optic communications [7.30], and a paper describing the use of low-rate convolutional codes to protect some levels in multilevel coding schemes [7.31].

The 1950s ended with the invention of one of the most important coding schemes still in use today, BCH codes, which were independently introduced by two groups [7.32, 33]. BCH codes are a class of cyclic error-correcting codes defined over a finite field that are able to correct multiple bit errors using simple syndrome decoding based on efficient closed-form solutions for computing the error locations [7.15, 34, 35]. After Hamming codes, single-error-correcting BCH codes of length 65 kbit, and two-error- and three-error-correcting BCH codes [7.36] were the next codes adopted in fiber-optic communication systems. These codes have overhead of less than 1%, while allowing the post-FEC bit error rate (BER) to be met with a channel pre-FEC BER of approximately 10^{-6}.

Shortly after the introduction of BCH codes, *Reed* and *Gustave Solomon* [7.37] introduced nonbinary cyclic error-correcting codes that are able to correct and/or detect a predetermined number of erroneous nonbinary symbols, consisting of groups of bits. The corresponding algebraic decoding algorithm was proposed by *Elwin Berlekamp* [7.38] and *James Massey* [7.39]. A widely used code is the Reed–Solomon code RS(255, 239) code (the notation will be introduced in Sect. 7.3) consisting of 255 nonbinary symbols, with each symbol composed of 8 bits. This code can correct up to eight symbol errors (or a single burst error of up to 57 bits) with 6.7% overhead and achieves a net coding gain (NCG) of 6.2 dB in SNR at a post-FEC BER of 10^{-15}. This means that we can tolerate an effective net SNR, taking into account the extra bandwidth required for the coding overhead, that is four times lower than without coding. This Reed–Solomon code was the first to find widespread use in fiber-optic communication systems, starting from submarine transmission [7.40], to metro and optical access systems [7.41]. Because of its low encoder and decoder implementation complexity, this code is still in use for applications with low coding gain requirements

and strict constraints on transceiver power consumption. It is included in several standards and described in detail in the ITU-T G.975 standard [7.42]. This code is commonly referred to as *first-generation* coding for light-wave systems [7.43].

In the decades that followed, coding research focused mostly on the construction of new, more powerful codes from the now existing schemes (i.e., following the same principle proposed originally by Elias with his PCs!). One such construction was the concatenation of simple component codes (e.g., convolutional codes, BCH codes, and Reed–Solomon codes), leading to long and powerful codes capable of achieving significant coding gains with low decoding complexity. Concatenated codes were first introduced by *Dave Forney* in 1965 in the form of serially concatenated codes (SCCs), with a cascaded *inner* and *outer* code [7.12, 44]. At the receiver, both component codes are decoded separately, resulting in much lower complexity compared with optimal ML decoding of the entire concatenated code, at the cost of some performance degradation. Typically, the inner code is a convolutional code decoded using SDD, while the outer code is a block code, usually a Reed–Solomon code, decoded using HDD. Concatenated codes soon became a popular code construction for satellite communications. A simple concatenated coding scheme was used already in the 1971 Mariner Mars orbiter mission, but regular use of concatenated codes for deep space communications began with the Voyager program. A concatenated code with an inner convolutional code decoded using Viterbi decoding and an outer Reed–Solomon code was first used in the 1977 NASA space mission Voyager 2, which was launched with the goal of studying the outer planets of our solar system. Concatenated codes are also used in the digital television broadcast standard DVB-S [7.45].

The so-called *second-generation* FEC schemes for fiber-optic communications are based mainly on PCs and concatenated codes, where the component codes are simple algebraic codes such as Hamming, BCH, or Reed–Solomon codes, or convolutional codes. The decoding of concatenated block codes is usually performed iteratively by alternately decoding each of the component codes, improving the overall result after each iteration. Almost all 6.7% overhead codes specified in the ITU-T G.975.1 standard [7.46] are concatenated codes. For instance, the code defined in [7.46, Appendix I.2] is composed of a convolutional inner code and a Reed–Solomon outer code. Similarly, the code defined in [7.46, Appendix I.5] is composed of an inner Hamming code and an outer Reed–Solomon code. All concatenated codes specified in [7.46] are very long codes, with block lengths of around 500 kbit. These codes achieve very good performance, with low error floors, and have moderate implementation complexity. For example, the concatenated code specified in [7.46, Appendix I.9], with inner and outer BCH codes, has an NCG of 9.24 dB and provides an output BER of less than 10^{-16} at an input BER of about 4×10^{-3}.

Second-generation coding schemes continue to evolve today for low-complexity applications such as metro networks [7.47] or data center interconnects. Three very recent developments in second-generation HDD FEC codes are continuously interleaved codes [7.48], staircase codes [7.49], and braided BCH codes [7.50], which can be considered generalizations of Elias' PCs with an additional convolutional-like structure. The NCG of the staircase code in [7.49] amounts to 9.41 dB at a BER of 10^{-15}, which is only 0.56 dB from the capacity of the hard-decision channel [7.49].

With the advent of coherent transmission schemes and high-resolution analog-to-digital converters (ADCs), SDD became an attractive means of increasing the NCGs and hence the transmission reach. Although SDD is the natural way to decode a code exploiting the full knowledge of the transmission statistics, HDD was prevalent in early fiber-optic communication systems due to the lack of ADCs and the requirements for simple high-speed receivers that carry out a binary 0/1 decision using analog radio-frequency (RF) circuits.

A breakthrough in the history of coding occurred in 1993, when Berrou and Glavieux introduced turbo codes, together with a simple iterative SDD algorithm, the *turbo principle* (due to its similarities to a turbo engine). Turbo codes are built from the parallel concatenation of two convolutional codes and were the first codes to approach capacity with acceptable decoding complexity. After the invention of turbo codes, a sudden interest in iteratively decodable codes arose in the research community. This led to the rediscovery of LDPC codes by *MacKay* [7.51, 52]. LDPC codes were originally introduced in 1960 by *Gallager* in his landmark PhD thesis [7.53, 54]. In his thesis, Gallager also proposed algorithms for both HDD and SDD. However, LDPC codes were not further investigated for many years because of the perceived higher decoding complexity for long codes. In the years that followed the invention of turbo codes and the rediscovery of LDPC codes, numerous publications from various researchers paved the way for a thorough understanding of these classes of codes, leading to numerous applications in various communication standards, such as WLAN (IEEE 802.11) [7.55], DVB-S2 [7.56], and 10G Ethernet (IEEE 802.3) [7.57].

In the realm of optical communications, FEC schemes with (iterative) SDD are typically classified as

third-generation FEC schemes. Today, there are largely two competing classes of codes that allow for SDD and that are attractive for implementation in optical receivers at decoding throughputs of 100 Gbit/s and above. The first class corresponds to the natural extension of the second-generation decoding schemes, i.e., PCs with SDD. Iterative low-complexity SDD of PCs was pioneered by *Ramesh Pyndiah* in 1998 [7.58], following the invention of turbo codes. Because of their similarities with turbo codes, PCs decoded using SDD are also often denoted as turbo product codes (TPCs) or block turbo codes (BTCs). The error floor of these codes tends to be low, with a steep BER slope. However, they require large block lengths to realize a small overhead, leading to a larger decoding latency. A further disadvantage is that flexibility with respect to varying frame sizes and overheads is more difficult to realize. Additionally, the decoder implementation requires the tweaking of many parameters to achieve the best possible performance. Furthermore, with overheads of more than 15% to 20%, these codes no longer perform well. The application of BTCs with iterative SDD in the context of fiber-optic communications was demonstrated in [7.59]. The second popular class of soft-decision decodable codes in optical communication systems is LDPC codes. One instance of an LDPC code has already been standardized in ITU-T G.975.1 ([7.46, Appendix I.6]). However, it was designed mostly for HDD. LDPC codes for SDD in optical communications were originally studied in [7.60], with work proposing complete coding schemes and frame structures shortly thereafter [7.61]. The popularity of LDPC codes is chiefly due to the existence of a suboptimal, conceptually very simple, low-complexity decoder for SDD: the belief propagation (BP) decoder. In addition to binary LDPC codes, nonbinary LDPC codes were promoted for optical communication systems in [7.62]. In short, nonbinary codes operate on symbols (e.g., modulation symbols), rather than bits. They come, however, at a cost of higher decoding complexity.

LDPC codes often suffer from an *error floor* at BERs of 10^{-8}–10^{-10} if no extra preventive measures are taken. Above a certain SNR, the BER no longer drops rapidly, but follows a curve with a smaller slope [7.63]. This effect is due to some structures in the code that lead to error patterns that cannot be recovered with BP decoding. Some modern high-performance FEC systems are therefore often constructed using an LDPC inner code decoded using SDD, which reduces the BER to a level of 10^{-3}–10^{-5}, and an outer block code (typically a BCH code or a Reed–Solomon code) decoded using HDD, which pushes the system BER to levels below 10^{-15} [7.61]. The implementation of a coding system with an outer cleanup code requires a thorough understanding of the LDPC code, the decoding algorithm, and its error mechanisms, and a properly designed interleaver between the LDPC code and the outer code so that errors at the output of the LDPC decoder—which typically occur in bursts—do not cause uncorrectable blocks after outer decoding. With increased computing resources, it is now also feasible to evaluate very low target BERs of LDPC codes and design LDPC codes that have very low error floors below the maximum acceptable BER [7.64].

Over the past decade, the field of coding theory has seen two important milestones: polar coding and spatial coupling. In 2008, *Erdal Arıkan* introduced the concept of channel polarization [7.65, 66], which led to the new concept of *polar codes*. The design of polar codes is based entirely on information-theoretic principles and differs radically from previously known code constructions. Despite the beauty of their construction and the fact that they are capacity-achieving with a simple *successive cancellation decoder*, which successively decodes the single bits of the codeword, polar codes are not immediately suited for practical application in optical communications. While polar codes show competitive performance for short block lengths with an improved *list decoder* [7.67] (which is one of the reasons they were chosen for transmitting control messages in the 5G radio standard [7.68]), for long block lengths, frequently used in optical communications to achieve high coding gains, their performance is not yet comparable to that of state-of-the-art LDPC codes. Furthermore, the decoder is inherently sequential, complicating their application in high-speed communication systems requiring high levels of parallelization. These drawbacks are currently topics of active research [7.69, 70].

The second breakthrough in the past decade was the introduction of the concept of *spatial coupling*. Spatial coupling itself has its own history. In 1999, *Alberto Jiménez Feltström* and *Kamil Zigangirov* investigated LDPC convolutional codes [7.71], a generalization of LDPC codes with a superimposed convolutional structure. It was only later fully realized that these codes, now known as *spatially coupled* LDPC (SC-LDPC) codes, possess outstanding performance. In fact, *Michael Lentmaier* and others realized that the asymptotic decoding performance of a certain class of SC-LDPC codes approaches the performance of optimal decoding [7.72–74] with simple suboptimal BP decoding. This effect, dubbed *threshold saturation*, was later rigorously proven in [7.75, 76]. The main advantage of spatially coupled codes is that it is possible to construct very powerful codes that can be decoded with a simple windowed decoder [7.77]. Spatial coupling is a very general concept that applies to other

classes of codes, such as turbo-like codes [7.78], as well as to other scenarios, including lossy compression [7.79], code-division multiple access [7.80], and compressed sensing [7.81]. Some product-like codes such as staircase codes and braided codes can also be seen as instances of spatially coupled codes.

Spatially coupled codes are now emerging in various forms. Two examples are the staircase code and the braided BCH code presented in [7.49] and [7.50], respectively, which are both codes with 6.7% overhead targeted for low-complexity applications with HDD (Sect. 7.5.3). Both codes can be classified as spatially coupled BCH PCs, and hence enable a natural windowed decoder implementation. Spatially coupled codes and polar codes are also being considered for SDD in optical communications [7.82–84]. However, as of today, the commercial application of polar codes in high-speed fiber-optic communications still seems further away than that of spatially coupled codes. The latter are a very general and broad concept that allows one to update many existing schemes and to reuse existing decoder hardware, at least to some extent. Therefore, we believe that the commercial application of spatially coupled codes is viable as of today.

The evaluation of the different generations of coding schemes for fiber-optic communications is shown in [7.85, Fig. 1.7]. In summary, we can say that within the span of only two decades, FEC in fiber-optic communication systems has evolved from the use of very simple schemes toward the implementation of state-of-the-art coding schemes in conjunction with high-speed decoding engines that operate at low complexity. Today, modern high-speed optical communication systems require high-performing FEC systems featuring throughput of 100 Gbit/s or multiples thereof, low power consumption, coding gains close to the theoretical limits, and adaptation to the peculiarities of the optical channel [7.86]. The stringent complexity requirements have even driven the investigation of new coding schemes, such as staircase codes [7.49], which are probably the first class of codes proposed specifically for optical transmission.

7.2 A Glimpse at Fundamental Limits and Achievable Rates

Before moving into the basics of coding, in this section we review the fundamental limits against which the performance of the coding schemes discussed in this chapter can be compared.

A simplified block diagram of a communication system is depicted in Fig. 7.1. The block of k source data bits $\boldsymbol{u} = (u_1, \ldots, u_k)$ is first encoded by an error-correcting code into a codeword $\boldsymbol{c} = (c_1, \ldots, c_n)$ consisting of n code bits, which is then modulated into a sequence of $\bar{n}$ constellation symbols $\boldsymbol{x} = (x_1, \ldots, x_{\bar{n}})$ that is transmitted over the channel. The received sequence $\boldsymbol{y}$ is demodulated and fed to the decoder, which aims at recovering the source data.

If the channel is memoryless, $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_i p_{Y|X}(y_i|x_i)$. In other words, the channel is completely specified by the conditional probability $p_{Y|X}(y|x)$ of receiving a single symbol y when a single symbol x is transmitted. The probability $p_{Y|X}(y|x)$ is often referred to as the *channel law*. Because of its prevalence and widespread use in modeling, throughout the chapter we will focus mainly on the case of memoryless channels with additive white Gaussian noise (AWGN), for which the channel law per real dimension (i.e., in each I/Q dimension in the case of complex constellation symbols) is expressed as

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(y-x)^2}{2\sigma^2}} , \tag{7.1}$$

where σ^2 is the noise variance. We note that when using complex constellations, we often assume that the noise of the real and imaginary parts are independent. In this case, the circularly symmetric complex channel law is expressed as

$$\begin{aligned} p_{Y|X}(y|x) &= p_{Y_I|X_I}(\mathrm{Re}\{y\}|\,\mathrm{Re}\{x\}) p_{Y_Q|X_Q}(\mathrm{Im}\{y\}|\,\mathrm{Im}\{x\}) \\ &= \frac{1}{\pi N_0} \mathrm{e}^{-\frac{\|y-x\|^2}{N_0}} \end{aligned}$$

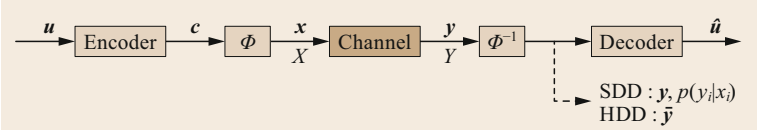


Fig. 7.1 Simplified block diagram of a communication system. For SDD, the decoder estimates the transmitted codeword based on $\boldsymbol{y}$ (equivalently the transition probabilities $p(y_i|x_i)$), while for HDD the decoder estimates the transmitted codeword based on the hard-detected sequence $\bar{\boldsymbol{y}}$

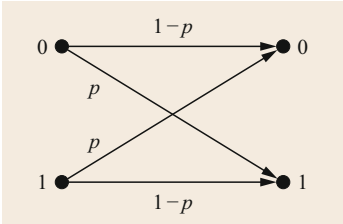


Fig. 7.2 The BSC. A bit is received correctly with probability $1-p$ and flipped with probability p

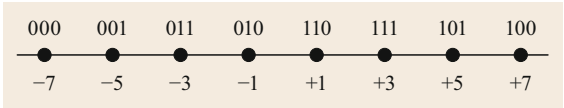


Fig. 7.3 8-PAM constellation with binary reflected Gray code labeling. The channel input X takes values on a set of 8 different symbols

The SNR is defined as $\mathrm{SNR} = E_\mathrm{s}/N_0$, where E_s is the average energy per symbol and $N_0/2 = \sigma^2$ the double-sided noise power spectral density.

We will also consider the binary symmetric channel (BSC), which is an important channel model for HDD. In fact, for an AWGN channel and binary phase-shift keying (BPSK) modulation, if hard detection is performed at the output of the channel, the resulting discrete memoryless channel between the encoder and the decoder can be modeled as a BSC with bits at the input and at the output, where a bit is flipped with probability

$$p = Q\left(\sqrt{\frac{2E_\mathrm{s}}{N_0}}\right) \tag{7.2}$$

and is received correctly with probability $1-p$. In (7.2), $Q(x) = (1/\sqrt{2\pi})\int_x^\infty \exp\left(-\tau^2/2\right)\mathrm{d}\tau$ is the tail probability of the standard normal distribution. The BSC is depicted in Fig. 7.2.

Let X be the channel input that takes values on the alphabet $\mathcal{X} = \{\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_M\} \subset \mathbb{C}$, i.e., $\mathcal{X}$ is the constellation of cardinality M imposed by the modulation (e.g., $M = 16$ for a 16-QAM format) and Y the channel output. Note that X and Y are random variables (x and y are their realizations). We assume that the channel output Y is a continuous random variable. As an example, consider transmission using the eight-pulse-amplitude modulation (PAM) constellation depicted in Fig. 7.3. In this case, X is a random variable that takes values on the set $\mathcal{X} = \{-7, -5, -3, -1, +1, +3, +5, +7\}$ according to a given distribution P_X. For conventional constellations, all symbols are equiprobable, i.e., $P_X(\mathsf{X}_i) = 1/M$ for all X_i.

For a given distribution P_X of the channel input, the mutual information (MI) between the channel input X and channel output Y,

$$\begin{aligned} I(X;Y) &\triangleq \sum_{x\in\mathcal{X}} P_X(x) \int p_{Y|X}(y|x) \\ &\quad \times \log_2 \frac{p_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} p_{Y|X}(Y|x')P_X(x')}\mathrm{d}y \\ &= \mathbb{E}_{XY}\left[\log_2\left(\frac{p_{Y|X}(Y|X)}{\sum_{x'\in\mathcal{X}} p_{Y|X}(Y|x')P_X(x')}\right)\right], \end{aligned} \tag{7.3}$$

determines the upper limit on the achievable rate. Here, by rate we mean the transmission rate, denoted by R, in bits per symbol (or bits per channel use). Note that in (7.3), the expectation is with respect to the PDF $P_X p_{Y|X}$. The rate R is sometimes referred to as spectral efficiency in the literature. This implicitly assumes ideal pulse shaping with infinitely extended $\sin(x)/x$ pulses and transmission at the Nyquist rate. In practice, the spectral efficiency (given in (bit/s)/Hz) and the transmission rate (given in bit/symbol or bit/channel use) often differ by a *spectral utilization factor* that depends on the pulse shaping, pre-emphasis, and other system parameters.

A rate R is called achievable if there exists a sequence of coding schemes such that the probability of a decoding error can be made arbitrarily small in the limit of infinitely large block lengths n. In other words, it is possible to transmit with vanishing error probability if $R < I(X;Y)$. Conversely, a sequence of coding schemes with vanishing error probability must have $R \le I(X;Y)$, and transmitting at a rate $R > I(X;Y)$ will always result in a probability of error bounded above zero. The highest possible achievable rate is given by the channel capacity, obtained by maximizing $I(X;Y)$ over all possible input distributions,

$$C \triangleq \max_{P_X} I(X;Y)\,. \tag{7.4}$$

Achieving the rate of (7.3) is possible with an optimal decoder or a typical sequence decoder for the channel $p_{Y|X}(Y|X)$ (in the limit of infinitely large block lengths). Furthermore, for some channels, suboptimal decoders may also achieve (7.3). In general, a communication system imposes some constraints on the transmitter and on the receiver (e.g., detection method, decoding strategy such as SDD or HDD), and thus for a specific communication system, the MI in (7.3) is not necessarily achievable. A lower bound on the true MI can be obtained using the mismatched decoding framework [7.87, 88]. The key idea behind mismatched decoding is to design a receiver that is optimal for an auxiliary channel (a good approximation of the true channel) and then computing (7.3) by averaging with respect to the true channel but using in the argument of the logarithm the conditional distribution of the auxil-

iary channel, i.e.,

$$\tilde{I}(X;Y) \triangleq \mathbb{E}_{XY}\left[\log_2\left(\frac{q_{Y|X}(Y|X)}{\sum_{x'\in\mathcal{X}} q_{Y|X}(Y|x')P_X(x')}\right)\right], \tag{7.5}$$

where $q_{Y|X}(Y|X)$ is the conditional distribution of the auxiliary channel and the expectation is taken with respect to $P_X p_{Y|X}$. The rate $\tilde{I}(X;Y)$ is achievable by using the optimal decoder for the auxiliary channel $q_{Y|X}(Y|X)$ over the real channel and is a lower bound of the MI, i.e., $I(X;Y) \geq \tilde{I}(X;Y)$.

Note that, in practice, the channel may not be memoryless. In this case, X and Y should be interpreted as vectors, $\boldsymbol{X}$ and $\boldsymbol{Y}$, and the conditional distribution $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})$ should be considered. The fiber-optic channel, indeed, has memory. In this section, to compute achievable rates, we make the common assumption that the memory of the fiber-optic channel can be neglected. In practice, most common receivers neglect the memory of the channel or remove it by sufficient interleaving; hence the memoryless assumption yields meaningful achievable rates for actual receivers.

To achieve high spectral efficiencies and performance close to the theoretical limits, fiber-optic communications employ coding in combination with higher-order constellations, a scheme known as coded modulation (CM). At the receiver side, both SDD and HDD may be used. For SDD, the decoder estimates $\boldsymbol{c}$ based on the full observation $\boldsymbol{y}$, i.e., $\boldsymbol{y}$ (or, equivalently, the transition probabilities $p(y_i|x_i)$) is fed to the decoder. For HDD, the demodulator takes hard decisions at the channel output, and the sequence of hard-detected symbols, denoted by $\bar{\boldsymbol{y}}$, is fed to the decoder (Fig. 7.1). Furthermore, binary codes and nonbinary codes may be employed, which lead to two decoding strategies, namely bit-wise decoding and symbol-wise decoding, respectively. In the following, we briefly discuss achievable rates for SDD and HDD with both bit-wise and symbol-wise decoding. We restrict ourselves to the conventional uniform input distribution, i.e., $p_X(x) = 1/M$ for all $x \in \mathcal{X}$. A brief discussion of CM with probabilistic shaping is provided in Sect. 7.6.

For further reading, achievable rates for fiber-optic communications assuming a memoryless channel have been addressed in [7.89, 90]. Achievable rates for probabilistic shaping are discussed in [7.91]. The capacity of the nonlinear optical channel with finite memory is addressed in [7.92, 93], where the joint effect of nonlinearity and dispersion was modeled as a finite-state machine and the capacity was estimated. Capacity lower bounds for a variety of scenarios were obtained in [7.94]. Achievable rates for a nonlinear wavelength-division-multiplexed (WDM) fiber-optic system using the mismatched decoding framework were obtained in [7.95].

7.2.1 Achievable Rates for Soft-Decision Decoding

In the following, we review some achievable rates for SDD. We consider both symbol-wise and bit-wise decoding, suitable for nonbinary codes and binary codes, respectively, which lead to different achievable rates.

Symbol-Wise Soft-Decision Decoding

Consider first a CM scheme that employs a nonbinary code to encode the data and symbol-wise SDD (SW-SDD). Using mismatched decoding, an achievable rate is given by

$$R_{\text{SW-SDD}} = \sup_{s\geq 0} \mathbb{E}_{XY}\left[\log_2\left(\frac{q_{Y|X}(Y|X)^s}{\frac{1}{M}\sum_{x\in\mathcal{X}} q_{Y|X}(Y|x)^s}\right)\right], \tag{7.6}$$

where sup stands for supremum (which can be interpreted as a maximization), s is an optimization parameter, and the expectation is taken with respect to $P_X(x)p_{Y|X}(y|x)$, where $P_X(x) = 1/M$ assuming equiprobable symbols. To obtain the results in Sect. 7.2.3 below, we will assume that the auxiliary channel $q_{Y|X}(Y|X)$ is an AWGN channel, which is a good approximation of the fiber-optic channel under certain conditions. Note that for transmission over the AWGN channel, (7.6) boils down to (7.3), i.e., for transmission over the AWGN channel, the MI is indeed an achievable rate using nonbinary coding and SW-SDD.

Bit-Wise Soft-Decision Decoding

In fiber-optic communications, CM is typically implemented according to the bit-interleaved coded modulation (BICM) paradigm [7.96], where a binary code is used in combination with a higher-order constellation, and the code bits are interleaved prior to the mapping to constellation symbols (BICM and other CM schemes are discussed in Sect. 7.6.3). Accordingly, bit-wise decoding is performed at the receiver. While the rate (7.6) is achievable by a symbol-wise soft-decision decoder, it is not necessarily achievable by a bit-wise decoder. In the following, we give an achievable rate for bit-wise decoding.

The BICM decoder treats the different bits of a given symbol as independent. Note that the bits are independent only in the case of perfect interleaving. However, the BICM decoder treats bits as though they were truly independent, which simplifies the decoding. We denote by $m = \log_2 M$ the number of bits

used to label each constellation symbol, and by $\mathsf{L}(x) = (b^{(1)}(x), \ldots, b^{(m)}(x))$ the m-bit labeling associated with symbol $x \in \mathcal{X}$. An example of 8-PAM with the binary reflected Gray code labeling is shown in Fig. 7.3. In the figure, $\mathcal{X} = \{-7, -5, -3, -1, +1, +3, +5, +7\}$ and $\mathsf{L}(+1) = (1, 1, 0)$. To simplify notation, we will sometimes write $\mathsf{L}(x) = (b_1, \ldots, b_m)$, hence omitting the argument x on the $b^{(i)}$'s. Also, we denote by $\mathcal{X}_b^{(i)}$, $i = 1, \ldots, m$, the subset of constellation points such that $b^{(i)} = b$, i.e., $\mathcal{X}_b^{(i)} \triangleq \{x \in \mathcal{X} : b^{(i)}(x) = b\}$. For the example in Fig. 7.3, $\mathcal{X}_0^{(2)} = \{-7, -5, +5, +7\}$ and $\mathcal{X}_1^{(2)} = \{-3, -1, +1, +3\}$.

Similar to symbol-wise decoding, we can compute an achievable rate using the mismatched decoding framework. In particular, an achievable rate for BW-SDD is given by the generalized mutual information (GMI),

$$R_{\text{BW-SDD}} = \text{GMI}_{\text{BW-SDD}} \triangleq \sup_{s>0} \mathbb{E}\left[\log_2\left(\frac{q(X,Y)^s}{\frac{1}{M}\sum_{x\in\mathcal{X}} q(x,Y)^s}\right)\right]. \tag{7.7}$$

Assuming that bit levels are independent (i.e., assuming that the random variables $B^{(i)}$, $i = 1, \ldots, m$, are independent), the GMI of the BICM mismatched decoder is equal to the sum of the GMI of the independent binary-input parallel channels [7.97, Th. 2],

$$R_{\text{BW-SDD}} = \sup_{s>0} \sum_{i=1}^{m} \mathbb{E}_{B^{(i)}Y}\left[\log_2\left(\frac{q_i(b^{(i)},Y)^s}{\frac{1}{2}\sum_{b'\in\{0,1\}} q_i(b',Y)^s}\right)\right], \tag{7.8}$$

where the expectation is over $P_{B^{(i)}}(b^{(i)})p_{Y|B^{(i)}}(y|b^{(i)})$, with $P_{B^{(i)}}(b^{(i)}) = 1/2$, and where the i-th bit decoding metric $q_i(b, y)$, $i = 1, \ldots, m$, is given by

$$q_i(b,y) \triangleq q_i(b^{(i)}(x) = b, y) = \sum_{x'\in\mathcal{X}_b^{(i)}} q_{Y|X}(y|x'), \tag{7.9}$$

with the auxiliary channel $q_{Y|X}(Y|X)$ being the AWGN channel.

7.2.2 Achievable Rates for Hard-Decision Decoding

HDD consists of two steps. First, hard decisions on the received symbols are performed, and then the decoder employs the Hamming distance metric to decode. By doing so, the discrete-input, continuous-output channel with discrete input $x \in \mathcal{X}$ and continuous output y is turned into an M-ary input, M-ary output discrete memoryless channel with input $x \in \mathcal{X}$ and output $\bar{y} \in \mathcal{X}$ (the hard-detected symbols) with transition probabilities $P_{\bar{Y}|X}(\bar{Y} = \bar{y}|X = x)$. We denote by $\boldsymbol{x} = (x_1, \ldots, x_n)$, $x_i \in \mathcal{X}$ the codeword of modulated symbols and by $\bar{\boldsymbol{y}} = (\bar{y}_1, \ldots \bar{y}_n)$, $\bar{y}_i \in \mathcal{X}$, the sequence at the output of the hard detector.

Symbol-Wise Hard-Decision Decoding

A possible strategy for computing an achievable rate is to resort to the mismatched decoding framework [7.87, 88]. With optimal Hamming-metric decoding, the decoding rule is

$$\hat{\boldsymbol{x}}_{\text{SW-HDD}} = \arg\min_{\boldsymbol{x}\in C} d_{\text{H}}(\boldsymbol{x}, \bar{\boldsymbol{y}}), \tag{7.10}$$

where $d_{\text{H}}(\boldsymbol{x}, \bar{\boldsymbol{y}})$ is the Hamming distance between vectors $\boldsymbol{x}$ and $\bar{\boldsymbol{y}}$, i.e., the number of positions in which the two vectors differ (see Definition 7.4 in Sect. 7.3 for a formal definition).

For symbol-wise decoding, assuming that the channel is memoryless, employing HDD based on the Hamming decoding metric is equivalent to maximizing the codeword mismatched decoding metric [7.90]

$$q(\boldsymbol{x}, \bar{\boldsymbol{y}}) = \prod_{i=1}^{n} q(x_i, \bar{y}_i) \tag{7.11}$$

with

$$q(x, \bar{y}) = \begin{cases} 1-\xi & \text{if } \bar{y} = x \\ \dfrac{\xi}{M-1} & \text{otherwise} \end{cases}, \tag{7.12}$$

where ξ is an arbitrary value in $(0, (M-1)/M)$. In fact (see [7.90] for details),

$$\begin{aligned} \hat{\boldsymbol{x}}_{\text{SW-HDD}} &= \arg\max_{\boldsymbol{x}\in C} \prod_{i=1}^{n} q(x_i, \bar{y}_i) \\ &= \arg\max_{\boldsymbol{x}\in C} \left(\frac{\xi}{(1-\xi)(M-1)}\right)^{d_{\text{H}}(\boldsymbol{x},\bar{\boldsymbol{y}})} \\ &\overset{\text{(a)}}{=} \arg\min_{\boldsymbol{x}\in C} d_{\text{H}}(\boldsymbol{x}, \bar{\boldsymbol{y}}), \end{aligned} \tag{7.13}$$

where (a) holds if and only if $0 < \xi < (M-1)/M$.

The metric in (7.12) corresponds to the optimal (i.e., ML) metric for an M-ary symmetric channel with symbols $x \in \mathcal{X}$ at its input and symbols $\bar{y} \in \mathcal{X}$ at its output and error probability ξ (i.e., the probability that symbol $x \in \mathcal{X}$ is received erroneously is $\Pr(\bar{y} \neq x|x) = \xi$). Implicitly, a Hamming distance metric decoder treats the channel $P_{\bar{Y}|X}$ as a symmetric channel, ignoring the

information provided by the actual channel transition probabilities.

An achievable rate for symbol-wise HDD is given by the GMI [7.87, 98]

$$R_{\text{SW-HDD}} = \text{GMI}_{\text{SW-HDD}} \triangleq \sup_{s>0} \mathbb{E}_{X\bar{Y}} \left[\log_2 \left(\frac{q(X,\bar{Y})^s}{\frac{1}{M}\sum_{x\in\mathcal{X}} q(x,\bar{Y})^s} \right) \right], \tag{7.14}$$

where the expectation is over $P_{X\bar{Y}}$, with $P_X(x) = 1/M$, and $q(x,\bar{y})$ is the (symbol) mismatched decoding metric. Using (7.12) as the mismatched metric in (7.14), after some simple derivations, the achievable rate for SW-HDD is obtained as [7.90]

$$R_{\text{SW-HDD}} = \log_2 M - h_{\text{b}}(\delta) - \delta \log_2(M-1), \tag{7.15}$$

where

$$h_{\text{b}}(\delta) = -\delta \log_2 \delta - (1-\delta)\log_2(1-\delta) \tag{7.16}$$

is the binary entropy function evaluated in δ, with δ being the pre-FEC channel symbol error probability, i.e., $\delta = \Pr(\bar{Y} \neq X)$.

Bit-Wise Hard-Decision Decoding

As before, let $\mathsf{L}(x) = (b^{(1)}, \ldots, b^{(m)})$ be the m-bit labeling associated with symbol $x \in \mathcal{X}$. Also, denote by $\mathsf{L}(\bar{y})$ the binary labeling associated with the hard-detected symbol $\bar{y}$. Assuming BICM, we can model the m-bit level channels as m parallel independent BSCs. Let ϵ_i be the bit error probability of the i-th channel, i.e., $\epsilon_i = \Pr(\hat{B}^{(i)} \neq B^{(i)})$. Analogous to SW-HDD, an achievable rate for BW-HDD is given by the GMI (7.14). For bit-wise decoding employing the Hamming distance metric decoding, the mismatched metric is

$$q(x,\bar{y}) = \xi^{d_{\text{H}}(\mathsf{L}(x),\mathsf{L}(\bar{y}))}, \tag{7.17}$$

where ξ is an arbitrary value in $(0,1)$, and $d_{\text{H}}(\mathsf{L}(x),\mathsf{L}(\bar{y}))$ is the Hamming distance between the binary vectors $\mathsf{L}(x)$ and $\mathsf{L}(\bar{y})$.

Using (7.17) in (7.14), after some simple derivations the achievable rate for BW-HDD is obtained as [7.90]

$$R_{\text{BW-HDD}} = m[1 - h_{\text{b}}(\bar{\epsilon})], \tag{7.18}$$

where $h_{\text{b}}(\bar{\epsilon})$ is the binary entropy function (7.16) evaluated in $\bar{\epsilon}$ and

$$\bar{\epsilon} = \frac{1}{m}\sum_{i=1}^{m} \epsilon_i . \tag{7.19}$$

We note that, rather than taking hard decisions over the received symbols, one may first compute bit-wise log-likelihood ratios (LLRs) and then take hard decisions at the bit level. This leads to an achievable rate different from the one in (7.18). However, the two achievable rates are very similar.

7.2.3 Comparison of Achievable Rates for Bit-Wise and Symbol-Wise Decoding

In Fig. 7.4, we plot the achievable rates (7.6), (7.8), (7.15), and (7.18) for transmission over the AWGN channel using 16-QAM and 64-QAM modulation as a function of the SNR. The achievable rate curves have to be interpreted as follows: for each SNR, they determine the maximum spectral efficiency that *can be achieved* (with vanishing probability of error for infinite code length and ideal pulse shaping) by a given decoding strategy (SW-SDD, BW-SDD, SW-HDD, or BW-HDD). Alternatively, for a given spectral efficiency, a given decoding strategy can yield vanishing error probability if operating at an SNR of at least the value provided by the curve. It is observed that for SDD, the achievable rates for bit-wise and symbol-wise decoding are similar. More precisely, the achievable rate for bit-wise decoding is slightly smaller, especially for low spectral efficiencies. However, bit-wise decoding entails virtually no loss in the region of interest (i.e., for spectral efficiencies where the curve starts bending toward its maximum

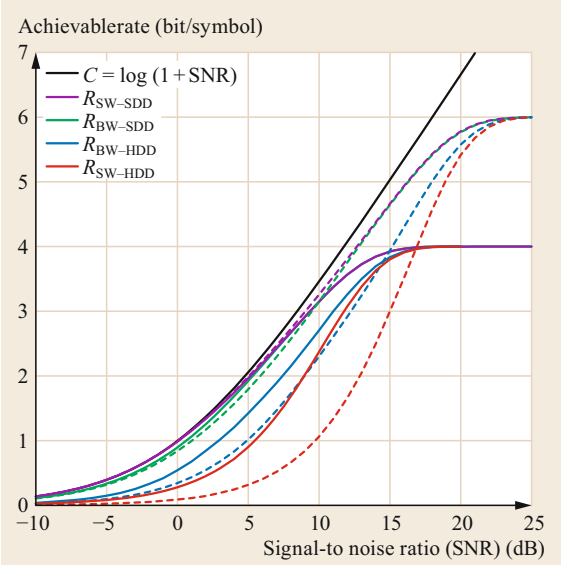


Fig. 7.4 Achievable rates for transmission over the AWGN channel with 16-QAM modulation (*solid curves*) and 64-QAM modulation (*dashed curves*) for SDD and HDD with both bit-wise and symbol-wise decoding

Table 7.1 Fiber and simulation parameters for the SSFM

Attenuation (α)	0.2 dB/km
Dispersion (D)	17 ps/nm/km
Nonlinear coefficient (γ)	1.3 $(\text{Wkm})^{-1}$
Wavelength λ	1550 nm
Symbol rate	32 Gbaud
Span length	80 km
EDFA noise figure	4.5 dB
SSFM step size	0.1 km

value m) with respect to symbol-wise decoding. On the other hand, as shown in [7.90], for HDD the achievable rates for bit-wise decoding (using BICM) are higher than those for symbol-wise decoding. As can be seen in the figure, the penalty incurred by using symbol-wise decoding yields a significantly lower achievable rate for low-to-medium SNRs. Equivalently, to achieve a given rate, the symbol-wise decoder requires a significantly higher SNR. According to these results, for HDD, binary FEC codes and bit-wise decoding are preferable to nonbinary FEC codes and symbol-wise decoding [7.90]. Furthermore, the decoding complexity of binary codes is lower than that of nonbinary codes. In the figure, we also plot the channel capacity. It is very important to realize that for high spectral efficiencies, there is a gap between the achievable rate $R_{\text{SW-SDD}}$, which coincides with the MI for the AWGN channel as explained in Sect. 7.2.1, *Symbol-Wise Soft-Decision Decoding*, and the capacity. This gap, referred to as the *shaping loss*, occurs because QAM constellations with equiprobable symbols are not capacity-achieving. The shaping loss induced by standard constellations is discussed in Sect. 7.6.5, together with methods to reduce it.

In Fig. 7.5, we plot the achievable rates as a function of the transmission distance for a polarization-multiplexed (PM) single-channel transmission system with electronic dispersion compensation (EDC) to compensate for the chromatic dispersion for 64-QAM and 256-QAM. Here, we neglect the polarization mode dispersion and the state of polarization (SOP) drift. Therefore, the two polarizations are independent of each other. The parameters of the optical fiber are summarized in Table 7.1. The results are for optimal launch power. The span loss is compensated for using erbium-doped fiber amplifiers (EDFAs), and a root-raised cosine pulse with a roll-off factor of 0.25 is used. The split-step Fourier method is used to simulate the fiber-optic channel and estimate $p_{Y|X}(y|x)$ for SDD and the transition probabilities $P_{\bar{Y}|X}(\bar{y}|x)$, the error probability $\bar{\epsilon}$ in (7.19), and the pre-FEC channel symbol error probability δ for HDD, and correspondingly calculate the achievable rates.

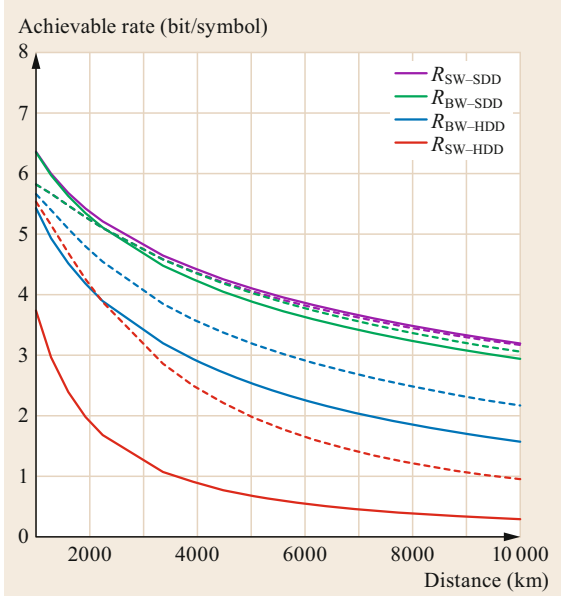


Fig. 7.5 Achievable rates for a PM single-channel transmission system with EDC as a function of the transmission distance for SDD and HDD with both bit-wise and symbol-wise decoding and optimal launch power. *Dashed curves* correspond to 64-QAM modulation and *solid curves* to 256-QAM modulation

In Fig. 7.6, we plot the achievable rates for a WDM transmission system with 81 channels and the parameters of Table 7.1. The optical channel for such a system is well approximated by the AWGN channel, a model known as the GN model in the literature [7.99, 100]. The achievable rates for the middle channel with 64-QAM and 256-QAM are shown in Fig. 7.6.

The results in Figs. 7.5 and 7.6 for the PM single-channel and WDM systems lead to similar conclusions as those of the results for the AWGN channel. We observe in Fig. 7.6 that for SDD, the bit-wise and symbol-wise decoding achievable rate curves for 16-QAM are indistinguishable, while for 64-QAM, bit-wise decoding incurs only a small penalty compared with symbol-wise decoding for 64-QAM for low spectral efficiencies. For 256-QAM, the achievable transmission reach for symbol-wise decoding is higher than that for bit-wise decoding. However, for the spectral efficiencies of interest, bit-wise SDD and symbol-wise SDD yield almost identical transmission reach. Similar results are observed in Fig. 7.5. In contrast, for HDD, bit-wise decoding yields larger achievable rates than symbol-wise decoding, which leads to a significant optical reach enhancement.

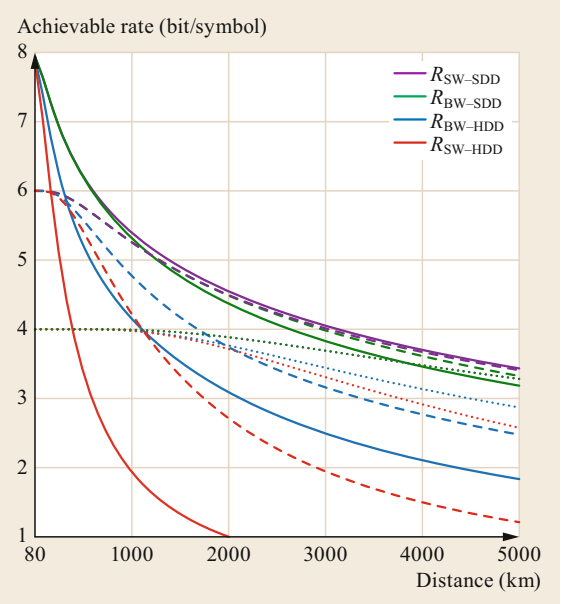


Fig. 7.6 Achievable rates for a WDM transmission system using the GN model as a function of the transmission distance for SDD and HDD with both bit-wise and symbol-wise decoding and optimal launch power. *Dotted curves* correspond to 16-QAM modulation, *dashed curves* to 64-QAM modulation, and *solid curves* to 256-QAM modulation

7.2.4 Achievable Rates and Actual Code Performance

It is important to stress that the achievable rates provided by information theory, as those discussed above, determine a rate that can be achieved, i.e., *there exists* a code that, with infinite block length, achieves that particular transmission rate. However, achievable rates tell little about the actual performance of particular FEC codes in general. A given code, decoded using a certain (suboptimal) decoding algorithm, does not necessarily achieve them. For example, regular LDPC codes (with growing degree) are capacity-achieving for binary transmission over the AWGN channel under maximum a posteriori (MAP) decoding, but not under practical iterative decoding. Similarly, staircase codes decoded with (suboptimal) iterative bounded distance decoding (Sect. 7.5.6) are not capacity-achieving. Therefore, in general, practical codes (in the sense of codes decoded using suboptimal decoding algorithms) do not achieve the achievable rates discussed above, and a performance gap will be observed. Furthermore, the gap with respect to the achievable rate may depend on the code rate (see (7.22) in Sect. 7.3 for the formal definition of code rate). For example, for staircase codes, the gap relative to the achievable rate increases for low rates. Nonetheless, achievable rates provide a very useful benchmark for comparing the performance of practical codes and are still informative.

Note that above, by capacity-achieving we mean a code that achieves the Shannon limit. For instance, for binary transmission using BPSK and SDD, the ultimate limit, also called the Shannon limit, is the corresponding MI. A code that achieves the MI for binary transmission over the AWGN channel does not achieve the capacity of the AWGN channel (since this requires Gaussian input). However, with some abuse of language, in coding theory jargon, such a code is called capacity-achieving.

As we will discuss in Sects. 7.4 and 7.5, for LDPC codes, generalized product codes, and modern codes in general, the performance limits (in the limit of infinite block length) for a certain code ensemble are given by density evolution. The so-called decoding threshold obtained from density evolution can be seen as the *effective capacity* (i.e., the effective performance limit) for a given code ensemble. One should then compare the threshold of a given code ensemble with the corresponding achievable rate to see how much is lost because of the choice of a particular code and decoding algorithm. Designing codes that approach the fundamental limits given by information theory is the ultimate goal of coding theory.

7.3 Basics of Forward Error Correction

Forward error correction (FEC), also known as *error-correcting coding*, *channel coding*, or simply *coding*, is an indispensable technique for achieving reliable communication and storage that has become ubiquitous in any communication and storage system. The role of FEC is to protect the data to be transmitted (or stored) from the channel impairments. It does so by following a basic principle: adding redundancy.

In this section, we introduce the basic concepts of FEC. Our focus is on binary codes, since their lower decoding complexity with respect to nonbinary codes makes them by far the most popular and widely used (except, perhaps, Reed–Solomon codes). However, the concepts introduced in this section can be extended to nonbinary codes in a straightforward manner.

We first give a layman's definition of an error-correcting code.

Definition 7.1 Error-correcting code
An error-correcting code is a device that adds redundancy to a block of data bits and uses this redundancy to correct potential transmission errors.

In this chapter, we focus on the most important and ubiquitous class of codes, *block codes*, and give the most basic definition of a block code.

Definition 7.2
A binary block code C of code length n and dimension k, denoted by $C(n,k)$, is a collection of 2^k binary vectors of length n bits, called n-vectors, i.e.,

$$C(n,k) \triangleq \left\{\boldsymbol{c}^{[1]}, \boldsymbol{c}^{[2]}, \ldots, \boldsymbol{c}^{[2^k]}\right\}, \quad \boldsymbol{c}^{[m]} \in \{0,1\}^n \,. \tag{7.20}$$

The n-vectors $\boldsymbol{c}^{[m]}$ are called *codewords*, and k is also referred to as the *information block length* of the code. The set of binary numbers is also often denoted as the *Galois field* of two elements and notated as $\mathrm{GF}(2) \triangleq \{0,1\}$. Note that the notion of Galois field also defines the operations of addition and multiplication of its elements. Loosely speaking, $\mathrm{GF}(q)$ is a finite set of q objects over which addition and multiplication are defined with similar properties as addition and multiplication of real numbers. For details and rigorous definitions, see [7.101, Chap. 2].

Example 7.1
Consider the 3-repetition code with parameters $n = 3$ and $k = 1$, consisting of $2^k = 2$ codewords of length 3 bits,

$$C_{\mathrm{rep}}(n=3, k=1) = \{(0,0,0), (1,1,1)\} \,. \tag{7.21}$$

One may consider an *encoder* that assigns to the information bit $u = 0$ the codeword $(0,0,0)$ and to the information bit $u = 1$ the codeword $(1,1,1)$. The assignment of codewords to information words (in this case to a single bit) is usually referred to as *encoding*.

An important parameter is the *code rate* R, or simply rate, defined as

$$R \triangleq \frac{k}{n} < 1 \,. \tag{7.22}$$

The rate is a measure of the redundancy of the code. Since $R < 1$, then $n > k$, i.e., the information words are encoded into longer sequences, thereby introducing redundancy. In contrast to the field of coding theory, in the optical communications literature the code rate is rarely used. Therein, the *code overhead*, or simply overhead, defined as

$$\mathrm{OH} \triangleq \frac{1}{R} - 1 = \frac{n-k}{k} \,, \tag{7.23}$$

is frequently used instead. The overhead denotes the relative portion of redundant bits that are added to the information bits. Note that the rate of the code is expressed in information bits per transmitted bit and is thus a measure of the *bandwidth efficiency* of the code. For example, the rate of the 3-repetition code of Example 7.1 is $R = 1/3$, which corresponds to overhead of $\mathrm{OH} = 200\%$. For fiber-optic communications, practical codes with much lower overhead are desired.

We now define one of the most important parameters of error-correcting codes, the minimum Hamming distance. We will first need the definition of Hamming weight and Hamming distance.

Definition 7.3 Hamming weight
For a binary vector $\boldsymbol{c} = (c_1, \ldots, c_n)$ of length n, the Hamming weight, denoted by $w_{\mathrm{H}}(\boldsymbol{c})$, is the number of entries in which $c_i = 1$, i.e.,

$$w_{\mathrm{H}}(\boldsymbol{c}) \triangleq |\{i \,:\, c_i = 1\}| \,. \tag{7.24}$$

Definition 7.4 Hamming distance
For any two binary vectors $\boldsymbol{c}$ and $\tilde{\boldsymbol{c}}$ of length n, the Hamming distance, $d_{\mathrm{H}}(\boldsymbol{c}, \tilde{\boldsymbol{c}})$, is the number of entries in which $\boldsymbol{c}$ and $\tilde{\boldsymbol{c}}$ differ, i.e.,

$$d_{\mathrm{H}}(\boldsymbol{c}, \tilde{\boldsymbol{c}}) \triangleq |\{i \,:\, c_i \neq \tilde{c}_i\}| \,. \tag{7.25}$$

Definition 7.5 Minimum Hamming distance
The minimum Hamming distance of a code C, denoted by $d_{\min}(C)$, is the smallest Hamming distance between any two distinct codewords of the code and is defined as

$$d_{\min}(C) \triangleq \min_{\substack{\boldsymbol{c}, \tilde{\boldsymbol{c}} \in C \\ \boldsymbol{c} \neq \tilde{\boldsymbol{c}}}} d_{\mathrm{H}}(\boldsymbol{c}, \tilde{\boldsymbol{c}}) \,. \tag{7.26}$$

As we will later see, the minimum Hamming distance of a code is related to its *error-correcting capability*. Since the notion of minimum Hamming distance is so important in coding, an (n,k) block code with minimum Hamming distance $d_{\min}$ is sometimes denoted as an $(n,k,d_{\min})$ block code.

The following simple but powerful upper bound on the minimum Hamming distance holds.

Theorem 7.1 Singleton bound
The minimum Hamming distance of an $(n, k, d_{\min})$ code C over GF(q) is upper bounded by $d_{\min} \leq n - k + 1$.

Codes whose minimum Hamming distance attains the Singleton bound achieve the best possible error-correcting capability for a given overhead and are referred to as maximum distance separable (MDS) codes. A very important class of MDS codes are Reed–Solomon codes, which are briefly discussed in Sect. 7.5.1. In contrast, there are no binary MDS codes, except trivial codes—the repetition code and the parity-check code.

7.3.1 Linear Block Codes

Linear block codes are a very important class of codes. Indeed, almost all practical codes in use today are linear codes. The reason is that linear codes are sufficient to achieve capacity [7.23, 102], and linearity allows for efficient encoding and decoding circuits. All codes considered in this chapter are linear codes.

Definition 7.6 Linear block code
A binary block code $C(n, k)$ is a linear block code if and only if its codewords $\boldsymbol{c}^{[1]}, \ldots, \boldsymbol{c}^{[2^k]}$ form a k-dimensional subspace of the n-dimensional vector space $\{0, 1\}^n$.

For a linear block code, the *all-zero* codeword $(0, 0, \ldots, 0)$ is always a codeword, i.e., $(0, 0, \ldots, 0) \in C$, and the (binary) addition of two codewords is also a codeword, i.e., for $\boldsymbol{c} \in C$ and $\tilde{\boldsymbol{c}} \in C$, then $\boldsymbol{c} + \tilde{\boldsymbol{c}} \in C$, with $+$ denoting the binary (modulo-2) addition. In other words, the set of codewords is closed under component-wise binary addition, where binary addition corresponds to the logical exclusive OR (XOR) operation, and binary multiplication corresponds to the logical AND operation. For any two binary vectors $\boldsymbol{c}$ and $\tilde{\boldsymbol{c}}$,

$$d_{\mathrm{H}}(\boldsymbol{c}, \tilde{\boldsymbol{c}}) = w_{\mathrm{H}}(\boldsymbol{c} + \tilde{\boldsymbol{c}}) \,, \tag{7.27}$$

and the minimum Hamming distance of a linear block code is equal to the minimum codeword weight, i.e.,

$$d_{\min}(C) = \min_{\substack{\boldsymbol{c} \in C \\ \boldsymbol{c} \neq \boldsymbol{0}}} w_{\mathrm{H}}(\boldsymbol{c}) \,. \tag{7.28}$$

Linear codes have appealing properties for analysis and implementation. It is easy to see that the 3-repetition code of Example 7.1 is a linear code. For this code, the three-dimensional subspace representation is shown in Fig. 7.7.

Since a linear block code is a k-dimensional subspace, we can find k linearly independent basis vectors $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k$ in $\{0, 1\}^n$ that span C, i.e., every codeword in C is a linear combination of these k basis vectors. Then, the codeword $\boldsymbol{c} = (c_1, \ldots, c_n)$ for the message $\boldsymbol{u} = (u_1, \ldots, u_k)$ is obtained as the product of $\boldsymbol{u}$ and $\mathbf{G}$ as

$$\boldsymbol{c} = \boldsymbol{u}\mathbf{G} \,, \tag{7.29}$$

where $\mathbf{G}$ is a $k \times n$ binary matrix with rows $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k$. Since $\mathbf{G}$ spans (i.e., generates) the code C, it is usually referred to as the *generator matrix* of the code. Note that the basis vectors $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k$ are also codewords.

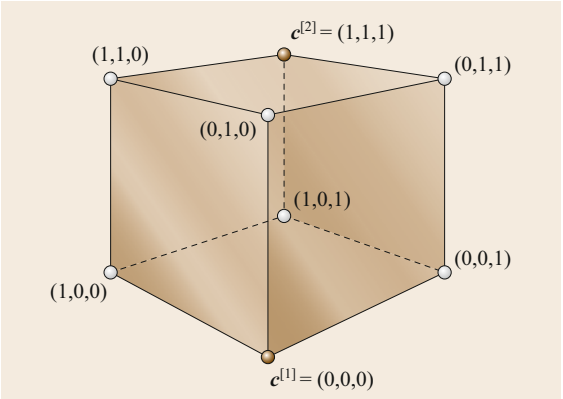


Fig. 7.7 Subspace representation of the $(3, 1)$-repetition code

The code, as well as the encoder (i.e., the mapping of information words to codewords), is completely specified by the generator matrix $\mathbf{G}$. Alternatively, a linear block code can be defined through its $(n-k) \times n$ parity-check matrix $\mathbf{H}$. Since a binary (n, k) linear block code C is a k-dimensional subspace of $\{0, 1\}^n$, its *null* (or *dual*) space, i.e., the set of all binary words of length n that are orthogonal to every element in C, is an $(n-k)$-dimensional subspace of $\{0, 1\}^n$. Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are orthogonal if their inner product is zero, i.e., $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}\boldsymbol{b}^{\top} = 0$. We denote the null space of C by $C_{\perp}$ and write

$$C_{\perp} = \{\tilde{\boldsymbol{c}} : \langle \tilde{\boldsymbol{c}}, \boldsymbol{c} \rangle = 0 \quad \text{for all } \boldsymbol{c} \in C\} \,. \tag{7.30}$$

Note that $C_{\perp}$ is a binary $(n, n-k)$ linear block code, usually referred to as the *dual code* of C. Let $\mathbf{H}$ be a generator matrix of the code $C_{\perp}$, of dimensions $(n-k) \times n$. The rows of $\mathbf{H}$, $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{n-k}$, are linearly independent codewords of $C_{\perp}$, i.e., $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{n-k} \in C_{\perp}$, and every codeword in $C_{\perp}$ is a linear combination of these $n-k$ codewords. Since every codeword $\boldsymbol{c} \in C$ is orthogonal to every codeword $\tilde{\boldsymbol{c}} \in C_{\perp}$, it follows that

$$\boldsymbol{c}\mathbf{H}^{\top} = \mathbf{0} \,. \tag{7.31}$$

Note that (7.31) is satisfied *if and only if* $\boldsymbol{c} \in C$. Therefore, we can define a code C through the generator matrix of its dual code $C_{\perp}$, $\mathbf{H}$.

Definition 7.7

A binary vector $\boldsymbol{c}$ is a codeword of C if and only if $\boldsymbol{c}\mathbf{H}^\top = \mathbf{0}$, i.e., the code C is defined as the null space of $\mathbf{H}$,

$$C = \left\{\boldsymbol{c} : \boldsymbol{c}\mathbf{H}^\top = \mathbf{0}\right\}. \tag{7.32}$$

Therefore, a linear block code is uniquely specified by $\mathbf{G}$ or $\mathbf{H}$. Usually, the generator matrix is used for encoding, while the decoding is based on $\mathbf{H}$.

A way to interpret (7.31) is by noting that $\boldsymbol{c}\mathbf{H}^\top = \mathbf{0}$ is a set of $n-k$ linearly independent equations

$$\begin{array}{ccccccc} h_{1,1}c_1 & + & h_{1,2}c_2 & + \ldots + & h_{1,n}c_n & = 0 \\ h_{2,1}c_1 & + & h_{2,2}c_2 & + \ldots + & h_{2,n}c_n & = 0 \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ h_{(n-k),1}c_1 & + & h_{(n-k),2}c_2 & + \ldots + & h_{(n-k),n}c_n & = 0 \end{array} \tag{7.33}$$

where $h_{i,j}$ is the entry of the matrix $\mathbf{H}$ at row i and column j. These equations are known as *parity-check equations*, and are the equations that each codeword must satisfy. As a result, the matrix $\mathbf{H}$ is commonly referred to as the *parity-check matrix* of the code C.

A practically important class of (linear) codes is the class of *systematic* codes.

Definition 7.8 Systematic code

A *systematic code* is a code with an encoder that generates codewords $\boldsymbol{c}$ containing the information word $\boldsymbol{u}$ as verbatim copy in $\boldsymbol{c}$.

The generator matrix of a *systematic linear code* (assuming that the k information bits appear in the first k positions of the codeword) can always be written as $\mathbf{G} = (\mathbf{I}_k \ \mathbf{P})$, where $\mathbf{I}_k$ is a $k \times k$ identity matrix. The codewords of a systematic code can then be written as

$$\begin{aligned} \boldsymbol{c} &= (c_1 \ldots, c_n) = (u_1, , \ldots, u_k, p_1, \ldots p_{n-k}) \\ &= (\boldsymbol{u} \ \boldsymbol{p}), \end{aligned} \tag{7.34}$$

i.e., the first k bits of the codeword are a replica of the information bits. The code bits $p_1, p_2, \ldots, p_{n-k}$ are referred to as the *parity bits* of the code.

Example 7.2

Consider the $(7, 4)$ binary block code defined by the 4×7 generator matrix

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}. \tag{7.35}$$

Table 7.2 Codewords of the $(7, 4)$ Hamming code

u_1	u_2	u_3	u_4	p_1	p_2	p_3
0	0	0	0	0	0	0
0	0	0	1	1	0	1
0	0	1	0	1	1	1
0	0	1	1	0	1	0
0	1	0	0	0	1	1
0	1	0	1	1	1	0
0	1	1	0	1	0	0
0	1	1	1	0	0	1
1	0	0	0	1	1	0
1	0	0	1	0	1	1
1	0	1	0	0	0	1
1	0	1	1	1	0	0
1	1	0	0	1	0	1
1	1	0	1	0	0	0
1	1	1	0	0	1	0
1	1	1	1	1	1	1

Table 7.2 lists all codewords of this code, which is the famous $(7, 4)$ Hamming code. We observe that the first four bits of each codeword correspond to the information bits, i.e., the code is systematic.

7.3.2 Error Detection and Error Correction Capability of a Block Code over the Binary Symmetric Channel

Consider an $(n, k, d_{\min})$ block code that is used for transmission over the AWGN channel using BPSK modulation. As discussed in Sect. 7.2, if hard decisions on the received vector are taken, the discrete channel between the encoder and the decoder can be modeled as a binary symmetric channel (BSC).

For transmission over the BSC, a block code with minimum Hamming distance $d_{\min}$ can *correct* all error patterns with

$$t = \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor \tag{7.36}$$

or fewer errors, where $\lfloor \tau \rfloor$ denotes the largest integer smaller than or equal to τ. The parameter t is usually referred to as the *error correction capability* of the code, and a code that can correct up to t errors is called a t-error-correcting code. Furthermore, a block code with minimum Hamming distance $d_{\min}$ can *detect* all error patterns with

$$d = d_{\min} - 1 \tag{7.37}$$

or fewer errors.

Example 7.3

The minimum Hamming distance of the 3-repetition code (Example 7.1) and the $(7,4)$ Hamming code in Example 7.2 is $d_{\min} = 3$; hence both codes can correct up to $t = 1$ error and detect up to $d = 2$ errors. The repetition code has rate $R = 1/3$, while the rate of the Hamming code is $4/7$; hence the latter is spectrally more efficient.

The minimum Hamming distance of a code is a very relevant parameter that determines how many errors the code can correct (or detect) for transmission over the BSC. For other channels, the minimum Hamming distance does not completely determine the code performance, but it dominates the code performance (under ML decoding) for high SNRs. Thus, to achieve low error rates, one should construct codes with good minimum Hamming distance.

7.3.3 Optimal Decoding of Block Codes

We have seen that we can characterize a block code in terms of its minimum Hamming distance. We now derive the optimal decoder for block codes for both SDD and HDD. We consider binary transmission using BPSK modulation (i.e., $\mathcal{X} = \{\pm 1\}$) by modulating every code bit c_i, $i = 1, \ldots, n$, of the codeword as $x_i = (-1)^{c_i}$, i.e., without loss of generality we assume the mapping $0 \mapsto +1$ and $1 \mapsto -1$. For example, for $\boldsymbol{c} = (0,0,1,1,0)$, we get $\boldsymbol{x} = (+1,+1,-1,-1,+1)$. Instead of BPSK, we can also consider quadrature phase-shift keying (QPSK) by independently modulating the I and Q components of the constellation. We assume transmission over a memoryless AWGN channel with zero mean and noise variance σ^2. At the output of the channel, we observe a (noisy) vector $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{x} = (x_1, \ldots, x_n)$ is the transmitted (modulated) codeword and $\boldsymbol{n}$ is the noise vector.

Based on the noisy observation $\boldsymbol{y}$, the role of the decoder is to optimally *estimate* the transmitted codeword $\boldsymbol{c}$. Decoding is a probabilistic *inference* problem: the decoder estimates the *most likely* transmitted codeword $\boldsymbol{c}$ (from which it can reverse-lookup the information bit sequence) given the noisy channel output $\boldsymbol{y}$. The most natural decoding criterion is that of minimizing the probability of error. If the decoding criterion is to minimize the probability that the decoder fails to decode to the correct codeword, i.e., minimize the *codeword error probability*, it can be shown that this is equivalent to maximizing the a posteriori probability $P_{C|Y}(\boldsymbol{c}|\boldsymbol{y})$. The *maximum a posteriori* (MAP) decoding rule is therefore

$$\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}\in C} P_{C|Y}(\boldsymbol{c}|\boldsymbol{y}) = \arg\max_{\boldsymbol{c}\in C} \frac{p_{Y|C}(\boldsymbol{y}|\boldsymbol{c}) P_C(\boldsymbol{c})}{p_Y(\boldsymbol{y})} , \tag{7.38}$$

where in the last equality we used Bayes' law. Typically, we can assume that all codewords are equiprobable, i.e., $P_C(\boldsymbol{c}_i) = 1/|C|$, for all i. In this case, and since $p_Y(\boldsymbol{y})$ is independent of $\boldsymbol{c}$, (7.38) can be rewritten as

$$\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}\in C} p_{Y|C}(\boldsymbol{y}|\boldsymbol{c}) , \tag{7.39}$$

which is referred to as the ML decoding rule. Note that the situation changes if probabilistic amplitude shaping (PAS), as described in Sect. 7.6.5, is used. In this case, not all codewords are equiprobable. For a detailed exposure, see [7.91].

In the next two sections, we expand the ML decoding rule (7.39) for SDD and HDD.

Soft-Decision Decoding

Starting from the ML decoding rule (7.39) and assuming a memoryless channel, i.e., $p_{Y|C}(\boldsymbol{y}|\boldsymbol{c}) = \prod_{i=1}^{n} p_{Y|C}(y_i|c_i)$, we can proceed as

$$\begin{aligned}
\hat{\boldsymbol{c}} &= \arg\max_{\boldsymbol{c}\in C} \prod_{i=1}^{n} p_{Y|C}(y_i|c_i) \\
&\overset{(a)}{=} \arg\max_{\boldsymbol{c}\in C} \ln \prod_{i=1}^{n} p_{Y|C}(y_i|c_i) \\
&= \arg\max_{\boldsymbol{c}\in C} \sum_{i=1}^{n} \ln p_{Y|C}(y_i|c_i) ,
\end{aligned} \tag{7.40}$$

where (a) follows because $\ln(\cdot)$ is a monotonically increasing function. To perform ML decoding, we need to have access to the channel law $p_{Y|C}(y_i|c_i)$, which is the conditional probability (density) that we have received y_i given that c_i was transmitted. If known, the true channel law can be used in (7.40). However, we usually resort to models. In the case of coherent optical communications, the channel is reasonably well modeled by an AWGN channel, as predicted by the GN models and their extensions [7.99, 103]. Another example is the case of amplitude-modulated solitons, where the channel law can be modeled as a noncentral chi-squared distribution with four degrees of freedom [7.104]. Furthermore, when higher-order modulation formats with bit-wise decoding (often called BICM) are used (see also Sect. 7.6.3), the equivalent noise that the FEC decoder observes does not necessarily follow a Gaussian distribution [7.105] but can be closely approximated by one.

We will now focus on the case of AWGN, where the channel law is given in (7.1). Continuing the derivation, we have with $x_i = (-1)^{c_i}$,

$$\begin{aligned}
\hat{\boldsymbol{c}} &= \arg\max_{\boldsymbol{c}\in C} \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(y_i-(-1)^{c_i})^2}{2\sigma^2}} \right) \\
&= \arg\max_{\boldsymbol{c}\in C} \sum_{i=1}^{n} \left[\ln\left(\frac{1}{\sqrt{2\pi}\sigma} \right) - \frac{[y_i-(-1)^{c_i}]^2}{2\sigma^2} \right]
\end{aligned}$$

$$\stackrel{(b)}{=} \arg\max_{\boldsymbol{c}\in C} \sum_{i=1}^{n} \frac{-[y_i-(-1)^{c_i}]^2}{2\sigma^2}$$
$$\stackrel{(c)}{=} \arg\min_{\boldsymbol{c}\in C} \sum_{i=1}^{n} \frac{[y_i-(-1)^{c_i}]^2}{2\sigma^2}$$
$$\stackrel{(d)}{=} \arg\min_{\boldsymbol{c}\in C} \sum_{i=1}^{n} [y_i-(-1)^{c_i}]^2 \,, \tag{7.41}$$

where in (b) we exploit the fact that addition of a constant does not change the maximization, in (c) we use the fact that maximizing is equivalent to minimizing the negation, and finally in (d) dividing by the constant $2\sigma^2$ does not change the minimization. Equation (7.41) is the famous Euclidean distance metric, as we take the squared Euclidean distance between (modulated) codewords and the received values. Therefore, for SDD and transmission over the AWGN channel, the ML decoder chooses among all possible transmitted codewords the codeword $\boldsymbol{c}$ that minimizes the Euclidean distance $d_{\mathrm{E}}(\boldsymbol{x},\boldsymbol{y}) \triangleq \|\boldsymbol{x}-\boldsymbol{y}\|$ between $\boldsymbol{y}$ and the BPSK-modulated codeword $\boldsymbol{x} = (x_1,\ldots,x_n) = ((-1)^{c_1},\ldots,(-1)^{c_n})$, i.e.,

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}\in C} d_{\mathrm{E}}(\boldsymbol{x},\boldsymbol{y}) \,. \tag{7.42}$$

Note that the Euclidean distance follows solely from the assumption of Gaussian noise.

We can further simplify the expression (7.41) yielding

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}\in C} \sum_{i=1}^{n} \left[y_i^2 + (-1)^{2c_i} - 2y_i(-1)^{c_i}\right]$$
$$= \arg\min_{\boldsymbol{c}\in C} \sum_{i=1}^{n} \left[y_i^2 + 1 - 2y_i(-1)^{c_i}\right]$$
$$\stackrel{(e)}{=} \arg\min_{\boldsymbol{c}\in C} \sum_{i=1}^{n} \left[-2y_i(-1)^{c_i}\right]$$
$$\stackrel{(f)}{=} \arg\max_{\boldsymbol{c}\in C} \sum_{i=1}^{n} y_i(-1)^{c_i} \,, \tag{7.43}$$

where in (e) we used the fact that addition of y_i^2+1 does not change the maximization and in (f) we replaced minimization by maximization of the negation.

Hard-Decision Decoding

We now derive the optimal decoder for HDD. We start again from the ML decoding rule, assuming a memoryless channel,

$$\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}\in C} p_{Y|C}(\boldsymbol{y}|\boldsymbol{c})$$
$$= \arg\max_{\boldsymbol{c}\in C} \prod_{i=1}^{n} p_{Y|C}(y_i|c_i) \,. \tag{7.44}$$

Note that for HDD, the output sequence $\boldsymbol{y} = (y_1,\ldots,y_n)$ is also a binary sequence, and the channel can be modeled as a BSC with crossover probability p (see (7.2)), i.e.,

$$p_{Y|X}(y_i|x_i) = p_{Y|C}(y_i|c_i) = \begin{cases} p & y_i \neq c_i \\ 1-p & y_i = c_i \,. \end{cases} \tag{7.45}$$

Using this in (7.44),

$$\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}\in C} p^{d_{\mathrm{H}}(\boldsymbol{c},\boldsymbol{y})}(1-p)^{n-d_{\mathrm{H}}(\boldsymbol{c},\boldsymbol{y})}$$
$$= \arg\max_{\boldsymbol{c}\in C} \ln\left[p^{d_{\mathrm{H}}(\boldsymbol{c},\boldsymbol{y})}(1-p)^{n-d_{\mathrm{H}}(\boldsymbol{c},\boldsymbol{y})}\right]$$
$$= \arg\max_{\boldsymbol{c}\in C} d_{\mathrm{H}}(\boldsymbol{c},\boldsymbol{y}) \ln p + [n-d_{\mathrm{H}}(\boldsymbol{c},\boldsymbol{y})] \ln(1-p)$$
$$= \arg\max_{\boldsymbol{c}\in C} d_{\mathrm{H}}(\boldsymbol{c},\boldsymbol{y}) \ln\left(\frac{p}{1-p}\right)$$
$$= \arg\min_{\boldsymbol{c}\in C} d_{\mathrm{H}}(\boldsymbol{c},\boldsymbol{y}) \,, \tag{7.46}$$

where in the last equality we assume that $p < (1/2)$.

We see that for HDD, the ML decoder must choose among all possible transmitted codewords the codeword $\boldsymbol{c}$ that minimizes the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$.

We illustrate the ML decoder using a simple example.

Example 7.4

Assume that we want to transmit the data bit $u = 0$ using the 3-repetition code of Example 7.1. Information bit $u = 0$ is encoded onto codeword $\boldsymbol{c} = (0,0,0)$, which is modulated to $\boldsymbol{x} = (+1,+1,+1)$. After transmission over the AWGN channel, we receive $\boldsymbol{y} = (-0.2,+1.1,-0.7)$. We can see that two sign changes occurred during transmission, so a HDD looking purely at the sign (and hence minimizing the Hamming distance) will erroneously assume that the codeword $(1,1,1)$ was transmitted. However, when we carry out the SDD ML decoding rule, we observe that for codewords $(0,0,0)$ and $(1,1,1)$ we get

$$(0,0,0): \quad \sum_{t=1}^{3} y_t(-1)^0 = +0.2 \,, \tag{7.47}$$
$$(1,1,1): \quad \sum_{t=1}^{3} y_t(-1)^1 = -0.2 \,. \tag{7.48}$$

We can see that the sum is maximized for $(0,0,0)$; thus the decoder (correctly) decides for $\hat{\boldsymbol{c}} = (0,0,0)$ and hence $\hat{u} = 0$.

The example above highlights that HDD incurs a performance penalty with respect to SDD. On the other hand, HDD is usually significantly less complex than SDD.

Decoding Complexity

The complexity of the ML decoder scales exponentially with the number of data bits k. Therefore, ML decoding is only feasible for simple (short) codes and quickly becomes prohibitive as k increases. For practical codes, k is easily above 10 000, which means that the code can contain more than 2×10^{3010} codewords. If we assume that we have access to the 500 largest supercomputers worldwide, which have an aggregated computational capacity of 748 PFlop/s [7.106], and that we only require a single floating point operation to compute the sum required in (7.43) or the Hamming distance in (7.44), we would still require 2.7×10^{2992} seconds to evaluate all codewords. As the universe has existed for roughly 4.3×10^{17} seconds, we see that all efforts to carry out ML decoding for such codes are futile. Therefore, suboptimal decoding strategies which yield an acceptable decoding complexity are required.

The quest for long, powerful codes with low decoding complexity has given rise to so-called *modern codes*, such as LDPC codes and product-like codes, which build upon simpler component codes and are decoded using suboptimal (iterative) decoding algorithms, with reasonable decoding complexity and yet astonishing performance. Suboptimal decoding strategies are discussed in Sects. 7.4 and 7.5.

7.3.4 Coding Gain and Net Coding Gain

An important measure for comparing different coding schemes is the NCG. To determine the NCG, bipolar signaling and an AWGN channel are frequently used. However, the definition can be easily extended to other types of channels. Consider an AWGN channel with SNR E_s/N_0. Since the modulated symbol x is obtained as $x = (-1)^{c_i}$, all modulation symbols have the same energy, and hence E_s also denotes the *energy per coded bit* in this setting. We therefore have $E_s = \mathbb{E}[X^2] = 1$. In a transmission system, E_s is usually fixed (for example, to the value corresponding to the laser transmit power). Here, we assume a normalized transmit energy $E_s = 1$. Instead of the SNR E_s/N_0, for a fair comparison of coding schemes it is helpful to use the ratio E_b/N_0, where E_b denotes the average *energy per information bit*. For example, if a code of rate $R = 4/5$ is used, corresponding to overhead OH = 25%, the ratio of code bits n versus information bits k amounts to $n/k = 5/4 = 1.25$, i.e., 1.25 code bits are transmitted for each information bit. This means that if the code bits are transmitted each with energy E_s, the equivalent amount of energy conveyed by each information bit amounts to $E_b \triangleq E_s/R \geq E_s$. As E_b is normalized to the information bits of the transmission system, it allows us to evaluate the NCG, i.e., the gain that is actually achievable by taking into account the coding. The NCG is defined as the gain (in dB) of the coded transmission in terms of E_b/N_0 relative to uncoded transmission for a given output BER.

Figure 7.8 illustrates the NCG and the influence of E_b/N_0 versus E_s/N_0. First, we consider uncoded transmission ($R = 1$), where $E_b = E_s = 1$. In this case, the BER is obtained as $Q(\sqrt{2E_b/N_0}) = Q(\sqrt{2E_s/N_0})$. If we now apply coding with a rate $R = 4/5$, we get

$$\frac{E_b}{N_0} = \frac{E_s}{R N_0} = \frac{1}{R}\frac{E_s}{N_0}, \tag{7.49}$$

i.e., E_b/N_0 is scaled by a factor $1/R$, corresponding to $-10\log_{10}(R) = 0.969$ dB. Hence, by applying coding, the energy per information bit is scaled by this factor, i.e., each information bit carries 0.969 dB additional energy compared with an uncoded transmission. We may also take on a different perspective and fix the energy per information bit E_b (e.g., to $E_b = 1$). From this perspective, as coding increases the effective number of transmitted bits per information bit, the effective energy per transmitted symbol is decreased. In this case, when the receiver cannot apply decoding but needs to recover only the (systematic) information bits from the sequence of transmitted bits, it requires an SNR that is increased by 0.969 dB to achieve the same BER. By carrying out decoding, we can now significantly improve the post-decoding BER starting from this shifted curve, as can be seen in Fig. 7.8. This increase corresponds to the coding gain (sometimes also called *gross* coding gain), indicated by the bottom arrow in Fig. 7.8. The NCG is reduced by the 0.969 dB required for transmitting the extra parity bits. In this example, we see that the coding scheme employed here yields an NCG of 11.9 dB at a BER of 10^{-15}.

In optical communications, the optical signal-to-noise ratio (OSNR) is also frequently employed. The OSNR is the SNR measured in a reference optical bandwidth, where frequently a bandwidth B_{ref} of 12.5 GHz is used, corresponding to 0.1 nm-resolution wavelength at a carrier wavelength of 1550 nm. The OSNR relates to the E_s/N_0 as follows

$$\begin{aligned} 10\log_{10}\mathrm{OSNR} &= 10\log_{10}\frac{E_s}{N_0} + 10\log_{10}\frac{R_s}{B_{ref}} \\ &= 10\log_{10}\frac{E_b}{N_0} + 10\log_{10}\frac{mRR_s}{B_{ref}}, \end{aligned} \tag{7.50}$$

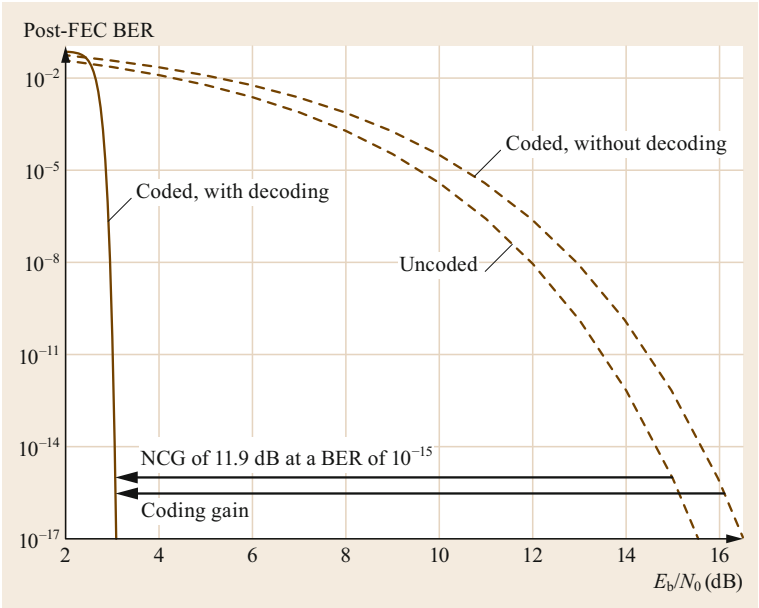


Fig. 7.8 Net coding gain definition

where B_{ref} is the reference bandwidth, R_s corresponds to the symbol rate of the transmission, R is the code rate, and m is the number of bits mapped to each modulation symbol.

7.3.5 Bounds on the Performance of Block Codes

From information theory (Sect. 7.2), we know that the required SNR for successful transmission, based on Shannon's capacity results, assumes infinite block lengths. This is usually not a valid assumption in practical communication systems. Unfortunately, selecting a finite (moderate) code length does not necessarily lead to good performance. Much work has been dedicated to obtaining precise bounds on the performance of transmission systems employing codes of finite block length n. A good overview of results can be found in [7.107] and [7.108], where the latter provides refined finite-length performance bounds. Shannon already provided a lower bound on the codeword error probability P_w, i.e., the probability that the decoder decodes an incorrect codeword $\hat{\boldsymbol{c}} \neq \boldsymbol{c}$, based on geometric arguments [7.107, 109].

In the following we use the normal approximation of the finite-block-length bound for the binary-input AWGN channel given in [7.110]. This bound allows us to compute an approximation of the gap relative to capacity for the binary-input AWGN channel (which describes BPSK and QPSK modulation with AWGN noise, see also at the beginning of Sect. 7.3.3). We assume that the decoder targets a codeword error probability of $P_w = 10^{-13}$, which means that one in 10^{13} codewords can be decoded erroneously. With this, the bit error probability P_b can be bounded as

$$\frac{1}{Rn} P_w \leq P_b \leq P_w \,, \tag{7.51}$$

where the upper bound assumes that all bits inside an erroneously decoded codeword are in error, and the lower bound assumes a single bit error only. Figure 7.9 shows

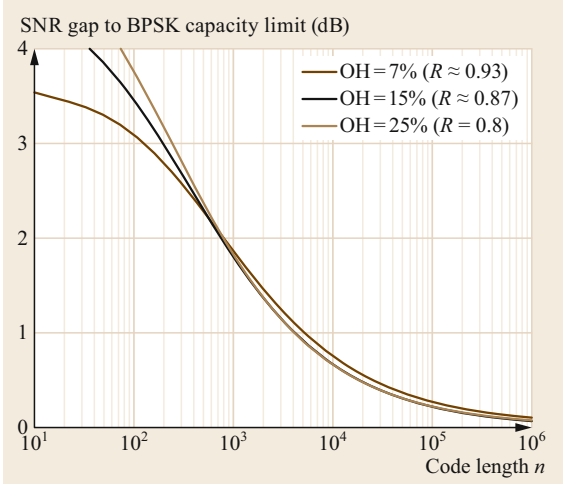


Fig. 7.9 Gap to the BPSK capacity as a function of the block size n for codes of different rates R and a codeword error probability $P_w = 10^{-13}$

the normal approximation for OH $\in \{7\%, 15\%, 25\%\}$ and a target codeword error rate of $P_w = 10^{-13}$, which is a value commonly used in fiber-optical communications (when the target P_b is around 10^{-15}). We often use Monte Carlo simulations to estimate the probabilities P_w and P_b. In this case, the estimated quantities are referred to as frame error rate (FER) and BER, respectively. In the remainder of this chapter, with some abuse of language, we will often use FER to refer to P_w and BER to refer to P_b.

We see in Fig. 7.9 that we need block sizes $n >$ 5 kbit to be able to operate within 1 dB from the channel capacity. Unfortunately, most classical coding schemes are only defined for small to medium n (usually less than 1 kbit) or do not perform well for large n. Code concatenation was a first step toward the realization of longer codes that proved to be very successful with HDD. In the next section, we will introduce different coding schemes that naturally allow the construction of long codes.

7.4 Soft-Decision Forward Error Correction

Contrary to the historical development of FEC in fiber-optic communications, we start by introducing SDD, and later discuss HDD in Sect. 7.5. The reason for doing so is that SDD immediately follows when optimally decoding many channels that are encountered in optical communications, such as the AWGN channel (Sect. 7.3.3). Furthermore, recently proposed hard-decision FEC schemes have strong ties with soft-decision FEC schemes (such as LDPC codes) and their analysis tools; hence it is somewhat natural to discuss these first. In this section, we introduce some popular coding schemes for SDD in fiber-optic communications, including LDPC codes, SC-LDPC codes, and polar codes. We discuss these codes and their variants in detail, and give detailed descriptions of the most popular decoding algorithms and construction methods. We further show the strengths and weaknesses of the different schemes by means of various simulation examples.

7.4.1 Low-Density Parity-Check Codes

We have seen in Sect. 7.3.5 that codes with large block sizes n are required to operate close to the achievable rate or capacity of the channel. For example, from Fig. 7.9, n should be $\geq 2\times10^4$ if we want to operate within 0.5 dB of the capacity of the binary-input AWGN channel with a code of overhead OH = 25%. As seen in Sect. 7.3.1, a code is completely specified by its parity-check matrix $\mathbf{H}$, of dimensions $m \times n$, with $m < n$. Most of classical coding theory focused on finding algebraic constructions for $\mathbf{H}$ (e.g., BCH codes), avoiding the need to store $\mathbf{H}$. However, in the general case, when $\mathbf{H}$ does not contain any *structure* and is random-like, storing it is difficult. The storage requirements for $\mathbf{H}$ scale as $O(n^2)$, and hence it may already not be feasible to store $\mathbf{H}$ for relatively small values of n.

We thus need practical codes that have enough structure to enable storage of $\mathbf{H}$ using data structures with low memory requirements and that allow operation close to the achievable rate. Inspired by the fact that random parity-check codes, i.e., codes where the entries $h_{i,j}$ of $\mathbf{H}$ are chosen to be *0* or *1* with probability 1/2, are capacity-achieving, to construct codes with high coding gains and efficient storage of $\mathbf{H}$, *Gallager* introduced LDPC codes in the 1960s [7.54].

An LDPC code is generally defined by a *sparse* binary parity-check matrix $\mathbf{H}$. Recall from Sect. 7.3 that each column of the parity-check matrix $\mathbf{H}$ corresponds to a code bit. Likewise, each row of $\mathbf{H}$ corresponds to a parity-check equation and ideally defines a single parity bit (if $\mathbf{H}$ has full rank). Sparse means that the number of *1*s in $\mathbf{H}$ is small compared with the number of zero entries. Practical codes usually have a fraction of *1*s that is below 1% by several orders of magnitude. Usually, provided that the parity-check matrix has full row rank, i.e., rank($\mathbf{H}$) $= m$, the number of information bits equals $k = n - m$. If the parity-check matrix $\mathbf{H}$ is rank-deficient, the number of information bits k can be larger, i.e., $k \geq n - m$. In the remainder of this chapter, we fix $k = n - m$, i.e., we do not actively use any possible additional information bits. In this case, the rate is given by $R = (n-m)/n$ and the overhead equals OH $= m/(n-m)$.

LDPC codes are often subdivided into two major classes, *regular* and *irregular* LDPC codes. We describe both classes in what follows.

Regular Low-Density Parity-Check Codes

Definition 7.9 Regular (d_v, d_c) LDPC Code
A regular (d_v, d_c) LDPC code is an LDPC code where each column of $\mathbf{H}$ is of weight d_v (i.e., it contains exactly d_v *1*s), and each row of $\mathbf{H}$ is of weight d_c. We assume $d_v \geq 2$ and $d_c \geq 2$.

The storage complexity of (d_v, d_c) LDPC codes is only $O(n)$, as the column weight is constant (and indepen-

dent of n), which enables efficient implementation of LDPC codes with low memory requirements.

As an example, consider the $(3, 6)$ LDPC code of rate $R = 1/2$, with $n = 18$, $m = 9$ defined by the parity-check matrix

$$\mathbf{H} = \begin{pmatrix} 0&1&0&0&1&0&0&0&1&0&1&0&0&0&1&0&1&0 \\ 1&0&0&0&0&1&1&0&0&0&0&1&0&1&0&1&0&0 \\ 0&0&1&1&0&0&0&1&0&1&0&0&1&0&0&0&0&1 \\ 0&1&0&0&0&1&0&0&1&0&1&0&0&0&1&1&0&0 \\ 0&0&1&0&1&0&0&1&0&0&0&1&0&1&0&0&1&0 \\ 1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&0&0&1 \\ 0&1&0&0&0&1&0&0&1&1&0&0&1&0&0&1&0&0 \\ 1&0&0&0&1&0&1&0&0&0&0&1&0&1&0&0&1&0 \\ 0&0&1&1&0&0&0&1&0&0&1&0&0&0&1&0&0&1 \end{pmatrix} . \tag{7.52}$$

We summarize in the following some of the properties of regular $(d_\mathrm{v}, d_\mathrm{c})$ LDPC codes:

- $n\,d_\mathrm{v} = m\,d_\mathrm{c}$ has to hold, as the total number of *1*s of $\mathbf{H}$ is identical regardless of the way we count them. Therefore, the rate R is given by

$$R = 1 - \frac{m}{n} = 1 - \frac{n\,d_\mathrm{v}}{n\,d_\mathrm{c}} = 1 - \frac{d_\mathrm{v}}{d_\mathrm{c}} \,. \tag{7.53}$$

- If d_v is even, then the summation of all rows of $\mathbf{H}$ is the all-zero row, i.e., there is at least one linearly dependent row, i.e., $\mathrm{rank}(\mathbf{H}) \leq m - 1$.
- With high probability, the parity-check matrix of a regular LDPC code has asymptotically full rank [7.111], i.e.,

$$\lim_{n\to\infty} P(\mathrm{rank}(\mathbf{H}) = m - \mathbb{1}_{\{d_\mathrm{v}\ \mathrm{even}\}}) = 1 \,. \tag{7.54}$$

Already for small values of n, the limit in (7.54) is approached.

Irregular Low-Density Parity-Check Codes

In some cases it can be advantageous to relax the regularity and allow codes where the *distribution* of *1*s in the parity-check matrix is *irregular* [7.112]. This gives rise to *irregular* LDPC codes, which are formally defined as follows.

Definition 7.10 Irregular $(\boldsymbol{\delta}_\mathrm{v}, \boldsymbol{\delta}_\mathrm{c})$ LDPC Code

An irregular $(\boldsymbol{\delta}_\mathrm{v}, \boldsymbol{\delta}_\mathrm{c})$ LDPC code, with $\boldsymbol{\delta}_\mathrm{v} = (\delta_{\mathrm{v},1}, \delta_{\mathrm{v},2}, \ldots, \delta_{\mathrm{v},d_{\mathrm{v,max}}})$ and $\boldsymbol{\delta}_\mathrm{c} = (\delta_{\mathrm{c},1}, \delta_{\mathrm{c},2}, \ldots, \delta_{\mathrm{c},d_{\mathrm{c,max}}})$ is an LDPC code with $\delta_{\mathrm{v},i}\,n$ columns of $\mathbf{H}$ of weight i and $\delta_{\mathrm{c},j}\,m$ rows of $\mathbf{H}$ of weight j. We assume that $\delta_{\mathrm{v},1} = \delta_{\mathrm{c},1} = 0$ and that $\delta_{\mathrm{v},i}\,n$ and $\delta_{\mathrm{c},j}\,m$ are integers for all $i \in \{2, \ldots, d_{\mathrm{v,max}}\}$ and $j \in \{2, \ldots, d_{\mathrm{c,max}}\}$, respectively. Note that $\sum_{i=1}^{d_{\mathrm{v,max}}} \delta_{\mathrm{v},i} = \sum_{j=1}^{d_{\mathrm{c,max}}} \delta_{\mathrm{c},j} = 1$. The vectors $\boldsymbol{\delta}_\mathrm{v}$ and $\boldsymbol{\delta}_\mathrm{c}$ are called the *degree distribution* (from a node perspective) of the code.

Note that irregular codes comprise the class of regular LDPC codes. Indeed, a regular $(d_\mathrm{v}, d_\mathrm{c})$ LDPC code is described by the degree distribution vectors $\boldsymbol{\delta}_\mathrm{v}$ and $\boldsymbol{\delta}_\mathrm{c}$ with $\delta_{\mathrm{v},d_\mathrm{v}} = 1$ and $\delta_{\mathrm{v},i} = 0$, $\forall i \neq d_\mathrm{v}$, and $\delta_{\mathrm{c},d_\mathrm{c}} = 1$ and $\delta_{\mathrm{c},i} = 0$, $\forall i \neq d_\mathrm{c}$. The rate of an irregular LDPC code can be expressed as a function of the vectors $\boldsymbol{\delta}_\mathrm{v}$ and $\boldsymbol{\delta}_\mathrm{c}$, namely

$$R = 1 - \frac{\sum_{i=1}^{d_{\mathrm{v,max}}} i\,\delta_{\mathrm{v},i}}{\sum_{j=1}^{d_{\mathrm{c,max}}} j\,\delta_{\mathrm{c},j}} \,. \tag{7.55}$$

An important generalization of regular and irregular LDPC codes are multi-edge type (MET) LDPC codes [7.111, Sect. 7.1]. In this class of codes, the parity-check matrix is divided into different regions, called types. Each such region shares some common properties. Some high-performing codes with low error floors have been constructed using this approach. One example is the so-called AR4JA codes [7.113] which have been standardized for deep-space communications [7.114]. These codes, which often require punctured state variables, have not yet found widespread application in fiber-optic communications, as their convergence is relatively slow and a larger number of decoding iterations is required to reach the desired performance.

Graphical Representation of Low-Density Parity-Check Codes

A widespread representation of LDPC codes is based on the so-called *Tanner graphs*, introduced in 1981 by *Michael Tanner* [7.115] as a tool for building large codes from small component codes. Tanner graphs are useful for the design and analysis of LDPC codes and can also serve as a tool for deriving and understanding the most basic decoding algorithm. In what follows, we introduce Tanner graphs by means of an example.

Let C be a binary (LDPC) code with parity-check matrix $\mathbf{H}$ of size $m \times n$. The Tanner graph of the code C is a *bipartite graph*, i.e., it has two types of nodes, and every edge connects a node of the first type with a node of the second type (i.e., there are no connections between nodes of the same type):

- The first type of nodes are called variable nodes (VNs) (they are also sometimes called *bit nodes*), and there are n of them, each one corresponding to a code bit c_i and thus to *column* i of $\mathbf{H}$. We draw the VNs by open circles ○.

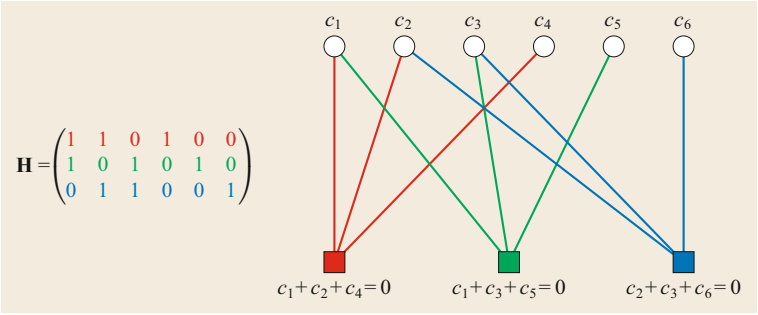


Fig. 7.10 Tanner graph representation of the $(6, 3)$ code given by the parity-check matrix $\mathbf{H}$ in (7.56)

- The second type of nodes are called check nodes (CNs), and there are m of them. Each CN corresponds to a parity-check constraint of the code and thus to a *row* of $\mathbf{H}$. We draw the CNs by filled squares ■.

CN j is connected by an edge to VN c_i if and only if $h_{j,i} = 1$, i.e., if $\mathbf{H}$ contains a 1 in row j and column i. We then say that code bit c_i participates in the j-th parity-check constraint.

Let us illustrate the construction of a Tanner graph using the following $(6, 3)$ code with parity-check matrix

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} . \tag{7.56}$$

The Tanner graph of this code is shown in Fig. 7.10. We use different colors to illustrate the relation between CNs and rows of $\mathbf{H}$ (which use the same coloring).

Consider as a second example the parity-check matrix in (7.52). The Tanner graph corresponding to this matrix is shown in Fig. 7.11. Each of the nine CNs corresponds to a row of the parity-check matrix $\mathbf{H}$. Every code bit has $d_v = 3$ adjacent edges, and every CN has $d_c = 6$ adjacent edges.

The Tanner graph can be described by the connection sets $\mathcal{M}(j)$ and $\mathcal{N}(i)$. The set $\mathcal{M}(j) = \{i : h_{j,i} \neq 0\}$ contains the positions (columns) of nonzero entries at row j of the parity-check matrix $\mathbf{H}$. For the exemplary matrix in (7.52) we have $\mathcal{M}(1) = \{2, 5, 9, 11, 15, 17\}$ and $\mathcal{M}(4) = \{2, 6, 9, 11, 15, 16\}$. Additionally, the set $\mathcal{N}(i) = \{j : h_{j,i} \neq 0\}$ contains the positions (rows) of nonzero entries at column i of the parity-check matrix $\mathbf{H}$. For the matrix in (7.52) we have $\mathcal{N}(1) = \{2, 6, 8\}$ and $\mathcal{N}(2) = \{1, 4, 7\}$.

Decoding Behavior of Low-Density Parity-Check Codes

Before discussing the decoder implementation in detail, we give a high-level overview of the post-FEC BER behavior of LDPC codes as a function of the channel quality (e.g., E_b/N_0). The decoding behavior of LDPC codes is visualized in Fig. 7.12 for an example code. We can distinguish three regions of the BER curve. Up to a specific E_b/N_0 value, the post-FEC BER does not change significantly. For E_b/N_0 values above this value, in the so-called *waterfall region* (highlighted in gray color in the figure), the post-FEC BER decreases rapidly. The slope of this rapidly decreasing BER curve depends mainly on the length of the code n, with longer codes leading to narrower waterfall regions. Finally, at a certain E_b/N_0, the waterfall region ends, the slope of the BER curve abruptly changes, and we observe an *error floor*. In the error floor region, the post-FEC BER only slowly decreases with increasing E_b/N_0. This effect is due mainly to the presence of some structures in

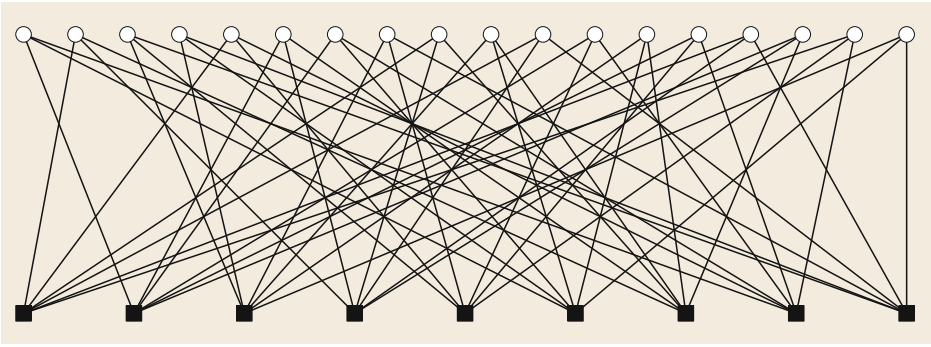


Fig. 7.11 Tanner graph representation of the code with parity-check matrix given by (7.52)

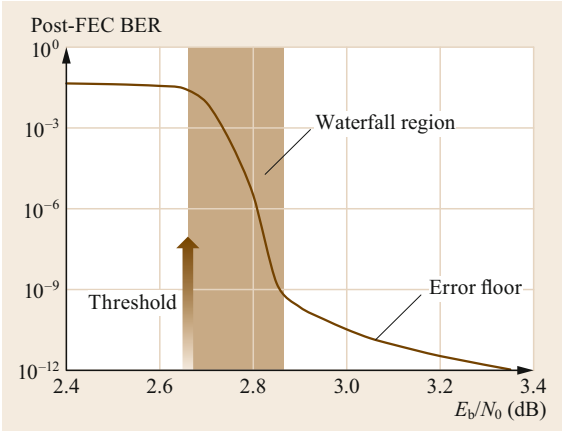


Fig. 7.12 Description of the post-FEC BER behavior for LDPC codes with waterfall region and error floor

the parity-check matrix and Tanner graph of the code that cause the decoder to stall for some specific error patterns [7.63, 116].

The value of E_b/N_0 where the LDPC code performance curve bends into the characteristic waterfall behavior, to which with some abuse of language we refer here as *threshold* (Fig. 7.12), depends on the parameters of the code (i.e., $\boldsymbol{\delta}_v$ and $\boldsymbol{\delta}_c$) and the channel. It can be determined using a technique called *density evolution* [7.117] or approximations thereof [7.118]. Formally speaking, the threshold denotes the E_b/N_0 value at which decoding will be successful (the BER tends to zero), with probability approaching 1 in the asymptotic limit of large block length ($n \to \infty$), provided that the decoder complexity is not constrained. For a detailed discussion, we refer the interested reader to [7.117]. Thresholds of some codes for the binary-input AWGN channel are given in Table 7.3, together with the channel capacity.

From Table 7.3, we see that the thresholds of some irregular LDPC codes are very close to the capacity limits of the binary-input AWGN channel. Indeed, irregular LDPC codes are a very important class of capacity-approaching codes [7.119]. Unfortunately, capacity-approaching irregular LDPC codes typically require parity-check matrices with many weight-2 columns (i.e., $\delta_{v,2} > 0$). Such codes usually have error floors in the range of 10^{-4} to 10^{-8}, depending also on n, that are unacceptable in fiber-optic communications. Hence, there is a trade-off between regular LDPC codes, which usually have low error floors but thresholds bounded away from capacity, and irregular codes, which have thresholds closer to capacity, but often entail unacceptably high error floors. Irregular codes have found important application in fiber-optic communications, especially when designed jointly with iterative decoding and demodulation schemes [7.120, 121]. Later, in Sect. 7.4.5, we will show a simulation example illustrating irregular LDPC codes.

Due to the rapid decrease in BER in the waterfall region, and because of other outstanding properties, LDPC codes for SDD in fiber-optic communications have been studied in multiple publications [7.60, 61, 86, 121, 122]. Due to the error floor, some modern high-performance FEC systems are constructed using a soft-decision LDPC inner code, which reduces the BER to a level of 10^{-3} to 10^{-5}, and a hard-decision outer code which pushes the system BER to levels below 10^{-12} [7.61]. The implementation of a coding system with an outer cleanup code requires a thorough understanding of the LDPC code and a properly designed interleaver between the LDPC and outer code so that the errors at the output of the LDPC decoder—which typically occur in bursts—do not cause uncorrectable blocks after outer decoding. With increasing computing resources, and in particular field-programmable gate array (FPGA)-based decoders, it has now also become feasible to evaluate very low target BERs of LDPC codes and optimize the codes to have very low error floors, below the target BER of the system [7.64, 86].

Construction of Low-Density Parity-Check Matrices

Unlike classical codes such as BCH codes or Reed–Solomon codes, there is no commonly accepted design rule for LDPC codes. Instead, a plethora of LDPC code design methodologies and heuristics exist, each with its own advantages and disadvantages. The goal of an

Table 7.3 Decoding thresholds of some regular and irregular LDPC codes on the AWGN channel

Code	Rate R	E_b/N_0 threshold (dB)	Capacity (E_b/N_0)
Regular (3, 6)	1/2	1.113 dB	0.187 dB
Regular (4, 8)	1/2	1.618 dB	0.187 dB
Irregular code from [7.119]	1/2	0.192 dB	0.187 dB
Regular (3, 15)	4/5	2.586 dB	2.039 dB
Regular (4, 20)	4/5	2.669 dB	2.039 dB
Irregular $\boldsymbol{\delta}_v = (0, 0, (9/10), 0, 0, 0, 0, 0, (1/10))$, $d_c = 18$	4/5	2.441 dB	2.039 dB
Irregular $\boldsymbol{\delta}_v = (0, 0.4, 0.36, 0, 0, 0, 0, 0.24)$, $d_c = 19$	4/5	2.232 dB	2.039 dB

LDPC code designer is to find a code that yields high coding gains and possesses some structure facilitating the implementation of the encoder and decoder. We point the interested reader to numerous articles published on this topic [7.111, 123, 124] and references therein. We provide in the following two ways of constructing parity-check matrices that can serve as a starting point for developing own coding schemes based on LDPC codes.

Regular $(d_\mathrm{v}, d_\mathrm{c})$ LDPC codes can be constructed using Gallager's method [7.54]. This simple method leads to codes that are sufficient for many purposes. In this method, we define the $m \times n$ matrix $\mathbf{H}$ by a *banded* structure. The rows in $\mathbf{H}$ are divided into d_v sets with m/d_v rows in each set (i.e., m must be an integer multiple of d_v). The first set of rows contains d_c consecutive *1*s ordered from left to right across columns, such that each row in the set contains exactly d_c *1*s. Note that this implies that n must be an integer multiple of d_c. Every other set is given by a *randomly chosen column permutation* of the first set. For example, a parity-check matrix of a $(3, 4)$ LDPC code of size 9×12 is constructed as

$$\mathbf{H} = \left(\begin{array}{cccccccccccc}
1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1\\
\hline
1&0&1&0&0&1&0&0&0&1&0&0\\
0&1&0&0&0&0&1&1&0&0&0&1\\
0&0&0&1&1&0&0&0&1&0&1&0\\
\hline
1&0&0&1&0&0&1&0&0&1&0&0\\
0&1&0&0&0&1&0&1&0&0&1&0\\
0&0&1&0&1&0&0&0&1&0&0&1
\end{array}\right) . \quad (7.57)$$

In practice, the permutations should be chosen such that any sub-matrix of the form $\left(\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right)$, i.e., two columns sharing *1*s in two common rows, is avoided. This is the so-called *row-column (RC) constraint*, which can improve the decoding performance significantly. Note that the RC constraint cannot be fulfilled for the matrix in (7.57), as, by the pigeonhole principle, a necessary condition to fulfill the RC constraint with Gallager's construction is that $m \geq d_\mathrm{v} d_\mathrm{c}$. Further note that due to the constraints of the banded Gallager design method, the rank of $\mathbf{H}$ is upper bounded as $\mathrm{rank}(\mathbf{H}) \leq m - d_\mathrm{v} + 1$.

Parity-check matrices can also be obtained via the Tanner graph. This method is particularly useful for irregular LDPC codes. We start with an empty graph and first place the n VNs and the m CNs. Note that each VN of degree i is equipped with i (empty) sockets that connect to an edge in the final graph. Similarly, each CN of degree j is equipped with j empty sockets. In the next step, we place the $n \sum_{i=1}^{d_{\mathrm{v,max}}} i\delta_{\mathrm{v},i}$ edges of the graph as follows: we connect each empty VN socket with a randomly chosen, empty CN socket by an edge such that no VN is connected to the same CN more than once (i.e., parallel edges in the Tanner graph are avoided). As there is a one-to-one correspondence between the Tanner graph and the parity-check matrix, we can construct $\mathbf{H}$ from the graph. Such a matrix will most likely not fulfill the RC constraint. In that case, we may either add a post-processing step that reshuffles some of the edges (i.e., entries of $\mathbf{H}$) such that the RC constraint is fulfilled without changing the degree distribution, or try to fulfill the RC constraint directly while picking the sockets.

These constructions can serve only as a starting point for good designs of parity-check matrices. A popular improvement to the graph-based construction described above is the progressive edge growth (PEG) algorithm [7.125], which adds edges sequentially such that the length of the shortest cycle in the graph is kept small. The ACE algorithm [7.126] also attempts to design LDPC codes with large minimum distance. It can be combined with the PEG algorithm [7.127] as well. In this chapter, we are able to just scratch the surface of all the possibilities and point to the vast body of literature describing different construction methods. For a deep treatment of LDPC codes together with an overview of construction methods, we refer the interested reader to [7.101, Chaps. 6, 10–12].

Quasi-Cyclic Low-Density Parity-Check Codes

The parity-check matrices $\mathbf{H}$ constructed using the methods highlighted in the previous section will be random-like due to the random permutations used in the construction. Unfortunately, when trying to construct a decoder application-specific integrated circuit (ASIC) or implement a decoder on an FPGA, the wiring and routing complexity will usually be prohibitively high even for moderate values of n. Hence, we need constructions with structures that can be leveraged to simplify the wiring and routing in the decoder. For this reason, most LDPC codes that are implemented in practice today are so-called quasi-cyclic (QC) LDPC codes. They have a parity-check matrix with a structure that allows for inherent parallelization of the decoder, reduced wiring complexity, and an efficient encoder realization. QC LDPC codes are constructed using a so-called lifting matrix $\mathbf{A}$. The lifting matrix has integer entries $a_{i,j} \in \{-1, 0, 1, \ldots, S-1\}$, where S denotes the *lifting factor*. The parity-check matrix $\mathbf{H}$ is constructed from $\mathbf{A}$ by replacing each element $a_{i,j}$ with either an all-zero matrix of size $S \times S$ or a cyclically permuted identity matrix of size $S \times S$. If $a_{i,j} = -1$, then the all-zero matrix of size $S \times S$ is used, and if $a_{i,j} \geq 0$, then $a_{i,j}$ denotes the number of cyclic right shifts of the identity matrix that are carried out. If $\dim(\mathbf{A}) = m' \times n'$, then $\dim(\mathbf{H}) = m \times n$, with $m = m'S$ and $n = n'S$.

Formally, we construct the parity-check matrix $\mathbf{H}$ as

$$\mathbf{H} = \begin{pmatrix} \mathbf{P}^{a_{1,1}} & \mathbf{P}^{a_{1,2}} & \dots & \mathbf{P}^{a_{1,n'}} \\ \mathbf{P}^{a_{2,1}} & \mathbf{P}^{a_{2,2}} & \dots & \mathbf{P}^{a_{2,n'}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}^{a_{m',1}} & \mathbf{P}^{a_{m',2}} & \dots & \mathbf{P}^{a_{m',n'}} \end{pmatrix}, \tag{7.58}$$

where $\mathbf{P}$ is the matrix of size $S \times S$ with $p_{i,i+1} = 1$, $\forall i \in \{1, \dots, S-1\}$, $p_{S,1} = 1$, and zeros everywhere else. We note that $\mathbf{P}^0 = \mathbf{I}$, the identity matrix, and

$$\mathbf{P}^j = \underbrace{\mathbf{P}\mathbf{P}\cdots\mathbf{P}}_{j \text{ times}} \tag{7.59}$$

equals the identity matrix cyclically shifted to the right by j. We further define $\mathbf{P}^{-1} \triangleq \mathbf{0}$, the all-zero matrix.

Typically, in modern FEC schemes, a small code with parity-check matrix $\tilde{\mathbf{H}}$ is first constructed, for example, based on Gallager's method or more advanced schemes such as the PEG and ACE algorithms [7.125, 126, 128], MET analysis and design [7.129], protographs [7.130], or finite geometries [7.101, Chap. 10]. See also [7.131] for an extensive overview and attempts at unifying construction methods. In a second step, we replace the nonzero elements of $\tilde{\mathbf{H}}$ by appropriately selected shift values. As an example, we can take the matrix of (7.57) and convert it to a lifting matrix to obtain, with $S = 16$,

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 3 & -1 & 7 & -1 & -1 & 1 & -1 & -1 & -1 & 3 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 & -1 & 3 & 6 & -1 & -1 & -1 & 5 \\ -1 & -1 & -1 & 5 & 9 & -1 & -1 & -1 & 9 & -1 & 2 & -1 \\ 7 & -1 & -1 & 3 & -1 & -1 & 12 & -1 & -1 & 10 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 & 0 & -1 & 4 & -1 & -1 & 7 & -1 \\ -1 & -1 & 14 & -1 & 14 & -1 & -1 & -1 & 8 & -1 & -1 & 9 \end{pmatrix}. \tag{7.60}$$

After replacing the entries of the lifting matrix by $\mathbf{P}^{a_{i,j}}$, we obtain the *lifted*, structured parity-check matrix $\mathbf{H}$ of size 144×192 shown in Fig. 7.13, where we added dashed lines to serve as visual guides. The *1*s are represented by black dots in Fig. 7.13, and the *0*s are omitted. The main task of the code designer is to find good values $a_{i,j}$. Many design rules for finding these values have been proposed in the literature. Commonly, the values $a_{i,j}$ are selected to maximize the so-called girth of the code, which is indirectly related to the decoding performance [7.132]. In the example above, we have selected the entries $a_{i,j}$ to yield a girth-10 code. Other constructions are aimed at maximizing a bound of the minimum distance [7.133], and use finite geometries [7.101, Chap. 10], matrix dispersions

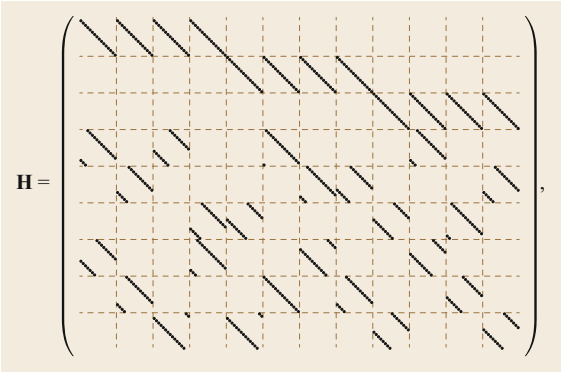


Fig. 7.13 Lifted structured parity-check matrix of size 144×192 obtained by lifting the entries of (7.60) with cyclic permutation matrices of size 16×16

of finite fields [7.101, Chap. 11], [7.131], combinatorial designs [7.101, Chap. 12], the Chinese remainder theorem [7.134], and other methods.

Encoding Low-Density Parity-Check Codes

In general, LDPC codes can be treated as any other block code when it comes to encoding. The generator matrix $\mathbf{G}$ of the LDPC code must be orthogonal to $\mathbf{H}$, i.e., $\mathbf{G} \cdot \mathbf{H}^\top = \mathbf{0}$. If $\mathbf{H}$ can be rearranged such that $\mathbf{H} = \left(\mathbf{P}^\top \ \mathbf{I}_{n-k}\right)$, then $\mathbf{G} = \left(\mathbf{I}_k \ \mathbf{P}\right)$.

Using Gauss–Jordan elimination, any $\mathbf{H}$ can in principle be converted to $\mathbf{H}' = \left(\bar{\mathbf{H}}^\top \ \mathbf{I}_m\right)$ with $\bar{\mathbf{H}}^\top$ an $m \times k$ binary matrix and $\mathbf{I}_m$ the identity matrix of size $m \times m$. The generator matrix is then obtained as $\mathbf{G} = \left(\mathbf{I}_k \ \bar{\mathbf{H}}\right)$. However, there is a fundamental problem with this simple approach: $\mathbf{G}$ will most likely *not be sparse*; thus the encoding complexity and the storage requirements will be $O(n^2)$, which we initially wanted to avoid when introducing LDPC codes.

Fortunately, there is a method to directly encode LDPC codes from $\mathbf{H}$ [7.111, App. A] that allows for an encoding complexity that scales (almost) linearly with the code length n. Instead of finding a generator matrix $\mathbf{G}$ for a given parity-check matrix $\mathbf{H}$, the encoding is carried out directly from $\mathbf{H}$. In a first step, $\mathbf{H}$ is transformed into an approximate triangular form: Using *only row and column permutations*, we construct $\mathbf{H}'$ from $\mathbf{H}$ with

$$\mathbf{H}' = \begin{pmatrix} \mathbf{A} & \mathbf{B} & \mathbf{T} \\ \mathbf{C} & \mathbf{D} & \mathbf{E} \end{pmatrix}, \tag{7.61}$$

where $\mathbf{T}$ is a *lower triangular matrix* of size $(m-g) \times (m-g)$, $\mathbf{B}$ is of size $(m-g) \times g$, and $\mathbf{A}$ is of size $(m-g) \times k$ (if $\mathbf{H}'$ is full rank). The parameter g denotes the number of rows left in $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{E}$ and is called the *gap* of the approximate representation. The smaller g is,

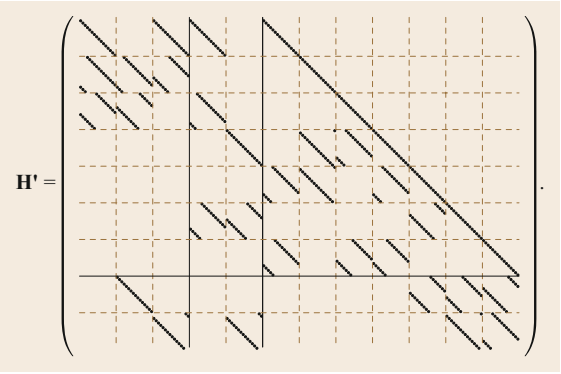


Fig. 7.14 Parity-check matrix of Fig. 7.13 after transformation into the form of (7.61)

the lower the encoding complexity will be. If the permutation operations are chosen well, we have $g \ll m$. A simple, greedy column- and row-swapping algorithm is presented in [7.111, App. A]. Applying this algorithm yields, for the running example, the matrix $\mathbf{H}'$ shown in Fig. 7.14.

Starting from $\mathbf{H}'$, we can clear $\mathbf{E}$ by Gaussian elimination, or equivalently

$$\begin{pmatrix} \mathbf{I}_{m-g} & \mathbf{0} \\ \mathbf{E}\mathbf{T}^{-1} & \mathbf{I}_g \end{pmatrix} \mathbf{H}' = \begin{pmatrix} \mathbf{A} & \mathbf{B} & \mathbf{T} \\ \widetilde{\mathbf{C}} & \widetilde{\mathbf{D}} & \mathbf{0} \end{pmatrix} \triangleq \widetilde{\mathbf{H}} \,, \tag{7.62}$$

with $\widetilde{\mathbf{C}} = \mathbf{E}\mathbf{T}^{-1}\mathbf{A} + \mathbf{C}$ and $\widetilde{\mathbf{D}} = \mathbf{E}\mathbf{T}^{-1}\mathbf{B} + \mathbf{D}$. Note that addition $+$ and subtraction $-$ are equivalent when operating over GF(2); hence, we can replace all subtractions by additions. Further note that inverting $\mathbf{T}$ is easy, as it is a lower triangular matrix [7.86, App. A]. Also, $\mathbf{T}^{-1}$ will retain its QC structure. Finally, the codeword $\boldsymbol{c}$ is divided into three parts

$$\boldsymbol{c} = (\boldsymbol{u} \;\; \boldsymbol{p}_1 \;\; \boldsymbol{p}_2) \,,$$

where $\boldsymbol{u}$ is the k-bit message, $\boldsymbol{p}_1$ holds the first g parity bits, and $\boldsymbol{p}_2$ the remaining $n-k-g$ parity bits. Vector $\boldsymbol{p}_1$ is then obtained from

$$\boldsymbol{p}_1^\top = \widetilde{\mathbf{D}}^{-1}\widetilde{\mathbf{C}}\boldsymbol{u}^\top \,. \tag{7.63}$$

Using back-substitution, $\boldsymbol{p}_2$ can be calculated as

$$\boldsymbol{p}_2^\top = \mathbf{T}^{-1}\left(\mathbf{A}\boldsymbol{u}^\top + \mathbf{B}\boldsymbol{p}_1^\top\right) = \mathbf{T}^{-1}\left(\mathbf{A} + \mathbf{B}\widetilde{\mathbf{D}}^{-1}\widetilde{\mathbf{C}}\right)\boldsymbol{u}^\top \,. \tag{7.64}$$

Although not obvious at first glance, this procedure has some complexity advantages. Many of the intermediate terms needed in the computation are *sparse* and can be precomputed and stored with low complexity. It can be shown that the total encoding complexity is $O(n+g^2)$, which means that if g is very small, the complexity approaches $O(n)$. Note that if the rank of the original matrix $\mathbf{H}$ is not m, some parity bits are linearly dependent, and the procedure must be modified slightly to accommodate this fact by either fixing the linearly dependent parity bits to some predetermined value or assigning additional information bits to them. For further details, see also [7.135].

Decoding Low-Density Parity-Check Codes

As the complexity of optimal MAP decoding grows exponentially with the number of information bits k, MAP decoding is usually not feasible for practical codes, and we need to resort to suboptimal decoding algorithms. The success of LDPC codes is largely due to the existence of very powerful suboptimal decoding algorithms that can be implemented with relatively low complexity. These decoding algorithms belong to the class of *message-passing* algorithms, which work by exchanging messages along the edges of the Tanner graph, with very simple operations to compute the messages inside the nodes. The BP algorithm is the most popular message-passing algorithm for decoding LDPC codes. It works by exchanging *beliefs* about the code bits along the edges of the Tanner graph. The BP algorithm is used in many neighboring fields, for instance in machine learning [7.101, 136, 137], but it was described already in the 1960s to decode LDPC codes [7.54].

We describe the decoder starting from the received, noisy codeword $\boldsymbol{y} = (y_1, \ldots, y_n)$. The goal of the decoder is to compute the bit-wise *a posteriori* probability based on the received codeword $\boldsymbol{y}$,

$$P_{C|\boldsymbol{Y}}(c_i = 1|\boldsymbol{y}) = 1 - P_{C|\boldsymbol{Y}}(c_i = 0|\boldsymbol{y}) \,, \tag{7.65}$$

which can be used to estimate the value of the bit c_i as $\hat{c}_i = \arg\max_{c_i \in \{0;1\}} P_{C|\boldsymbol{Y}}(c_i|\boldsymbol{y})$. Note that this rule is in slight contrast to the MAP rule in (7.38), which estimates the most probable codeword instead of the most probable bit. This rule is therefore also called the bit-wise MAP rule, and minimizes the bit error probability.

Most decoders for LDPC codes (which can include simplified versions such as binary message-passing decoders [7.138]) rely on the knowledge of the channel. The communication channel, or an equivalent communication channel comprising the physical channel as well as various inner receiver and signal processing stages, can be characterized by its *conditional probability mass function* $P_{Y|X}$, or equivalently $P_{Y|C}$, as described in Sect. 7.2. Continuous channels are characterized by their *conditional probability density function* $p_{Y|X}$. Here, we assume a binary-input AWGN channel with $x_i = (-1)^{c_i}$. However, the decoder can easily be

extended to other modulation formats using the respective bit-wise decoder (Sects. 7.2.1 and 7.6.3).

The beliefs that are passed by the BP decoder are conditional probabilities, but for numerical reasons, it is much more convenient to work in the logarithmic domain and use LLRs [7.139]. The use of LLRs also facilitates hardware implementation. The channel-related LLR l_i is defined as

$$l_i = \log \frac{p_{Y_i|C_i}(y_i|0)}{p_{Y_i|C_i}(y_i|1)} \,. \tag{7.66}$$

A simple computation reveals that for the AWGN channel with binary, antipodal ± 1 channel input ($x_i = (-1)^{c_i}$) and channel law given by (7.1), we have

$$l_i = \frac{2}{\sigma_n^2} y_i \triangleq L_c^{[\mathrm{AWGN}]}\, y_i \,. \tag{7.67}$$

Equation (7.67) means that the LLR l_i is obtained simply by multiplication of y_i with a constant $L_c^{[\mathrm{AWGN}]} \triangleq 2/\sigma_n^2$, which depends only on the noise variance. Estimating the noise variance at the receiver is an old problem, and many different solutions exist [7.140, 141]. In many cases, the true noise variance is not even necessary. For example, with the min-sum decoder introduced below, we can assume the noise variance to be constant, and hence the constant $L_c^{[\mathrm{AWGN}]}$ is predetermined and set to a value suitable for implementation. If higher-order modulation formats are used with BICM or PAS, we use the bit-wise decoder (see Sects. 7.6.3 and 7.6.5 for details) to compute the LLRs passed to the decoder.

A vast collection of different varieties of the BP message-passing decoding algorithm for LDPC codes exists. We refer the interested reader to [7.101, 142] for an in-depth treatment of these variations. In the following, we describe the two most popular variants: the flooding and the layered decoder.

The Flooding Sum–Product Decoder. The flooding sum-product decoder is perhaps the most widely used decoder for LDPC codes. It can be best explained using the Tanner graph representation of the code. The decoding algorithm itself is iterative, and each iteration can be split into two half-iterations. In the first half-iteration, each VN of degree d_v computes d_v outgoing messages, one for each of its outgoing edges. In the second half-iteration, each CN of degree d_c computes d_c messages, one for each of its outgoing edges. We repeat this procedure until a maximum number of iterations has been achieved or the bits have been successfully recovered.

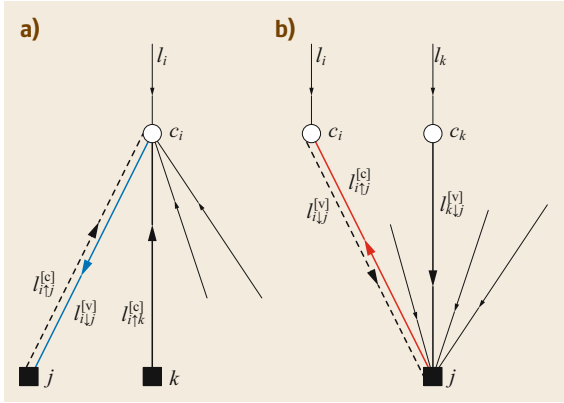


Fig. 7.15a,b Illustration of messages of the sum-product flooding decoding inside the Tanner graph. **(a)** VN perspective; **(b)** CN perspective

We denote by $l_{i\downarrow j}^{[v]}$ the message that is computed by VN i (corresponding to code bit c_i) and passed to CN j via the edge in the Tanner graph. Similarly, we denote by $l_{i\uparrow j}^{[c]}$ the message computed by CN j and passed (via the edge in the Tanner graph) to VN c_i. The superscript [v] or [c] describes the origin of the message (VN or CN), and the direction of the arrow ($\downarrow$ or $\uparrow$) denotes the *direction of the message flow* in the Tanner graph.

Figure 7.15 illustrates the flow of messages along the edges of the Tanner graph, from both a VN perspective (Fig. 7.15a) and a CN perspective (Fig. 7.15b).

We are now ready to state the sum-product algorithm:

1. *(Initialize)* For all $i \in \{1, \ldots, n\}$, set $l_{i\downarrow j}^{[v]} = l_i$, $\forall j \in \mathcal{N}(i)$, where l_i is calculated either using (7.67) (AWGN channel with binary inputs) or using the bit-wise decoder given by (7.93) in Sect. 7.6.3 (if higher-order modulation with BICM or PAS is used).
2. *(CN update)* For every CN j, with $j \in \{1, \ldots, m\}$, compute $|\mathcal{M}(j)| = d_c$ outgoing messages

$$l_{i\uparrow j}^{[c]} = 2 \tanh^{-1}\left[\prod_{i' \in \mathcal{M}(j)\setminus\{i\}} \tanh\left(\frac{l_{i'\downarrow j}^{[v]}}{2}\right)\right], \quad \forall i \in \mathcal{M}(j) \,. \tag{7.68}$$

This update equation follows from the probability that a single parity-check equation is fulfilled when the input bits are 0 or 1 with a certain probability and the conversion of probabilities to LLRs. A detailed derivation can be found for instance in [7.101, Sect. 5.4.3].

3. *(VN update)* For every VN c_i, with $i \in \{1, \ldots, n\}$, compute $|\mathcal{N}(i)|$ outgoing messages

$$l_{i\downarrow j}^{[\mathrm{v}]} = l_i + \sum_{j' \in \mathcal{N}(i)\backslash\{j\}} l_{i\uparrow j'}^{[\mathrm{c}]}, \qquad \forall j \in \mathcal{N}(i)\,. \quad (7.69)$$

4. *(A posteriori estimate)* For $i \in \{1, \ldots, n\}$, compute

$$\hat{c}_i = \begin{cases} 1 & \text{if } l_i + \sum_{j \in \mathcal{N}(i)} l_{i\uparrow j}^{[\mathrm{c}]} < 0, \\ 0 & \text{otherwise.} \end{cases} \quad (7.70)$$

If $\mathbf{H}\hat{\boldsymbol{c}}^{\top} = \mathbf{0}$, with $\hat{\boldsymbol{c}} = (\hat{c}_1, \ldots, \hat{c}_n)$ or if the maximum number of iterations is reached, then stop; otherwise go to step 2.

The notation $\mathcal{N}(i)\backslash\{j\}$ denotes all the elements of $\mathcal{N}(i)$ except j. For the example code of (7.52), we have $\mathcal{N}(2)\backslash\{4\} = \{1, 7\}$.

A straightforward hardware implementation of this algorithm is challenging due to the nature of (7.68). Hence, we will resort to simplified approximations to this update rule that only slightly affects the performance of the decoder. One particularly widely used approximation is the scaled min-sum rule [7.143], which is expressed as

$$l_{i\uparrow j}^{[\mathrm{c}]} = \alpha \left[\prod_{i' \in \mathcal{M}(j)\backslash\{i\}} \operatorname{sign}\left(l_{i'\downarrow j}^{[\mathrm{v}]}\right) \right] \min_{i' \in \mathcal{M}(j)\backslash\{i\}} \left| l_{i'\downarrow j}^{[\mathrm{v}]} \right|, \quad (7.71)$$

where α is an appropriately chosen scaling factor, which is usually determined offline and which depends on the code. The factor may also be iteration dependent and change during decoding [7.144]. An overview of other approximations and variants is presented in [7.145], [7.101, Sect. 5.5]. The theoretical foundations of this decoding algorithm are beyond the scope of this book. We refer the interested reader to [7.111] for an in-depth treatment of the underlying theory.

The Layered Decoder. In what follows, we describe the row-layered variant of the flooding sum-product decoder [7.146]. This variant is optimized for fast convergence and low complexity, and therefore is often used in optical communications. The row-layered decoder can be described directly from the parity-check matrix.

In a first step, the row-layered decoder copies the received LLRs l_j to a memory $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ of size n with $\mu_j = l_j$, which is continuously updated by the decoder. This memory will be referred to as *a posteriori memory*, as it comprises estimates of the bit-wise a posteriori probabilities. Additionally, the decoder requires an *extrinsic* memory consisting of $m \times d_{\mathrm{c}}$ entries (for a regular code, $m \sum_i i\delta_{\mathrm{c},i}$ for an irregular code), corresponding to the nonzero elements of $\mathbf{H}$. The memory cell $e_{j,i}$ corresponds to row j and column i of $\mathbf{H}$ and is only defined when $h_{j,i} = 1$. We initialize $e_{j,i} = 0, \forall j \in \{1, \ldots, m\}, i \in \mathcal{M}(j)$.

The row-layered decoding operation sequentially carries out three steps for each of the m rows of the parity-check matrix $\mathbf{H}$. After all m rows have been processed, a single decoding iteration has been carried out, and the decoder may restart with the first row to perform an additional iteration. Next we describe the steps for row j of $\mathbf{H}$:

1. *(Computation of input values)* Compute $|\mathcal{M}(j)|$ temporary values $t_{j,i} = \mu_i - e_{j,i}$, $\forall i \in \mathcal{M}(j)$, for all nonzero entries (indexed by i) of the j-th row of $\mathbf{H}$.
2. *(CN update)* The parity-check update rule is carried out to compute new extrinsic information using the full update rule

$$e_{j,i} = 2 \tanh^{-1}\left[\prod_{i' \in \mathcal{M}(j)\backslash\{i\}} \tanh\left(\frac{t_{j,i'}}{2}\right) \right], \ \forall i \in \mathcal{M}(j)\,, \quad (7.72)$$

where the product is again taken over all entries in $\mathcal{M}(j)$ except the one under consideration, i. Alternatively, e.g., for a hardware implementation, the scaled min-sum update rule

$$e_{j,i} = \alpha \left[\prod_{i' \in \mathcal{M}(j)\backslash\{i\}} \operatorname{sign}(t_{j,i'}) \right] \min_{i' \in \mathcal{M}(j)\backslash\{i\}} |t_{j,i'}|\,. \quad (7.73)$$

can also be used.
3. *(A posteriori update)* The $|\mathcal{M}(j)|$ a posteriori memory positions that take part in the parity-check equation are updated according to $\mu_i = t_{j,i} + e_{j,i}$, $\forall i \in \mathcal{M}(j)$.

After having carried out the three steps for row j, the LDPC decoder continues with decoding row $j+1$ or, if $j = m$ has been reached, it restarts either at $j = 1$ or terminates if the number of iterations has reached a predefined limit. After termination, we can get an estimate of the codeword by evaluating for every code bit

$$\hat{c}_i = \frac{1}{2}\left[1 - \operatorname{sign}(\mu_i)\right]. \quad (7.74)$$

If the code is systematic, the information portion can be recovered by taking the decision at the respective systematic positions.

Because of its sequential updates, this layered decoding algorithm is not immediately suited for massively parallel decoding of arbitrary LDPC codes. However, when used together with QC LDPC codes, we can see that groups of S consecutive rows of $\mathbf{H}$ are always non-interacting. In a decoder, the respective memory accesses are all to different locations, and hence the operations are non-interacting and can be parallelized. The cyclic shifts can be easily realized using barrel shifter circuits. See for instance [7.147] for further details.

7.4.2 Spatially Coupled Low-Density Parity-Check Codes

While so far we have focused mostly on block codes (e.g., LDPC codes), a significant number of practical codes were and still are based on simple convolutional codes using simple binary shift registers. About two decades ago, LDPC convolutional codes were introduced [7.71], which superimpose LDPC codes with a convolutional structure. The initial goal was to combine the advantages of convolutional and LDPC codes, and the resulting codes were called *LDPC convolutional codes*. A similar idea had already been pursued in the late 1950s and early 1960s with the concept of recurrent codes ([7.148] and references therein). The full potential of LDPC convolutional codes was not realized until a few years later, when *Lentmaier* et al. noticed that the performance of some LDPC convolutional code constructions asymptotically approached the performance of optimal ML decoding [7.72, 74]. Soon thereafter, this was formally proven for a wide range of channels by *Shrinivas Kudekar* et al. [7.75, 76]. As a proof technique, Kudekar et al. modified the original LDPC convolutional codes slightly and introduced a variation of the construction, denoted as spatially coupled codes. Because of the similarity between the two schemes, we use the term SC-LDPC codes to refer to both constructions for the remainder of this chapter. An alternative proof technique to the one in [7.75, 76] was proposed in [7.149, 150]. The analysis of nonbinary SC-LDPC codes was addressed in [7.151, 152].

Definition and Basic Construction

We now extend our description of LDPC codes to SC-LDPC codes. We first take on an encoder perspective. An LDPC encoder is a block encoder, i.e., it encodes a block of information bits $\boldsymbol{u}_t$ into a codeword $\boldsymbol{c}_t$. This is shown in Fig. 7.16a. The encoder of an SC-LDPC code can be interpreted as an encoder of a recurrent code [7.148] and is shown in Fig. 7.16b. The recurrent

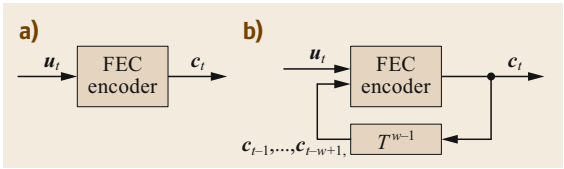


Fig. 7.16a,b Comparison of (a) conventional block LDPC encoder and (b) spatially coupled (SC) LDPC encoder

encoder generates *batches* $\boldsymbol{c}_t$ of a codeword which consists of a sequence of such batches. The batch $\boldsymbol{c}_t$ at time t depends not only on $\boldsymbol{u}_t$ but also on $w-1$ previous batches $\boldsymbol{c}_{t-1}, \ldots, \boldsymbol{c}_{t-w+1}$. The parameter w is called the *coupling width* of the code. The term *spatial coupling* highlights the dependence of $\boldsymbol{c}_t$ on previous batches along the *spatial* dimension t (this new, additional dimension was referred to as *spatial* dimension in [7.75]; however, note that it does *not* refer to a physical spatial dimension). Each batch is hence also called a *spatial position*.

At the beginning of the transmission (for $t = 1$), we assume that the contents of the feedback memory are zero, i.e., the previous batches at undefined times $t < 1$ are $\boldsymbol{c}_0 = \boldsymbol{c}_{-1} = \boldsymbol{c}_{-2} = \ldots = \mathbf{0}$. After starting the transmission, we could in principle run the encoding procedure without interrupting it. This is sometimes called *left-termination* or *unterminated* operation. Frequently, however, the encoding is terminated after the transmission of L batches, where L denotes the *replication factor* or *spatial length* of the code. Termination means that $\boldsymbol{c}_t = \mathbf{0}$ for $t > L$, provided that the input $\boldsymbol{u}_t = \mathbf{0}$ for $t > L$. Termination yields a block code again, where the length n of the block code amounts to Ln_b bits, with n_b being the size of one batch $\boldsymbol{c}_t$. Termination is realized by slightly reducing the number of information bits to $k'_\mathrm{b} < k_\mathrm{b}$ in the final $w-1$ batches, i.e., $\boldsymbol{u}_{L-w+2}, \ldots, \boldsymbol{u}_L$. The encoder then generates $k_\mathrm{b} - k'_\mathrm{b}$ extra parity bits that are required to enforce the termination condition. These extra parity bits lead to a *rate loss* (i.e., reduce the effective number of information bits). However, note that the effect of the rate loss becomes negligible if L is increased. Termination is often used to fit multiple FEC frames into a larger super-frame (e.g., an optical transport network (OTN) frame [7.153]) or a networking container [7.154], or to prevent error propagation.

Instead of interpreting the terminated encoding procedure as a recurrent procedure of L batches, it is often customary to group the L batches $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_L$ into the codeword $\boldsymbol{c}^{[\mathrm{SC}]} \triangleq (\boldsymbol{c}_1 \quad \boldsymbol{c}_2 \quad \ldots \quad \boldsymbol{c}_L)$. We denote the parity-check matrix of the corresponding *terminated* (and non-time-varying) SC-LDPC code with coupling

width w by $\mathbf{H}^{[\mathrm{SC},w]}$. It is given by

$$\mathbf{H}^{[\mathrm{SC},w]} = \begin{pmatrix} \mathbf{H}_1 & & & \\ \vdots & \mathbf{H}_1 & & \\ \mathbf{H}_w & \vdots & \ddots & \\ & \mathbf{H}_w & \ddots & \mathbf{H}_1 \\ & & \ddots & \vdots \\ & & & \mathbf{H}_w \end{pmatrix}, \tag{7.75}$$

with sub-matrices $\mathbf{H}_i$ of dimensions $m_\mathrm{b} \times n_\mathrm{b}$, with n_b denoting the length of one batch. The simplest case is obtained for $w = 2$ (which we sometimes call *unit-memory* coupling), for which we have

$$\mathbf{H}^{[\mathrm{SC},2]} = \begin{pmatrix} \mathbf{H}_1 & & & \\ \mathbf{H}_2 & \mathbf{H}_1 & & \\ & \mathbf{H}_2 & \ddots & \\ & & \ddots & \mathbf{H}_1 \\ & & & \mathbf{H}_2 \end{pmatrix}. \tag{7.76}$$

The parity-check matrix $\mathbf{H}^{[\mathrm{SC},w]}$ is of dimensions $(L + w - 1)m_\mathrm{b} \times Ln_\mathrm{b}$, and consequently, assuming full rank, the rate of the terminated SC-LDPC code is given by

$$R = 1 - \left(1 + \frac{w-1}{L}\right)\frac{m_\mathrm{b}}{n_\mathrm{b}} \leq 1 - \frac{m_\mathrm{b}}{n_\mathrm{b}}, \tag{7.77}$$

i.e., the rate is decreased by spatial coupling due to the extra rows in $\mathbf{H}^{[\mathrm{SC},w]}$ required for the termination. Note that this *rate loss* vanishes when $L \to \infty$, which is why we usually want to select L that is quite large. Multiple techniques for reducing the rate loss have been investigated [7.155–158].

Tanner Graph and Construction of Spatially Coupled Low-Density Parity-Check Codes

Due to the structured parity-check matrix $\mathbf{H}^{[\mathrm{SC},w]}$, the Tanner graph of an SC-LDPC code has a particularly simple structure. An SC-LDPC code and the corresponding Tanner graph can be constructed starting from an LDPC code. We show the construction of both the parity-check matrix $\mathbf{H}^{[\mathrm{SC},w]}$ and the Tanner graph by means of an example. We use the *edge-spreading* technique introduced in [7.72], which can be carried out directly in the Tanner graph of an LDPC code.

We start with a Tanner graph of an LDPC code. The first step consists in *coloring* the edges of the graph in w distinct colors or *edge types*. One possibility for coloring the edges is to carry out a random experiment for each edge and assign to it color w with probability $1/w$. This is called *uniform coupling*. The coloring probabilities can also be optimized, for example, to improve the performance of the code [7.158]. The coloring procedure is illustrated in Fig. 7.17. We start with the Tanner graph of a $(3, 6)$ LDPC code which is shown in Fig. 7.17a. We then apply the edge coloring where the two colors are represented by solid and dashed lines, respectively. This is shown in Fig. 7.17b. For each of the two edge colors, we can construct a new Tanner graph by removing all edges of different colors, and from this graph obtain the corresponding parity-check matrices $\mathbf{H}_1$ and $\mathbf{H}_2$. For the example in Fig. 7.17b with $w = 2$, we obtain

$$\mathbf{H}_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \quad \text{and}$$

$$\mathbf{H}_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{7.78}$$

Hence, by this technique, we have decomposed the matrix $\mathbf{H}$ into w *binary* matrices $\mathbf{H}_i$ of same dimension such that

$$\sum_{i=1}^{w} \mathbf{H}_i = \mathbf{H},$$

where the sum is taken over the natural numbers. We can also describe the edge-spreading technique in the following way: for every column i of $\mathbf{H}$, partition (either randomly or according to some deterministic procedure) the set $\mathcal{N}(i)$ into w pairwise disjoint subsets $\mathcal{N}_j(i)$ such that $\bigcup_{j=1}^{w} \mathcal{N}_j(i) = \mathcal{N}(i)$. Then use $\mathcal{N}_j(i)$ to set the *1* values of $\mathbf{H}_j$.

In the next step, we can construct the Tanner graph of the SC-LDPC code. For this, we start from our basic LDPC graph, and we first generate $w - 1$ additional copies of the VNs. This gives us w groups of VNs, each corresponding to one of the w colors. In the next step, we assign the edges of a single color to the corresponding group of VNs. For the colored graph in Fig. 7.17b, this leads to the graph shown in Fig. 7.18. Note that this graph corresponds to the parity-check matrix $(\mathbf{H}_2 \ \mathbf{H}_1)$. The Tanner graph of the final spatially coupled code with $w = 2$ is obtained by superimposing $L + 1$ of these Tanner graphs with an overlap of one spatial position (and combining the superimposed VNs into one new

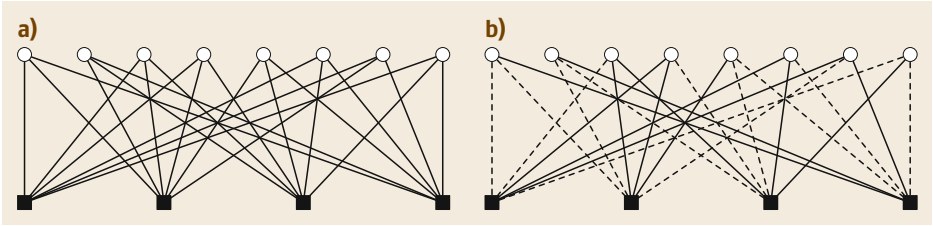


Fig. 7.17a,b Illustration of the edge-coloring principle for designing an SC-LDPC code from an LDPC code. **(a)** Tanner graph of a $(3, 6)$ LDPC code; **(b)** Tanner graph with edge coloring

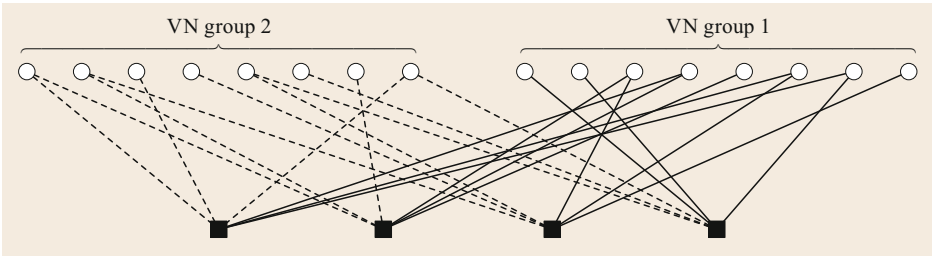


Fig. 7.18 Tanner graph representation of the example code with parity-check matrix $(\mathbf{H}_2 \ \mathbf{H}_1)$ of (7.78)

VN) and then removing the first and last n VNs together with their associated edges. For $L = 4$, this is illustrated in Fig. 7.19. We can clearly see the repetitive structure of the code. In the general case (i.e., for any $w \geq 2$), we superimpose $L + w - 1$ of the Tanner graphs corresponding to $(\mathbf{H}_w \ \cdots \ \mathbf{H}_2 \ \mathbf{H}_1)$ with an overlap of $w - 1$ spatial positions and then remove the first and last $(w - 1)n$ VNs with their associated edges.

Because of the edge-spreading technique, we have conserved $\boldsymbol{\delta}_\mathrm{v}$, i.e., the VN degree distribution (as well as $\boldsymbol{\delta}_\mathrm{c}$). The number of *1*s in each row of $\mathbf{H}^{[\mathrm{SC},w]}$ is the same everywhere as in the starting matrix $\mathbf{H}$, except in rows 1 to $(w - 1)m$ and $(L + 1)m$ to $(L + w - 1)m$, where we have lower degrees. These correspond to the lower-degree CNs at the boundaries of the Tanner graph (Fig. 7.19). This lack of *1*s, which is due to termination, is crucial for the outstanding performance of SC-LDPC codes [7.75, 159]. This leads to low-degree parity-check equations at the boundaries, which better protect the code bits at the boundaries. These, in turn, after decoding, protect their neighbors in a ripple- or wave-like fashion.

Encoding Spatially Coupled Low-Density Parity-Check Codes

As an SC-LDPC code is in principle an LDPC code with a structured parity-check matrix, the technique introduced in Sect. 7.4.1, can be directly applied to $\mathbf{H}^{[\mathrm{SC},w]}$. However, as the length of $\mathbf{H}^{[\mathrm{SC},w]}$ is usually relatively long, it can be advantageous to take on the recurrent coding perspective of Fig. 7.16b. In this case, we merely apply the triangularization to $\mathbf{H}_1$ and then encode the different $\boldsymbol{c}_t$s on a batch-by-batch basis. The encoding procedure can be simplified significantly by slightly modifying the construction procedure and enforcing that $\mathbf{H}_1$ has the structure $\mathbf{H}_1 \triangleq (\tilde{\mathbf{H}}_1 \ \mathbf{I}_m)$, where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix. In this case, systematic encoding (see Definition 7.8) of $\boldsymbol{u}_t$ into $\boldsymbol{c}_t = (\boldsymbol{u}_t \ \boldsymbol{p}_t)$ is accomplished by computing

$$\boldsymbol{p}_t^\top = \tilde{\mathbf{H}}_1 \boldsymbol{u}_t^\top + \sum_{i=2}^{w} \mathbf{H}_i \boldsymbol{c}_{t-i+1}^\top \tag{7.79}$$

during the first $L - w + 1$ batches with the initialization $\boldsymbol{c}_0 = \boldsymbol{c}_{-1} = \cdots = \mathbf{0}$. The last batch(es) require special

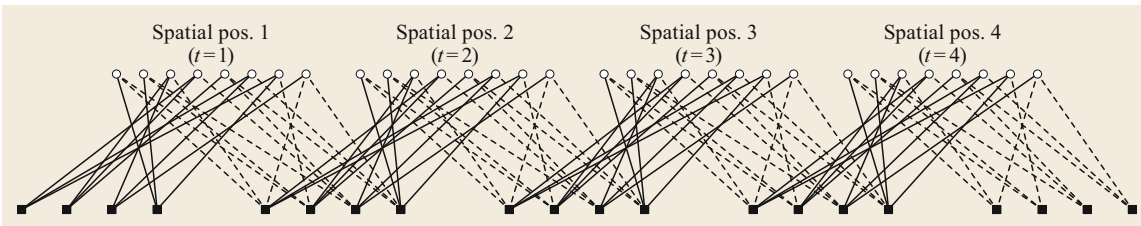


Fig. 7.19 Tanner graph representation of the example code with parity-check matrix $(\mathbf{H}_2 \ \mathbf{H}_1)$ of (7.78), with $L = 4$

treatment to fulfill the termination constraints, such that $\mathbf{H}^{[\mathrm{SC},w]}(\boldsymbol{c}^{[\mathrm{SC}]})^\top = 0$. We require up to $m(w-1)$ extra parity bits to fulfill these constraints. These are usually accommodated in the last batch or spread over the final $w-1$ batches, at a cost of a smaller number of information bits in these batches. A detailed, in-depth treatment of the encoder and related hardware architecture can be found in [7.71, 160, 161].

Decoding Spatially Coupled Low-Density Parity-Check Codes

In principle, an SC-LDPC code is merely an LDPC code with a parity-check matrix $\mathbf{H}$ structured in a particular way. Hence, decoding SC-LDPC codes is possible in a straightforward manner using the decoders given in Sect. 7.4.1. However, the decoding can be greatly simplified by exploiting the structure of the parity-check matrix and the decoding mechanics of such codes using a so-called windowed decoder, which operates on only a part of the codeword [7.77]. In the following, we first describe the conventional layered windowed decoder and then an extension that can improve the decoding performance without increasing the decoding latency.

Conventional Windowed Decoder. The conventional windowed decoder extends the flooding or layered decoders presented in Sect. 7.4.1 by some extra data flow and message handling to account for the memory of the code. We can define a decoding window size W_D and extract a portion $\mathbf{H}_\mathrm{W}$ of $m_\mathrm{b}W_\mathrm{D}$ rows of $\mathbf{H}^{[\mathrm{SC},w]}$ (with the corresponding nonzero columns) of size $\dim(\mathbf{H}_\mathrm{W}) = m_\mathrm{b}W_\mathrm{D} \times n_\mathrm{b}(W_\mathrm{D}+w-1)$ that has the form

$$\mathbf{H}_\mathrm{W} \triangleq \begin{pmatrix} \mathbf{H}_w & \mathbf{H}_{w-1} & \cdots & \mathbf{H}_1 & & \\ & \mathbf{H}_w & \mathbf{H}_{w-1} & \cdots & \mathbf{H}_1 & \\ & & \ddots & \ddots & \cdots & \ddots \\ & & & \mathbf{H}_w & \mathbf{H}_{w-1} & \cdots & \mathbf{H}_1 \end{pmatrix}. \tag{7.80}$$

We focus here only on the layered decoder, as it has most practical relevance due to its fast convergence. We initialize an a posteriori memory of size $n_\mathrm{b}(W_\mathrm{D}+w-1)$ as follows: $\mu_1 = \mu_2 = \cdots = \mu_{n_\mathrm{b}(W_\mathrm{D}+w-1)} = +\infty$. In practice, a sufficiently large constant (e.g., the largest value representable with the fixed-point representation) is used instead of $+\infty$. As in the layered decoder, we initialize an extrinsic memory as $e_{j,i} = 0$, $\forall j \in \{1,\ldots,m_\mathrm{b}W_\mathrm{D}\}$, $i \in \mathcal{M}_\mathrm{W}(j)$, where $\mathcal{M}_\mathrm{W}(j) = \{k : h_{\mathrm{W},j,k} \neq 0\}$ is the set of all nonzero entries of row j of $\mathbf{H}_\mathrm{W}$.

The windowed decoder then processes in each step a batch of n_b received values $\boldsymbol{y}_t$ that correspond to the transmitted batch $\boldsymbol{c}_t$, where t also denotes the spatial position. At the start of a codeword $\boldsymbol{c}^{[\mathrm{SC}]}$, we set $t = 1$. The windowed decoder is described as follows:

1. (*Receive and shift*) Upon reception of a new batch of n received values $\boldsymbol{y}_t = (y_{t,1},\ldots,y_{t,n})$ that correspond to the batch $\boldsymbol{c}_t$ of $\boldsymbol{c}^{[\mathrm{SC}]}$, shift the a posteriori memory by n positions and append the newly received LLRs $l_{t,i}$, i.e.,

$$\begin{aligned} &\mu_i \leftarrow \mu_{i+n_\mathrm{b}}, \quad \text{for } i \text{ from 1 to } (W_\mathrm{D}+w-2)n_\mathrm{b} \\ &\mu_{(W_\mathrm{D}+w-2)n_\mathrm{b}+i} \leftarrow l_{t,i}, \quad \text{for } i \text{ from 1 to } n_\mathrm{b}\,. \end{aligned}$$

   Note that the decoder must run even when $t > L$ until all batches of the codeword have been decoded. If $t > L$, there are no newly received values, and we define $l_{t,i} = +\infty$ for $t > L$ and $i \in \{1,\ldots,n_\mathrm{b}\}$. Similarly, we shift the extrinsic memory and initialize the unused part with zeros,

$$\begin{aligned} &e_{j,i} \leftarrow e_{j+m_\mathrm{b},i}, \quad \text{for } j \text{ from 1 to } (W_\mathrm{D}-1)m_\mathrm{b}, \forall i \\ &e_{(W_\mathrm{D}-1)m_\mathrm{b}+j,i} \leftarrow 0, \quad \text{for } j \text{ from 1 to } m_\mathrm{b}, \forall i\,. \end{aligned}$$

2. (*Decoding*) Run the layered decoder as described in Sect. 7.4.1, *Decoding Low-Density Parity-Check Codes* using $\mathbf{H}_\mathrm{W}$ as parity-check matrix for a predetermined number of iterations I with extrinsic memory $e_{j,i}$ and a posteriori memory μ_i.
3. (*Decision*) Recover the estimated code bits

$$\hat{c}_{t-W_\mathrm{D}-w+2,i} = \frac{1}{2}[1-\mathrm{sign}(\mu_i)], \quad \forall i \in \{1,\ldots,n_\mathrm{b}\} \tag{7.81}$$

   of the batch corresponding to $\boldsymbol{c}_{t-W_\mathrm{D}-w+2}$ of $\boldsymbol{c}^{[\mathrm{SC}]}$. Note the additional delay of $W_\mathrm{D}+w-2$ introduced by the windowed decoder, which adds extra latency to the decoding process.

Note that the number of effective iterations per spatial position is $I(W_\mathrm{D}+w-1)$, i.e., in order to compare SC-LDPC codes with conventional (block) LDPC codes, we need to choose the effective number of iterations accordingly. Also note that the windowed decoder introduces a decoding *delay* of $W_\mathrm{D}+w-2$ batches and hence potentially increases the decoding latency with respect to that of a conventional LDPC code, unless n_b is chosen to be significantly smaller than the length n of the LDPC code to compensate for this delay. Also note that the decoder will not output meaningful information in the first $W_\mathrm{D}+w-2$ batches and hence must run following the reception of all L transmitted batches for an additional $W_\mathrm{D}+w-2$ batches until all bits have been decoded. These extra executions must be addressed by

the decoder and require attention inside an FPGA or ASIC. However, as the windowed decoder essentially wraps around a conventional LDPC decoder, the effort of a very-large-scale integration VLSI implementation is potentially simple.

Multi-Engine Windowed Decoder. Usually $W_D > w$, and hence, as a decoding engine operates on some rows of $\mathbf{H}_W$, only some of the corresponding memory locations are updated, while a significantly large part of $\mathbf{H}_W$ is not considered due to the particular banded structure of $\mathbf{H}_W$. For example, all rows of $\mathbf{H}_W$ that are at least wm_b rows apart do not share common code bits and could hence be treated independently in parallel inside the decoder. For this reason we introduce a multi-engine layered decoder that consists of E engines that operate independently in parallel [7.83]. The only modification is step 2 of the windowed decoder. While the first engine processes the rows from 1 to $m_b W_D$ in that order, the second engine starts with an offset $O_2 \geq wm_b$ and operates on the sequence of rows $(O_2, \ldots m_b W_D, 1, \ldots, O_2 - 1)$, i.e., it wraps around the matrix. In general, engine i operates on the sequence of rows $(O_i, \ldots, m_b W_D, 1, \ldots, O_i - 1)$, and we define $O_1 = 1$. The offsets need to fulfill the following condition: $|O_i - O_j| \geq wm_b$, $\forall (i,j) \in \{1, \ldots, E\}^2$, $i \neq j$. By the pigeonhole principle, the number of engines is hence upper bounded as $E \leq \lfloor W_D / w \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x.

Optimization of Spatially Coupled Low-Density Parity-Check Codes

In contrast to LDPC codes, which are well understood and have now commonly accepted design rules, the design of high-performing SC-LDPC codes is still part of active research efforts. In this section, we highlight some of these activities and recent results that seem promising for the design and evaluation of future close-to-capacity SC-LDPC codes.

Despite their similarity to LDPC codes, SC-LDPC codes have significantly different error correction behavior. This behavior was rigorously analyzed in [7.75, 76, 159], where it was shown that decoding behaves like a *wave* that propagates through the codeword, which is also in line with the windowed decoding mechanism. To optimize codes during construction, knowing the speed of this wave is crucial. An analysis of the speed of this wave together with upper and lower bounds was provided in [7.162–164]. This analysis was used in [7.165] to show that we can design SC-LDPC codes with irregular VN degrees that lead to performance improvements when decoded with a complexity-constrained decoder.

While the asymptotic performance of SC-LDPC codes is now well understood, the influence of the various parameters on the practical, non-asymptotic decoding performance is much less clear. Some attempts to better understand the influence on the different parameters are given in [7.166, 167] which could help to design codes in the future. It turns out that for highly optimized codes, L should be small. In order not to reduce the rate given by (7.77) too much, w should be kept small as well, ideally $w \in \{2, 3\}$, to keep the rate loss and complexity and latency of windowed decoding small. Such codes have been called *weakly coupled* in [7.165]. To construct good codes in this case, *non-uniformly* coupled SC-LDPC codes were constructed in [7.158], which have both superior decoding performance and reduced rate loss.

While selecting a small coupling width w can have its advantages, the choice of a larger w together with very sparse matrices $\mathbf{H}_i$ can also be useful in some applications. In [7.165], the latter case was called *strongly coupled*. Although the windowed decoder needs to cover a potentially longer decoding window, this is compensated for by the fact that we can have a shorter n_b and that the sub-matrices $\mathbf{H}_i$ are very sparse, limiting the effective amount of operations to be done. Constructions for such strongly coupled codes, which closely resemble traditional convolutional codes with large syndrome-former memory, are given for instance in [7.168–170].

As already discussed, one of the main challenges for implementing SC-LDPC codes in practice is the rate loss due to termination. The rate loss, which scales with w, can be made arbitrarily small by increasing L; however, in some cases we want to avoid a large L, as it can worsen the finite-length performance of SC-LDPC codes [7.166]. Nevertheless, in some cases we can let L go to infinity and still achieve gains compared with LDPC codes, as we have demonstrated in [7.171]. For a fixed L, there exist multiple proposals for mitigating the rate loss [7.172, 173], which however often introduce extra structure at the boundaries. Such extra structure is usually undesired, as it complicates the decoder data flow. Other approaches use tail-biting codes (without termination) and mimic the termination (summoning the positive effects of termination) by shortening [7.155, 157, 174], energy shaping [7.175], using intelligent interleaving together with the properties of higher-order modulation formats [7.156, 176–178], or iterative demapping and equalization [7.179], just to name a few.

Another line of research is the optimization of the windowed decoder of SC-LDPC codes, which is itself not free of issues. Some of these issues are highlighted for instance in [7.166, 180]. In systems with heavy complexity constraints, the scheduling of the operations inside the decoder can lead to some performance im-

provements [7.181]. The windowed decoder itself can lead to unrecoverable burst errors, significantly degrading the system performance. These burst errors can be addressed as well by intelligent scheduling of the decoder operations [7.182]. A general introduction to SC codes and an overview of some further research problems is given in [7.183].

7.4.3 Polar Codes

In addition to spatially coupled codes, another coding scheme has emerged in the past few years that is asymptotically capacity-achieving and has appealing encoding and decoding complexity. This scheme, called *polar coding*, was invented by *Arıkan* in 2008 [7.65, 66] and was the first low-complexity coding scheme that could (provably) achieve the capacity over binary discrete memoryless symmetric channels (DMSCs), nonbinary DMSCs [7.184], and asymmetric channels [7.185]. In fact, polar codes had been proposed already in 2002 by *Norbert Stolte* [7.186] but did not receive much attention until Arıkan proved that they were capacity-achieving. A polar code encodes a vector $\tilde{\boldsymbol{u}}$ of length $n = 2^\eta$, containing the k information bits in specific, predetermined positions and (known) frozen values otherwise, by $\boldsymbol{c} = \tilde{\boldsymbol{u}}\mathbf{F} = \tilde{\boldsymbol{u}} \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)^{\otimes\eta}$, where $\otimes\eta$ denotes η successive applications of the Kronecker product. If the positions of the k information bits are carefully chosen, then after decoding with a *successive cancellation* decoder or variants thereof, these positions are unaffected by channel noise. More precisely, each component of $\tilde{\boldsymbol{u}}$ *sees* its own equivalent sub-channel, so that in total there are 2^η sub-channels. As η tends to infinity, the error probability of each sub-channel converges to zero (a *good* sub-channel) or one-half (a *bad* sub-channel). This phenomenon is referred to as *channel polarization*. The k information bits can be reliably transmitted over the k good sub-channels. Arıkan showed that the fraction of good sub-channels converges to the channel capacity under successive cancellation decoding. Figure 7.20 shows an example of a rate $R = 3/4$ polar encoder for a short toy code with $n = 8$ ($\eta = 3$ and hence $k = 6$). An offline procedure, as described for instance in [7.187], determines the $k = 6$ *good* sub-channels over which we transmit information, and the frozen bits, which we set to a predetermined value, e.g., *0*. Note that the codeword $\boldsymbol{c}$ is *not* systematic by default. Polar codes can, however, also be rendered systematic following the procedure given in [7.188].

The successive cancellation decoder of polar codes consists of relatively simple binary operations and is of complexity $O(n \log n)$. The elegance of the construction has made polar codes an attractive choice in research with many applications and scenarios, such as 5G. An early application to fiber-optic communications has been given in [7.82, 189]. However, there are multiple areas where polar codes can be further improved. Extremely long block lengths n are required to achieve low error probabilities with polar codes [7.66] and simple successive cancellation decoding. Recently, list decoding [7.67] in conjunction with a concatenated cyclic redundancy check CRC code was proposed to improve the error performance for finite code lengths, albeit at a cost of increased receiver complexity. Furthermore, the original successive cancellation decoder proposed by *Arıkan* [7.66] is not directly suited for parallelization, which is however mandatory if this class of codes is to be used in fiber-optic communications. Recent research efforts are aimed at studying new concepts for building hardware-efficient, parallelizable decoders [7.69, 70, 190, 191], which is still an active area of research. A promising line of research is the use of the sum-product algorithm developed for LDPC for polar codes, enabling efficient parallelization [7.192].

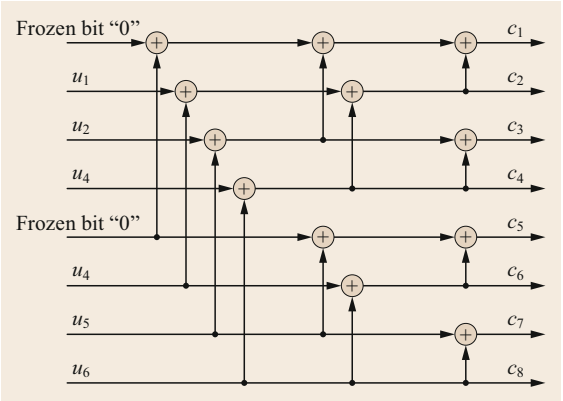


Fig. 7.20 Threefold application ($\eta = 3$) of the basic channel transform

7.4.4 Other Coding Schemes

In addition to LDPC, SC-LDPC, and polar codes, many other coding schemes exist for SDD. However, most of these have not found widespread application in optical communications. Perhaps the most prominent one is SDD of TPCs or concatenated BCH codes, which we will briefly discuss in the next section. In the 3G and 4G mobile wireless communication standards, turbo codes [7.193] (which are parallely concatenated codes based on convolutional component codes) are still the dominant error-correcting codes. However, they have found no application so far in optical communications. They suffer from the disadvantage of a non-negligible error floor and little to no available decoder implementations that can operate at the speeds required in optical

communications. In the upcoming 5G wireless communication standard, turbo codes will be at least partly superseded by LDPC codes (for some data channels) and polar codes (for some control channels) [7.68].

7.4.5 Performance Curves

Performance of Low-Density Parity-Check Codes

To show the potential of LDPC codes with the simple SDD algorithm described above, we generate two regular QC LDPC codes with parameters $d_v = 3$ and $d_v = 4$, and we fix d_c to obtain codes of rate $4/5$ (OH $= 25\%$). We generate small base matrices of size $n = 300$ using an algorithm similar to Gallager's algorithm and then use lifting with cyclically shifted identity matrices of size 128×128. We have selected the shifts to maximize the girth parameter of the code by avoiding *Fossorier*'s conditions for low-girth codes [7.132]. In Fig. 7.21, we show simulation results with the non-simplified CN update rule and the normalized min-sum decoding rule (with $\alpha = 3/4$) after transmission over a binary-input AWGN channel. We can see that the code with $d_v = 3$ outperforms the code with $d_v = 4$, and the performance loss due to scaled min-sum decoding is almost negligible. For $d_v = 4$, the performance loss due to scaled min-sum decoding is more pronounced, which is due to a mismatched choice of the scaling factor α. The increasing color scales in Fig. 7.21 indicate the number of iterations (1, 2, 3, 5, 10 and 30). We can see that after only 10 decoding iterations, we have achieved a very good performance, and increasing the number of iterations to 30 yields additional gains of only 0.1 dB, which may not be worth the effort.

Figure 7.22 illustrates the advantage of the layered decoder compared with the flooding decoder. We use the $(d_v = 3, d_c = 15)$ regular LDPC code at a channel quality of $E_b/N_0 = 2.9$ dB. We plot the post-FEC BER as a function of the iterations that are carried out inside the decoder. We can see that the BER decreases significantly faster in the case of the layered decoder, and when targeting a specific post-FEC BER, the layered decoder roughly halves the number of required iterations. Especially in high-speed fiber-optic communications, this can lead to dramatic reductions in VLSI complexity, as only a relatively small number of iterations need to be carried out. The convergence speed can be further improved by combining two layered decoders with different schedules, as highlighted in [7.194], at a cost of increased hardware complexity.

We consider as a second simulation example LDPC codes for transmission over the BSC. This case occurs for instance in systems without ADC at the receiver, using a simple thresholding device. Other examples are high-speed communications with different dedicated digital signal-processing (DSP) and FEC chips. In this case, the chip-to-chip interface may not be able to transfer the quasi-continuous decoder input LLRs to the FEC chip, and heavy quantization is necessary. Although the channel output undergoes a decision, we can still carry out SDD by passing *soft* messages inside the decoder and exploiting the channel statistics. Note that for the

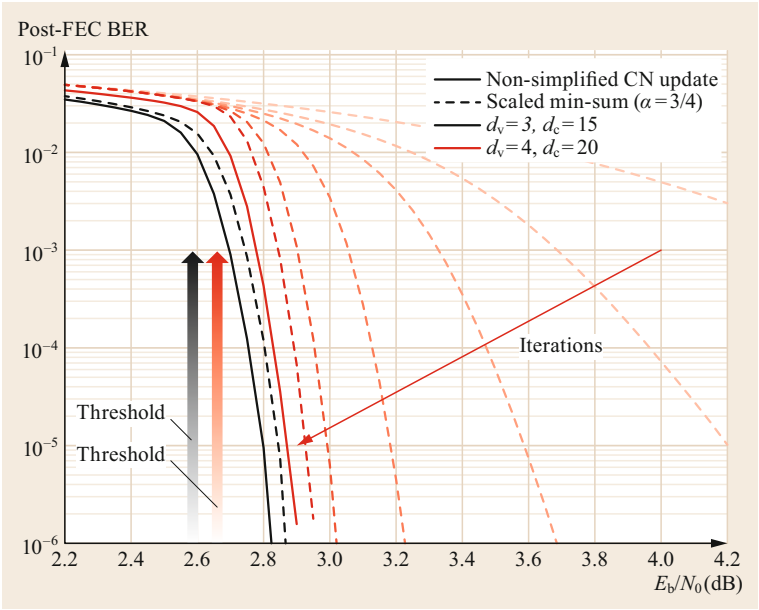


Fig. 7.21 Simulation of two regular LDPC codes of rate $R = 4/5$ (OH $= 25\%$) with parameters $(d_v = 3, d_c = 15)$ (*black lines*) and $(d_v = 4, d_c = 20)$ (*red lines*) with the full update rule (*solid lines*) and scaled min-sum (*dashed lines*) with layered decoding and 30 decoding iterations

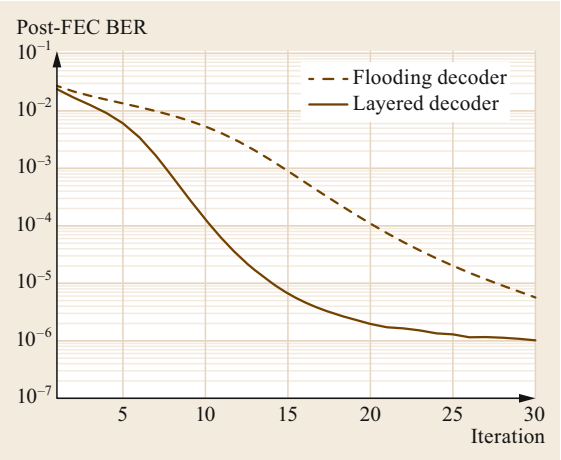


Fig. 7.22 Comparison of flooding decoder and layered decoder with min-sum update rule ($\alpha = 0.875$) for a regular LDPC code with parameters $d_v = 3$ and $d_c = 15$ ($R = 4/5$, 25% OH) at a channel quality of $E_b/N_0 = 2.9$ dB

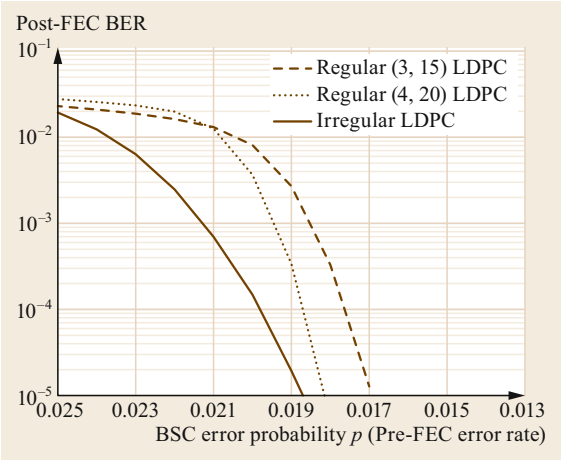


Fig. 7.23 Simulation of three LDPC codes of rate $R = 4/5$ (25% OH) with scaled min-sum update rule ($\alpha = 3/4$), layered decoding, and 30 decoding iterations

BSC, the achievable rates $R_{\text{SDD-BW}}$ and $R_{\text{HDD-BW}}$ coincide. We compute the channel-related LLRs as

$$l_i = \log \frac{p_{Y_i|X_i}(y_i|+1)}{p_{Y_i|X_i}(y_i|-1)} = y_i \log \frac{1-p}{p} \triangleq y_i L_c^{[\text{BSC}]}, \tag{7.82}$$

where we assume BPSK modulation $x_i = (-1)^{c_i}$ and received values $y_i \in \{\pm 1\}$. The error probability of the BSC is given by p and is known at the receiver. We consider a normalized min-sum decoder with $\alpha = 3/4$. Figure 7.23 shows simulation results for three different codes of rate $R = 4/5$: two regular LDPC codes with parameters $d_v \in \{3, 4\}$ and one irregular LDPC code with $\delta_{v,3} = 4/5$, $\delta_{v,10} = 1/5$, and $d_c = 22$ ($\delta_{c,22} = 1$). We keep $\delta_{v,2} = 0$ in order not to increase the error floor. We can see that in this case, the code with $d_v = 4$ outperforms the code with $d_v = 3$ (contrary to the case of the AWGN channel shown in Fig. 7.21). The irregular code enables further coding gains, especially at post-FEC BERs above 10^{-5}, but the slope of the BER curve is less steep than that of the BER for the two regular codes.

Comparison of Low-Density Parity-Check and Polar Codes

In [7.165], LDPC codes, SC-LDPC codes, and polar codes with successive cancellation decoding were compared, and it was concluded that as of today, SC-LDPC codes are the most promising for high-speed optical transponders. In Fig. 7.24, we compare the performance of a ($d_v = 3, d_c = 15$) regular LDPC code of length 38 400 and a polar code of length $n = 2^{15} = 32\,768$, both of rate $R = 4/5$. Note that we use the full update equation (7.72) for both cases to have a fair comparison. We use the successive cancellation list decoder described in [7.67] with list sizes of $L_S = 4$ and $L_S = 32$ and additionally concatenate the polar code with a short CRC (adding 32 parity bits), which is used to improve the selection of candidate codewords kept in the list. We can see that with list-based decoding, polar codes now have comparable performance as LDPC codes. However, the slope of the polar code is significantly less steep than that of the LDPC code. Hence, if low BERs of around 10^{-15} are targeted, LDPC codes will still show better performance. Polar codes may have an advantage for low-error-rate applications due to the absence of error floors under successive cancellation decoding [7.195]. In addition, note that the complexity of the list decoder, required for achieving good performance, scales linearly with the list size L_S. Furthermore, the baseline complexity of the decoder is significantly larger than that of an LDPC decoder. Hence, polar codes may become attractive in the future when highly parallel and efficient low-complexity decoders become available.

Performance of Spatially Coupled Low-Density Parity-Check Codes

In this section, we compare a few constructions of SC-LDPC codes. The possibilities for optimizing SC-LDPC codes are vast, and depending on the application, different setups might be useful. We again summarize the parameters of SC-LDPC codes and their influence on performance:

- The degree distribution will affect the performance. Although regular codes already achieve outstand-

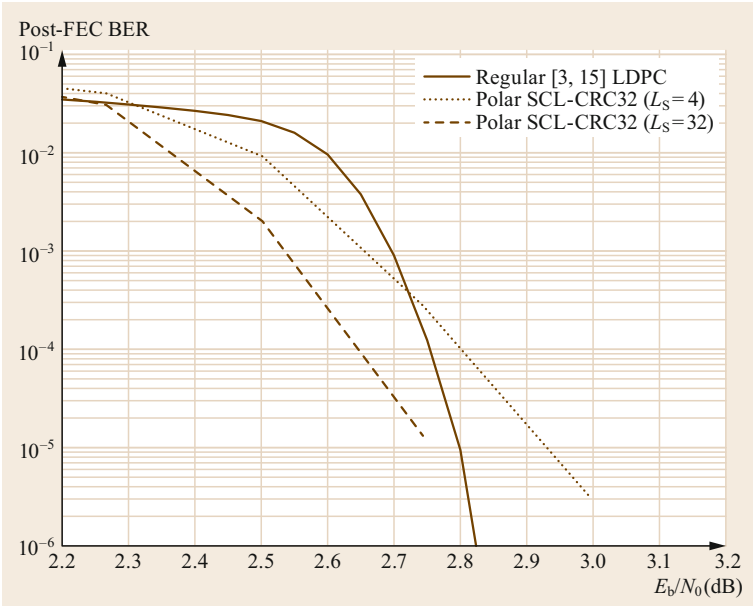


Fig. 7.24 Comparison of LDPC codes with polar codes with successive cancellation list decoding

ing performance, some improvements are expected with irregular codes, especially if the decoder complexity is constrained, as shown in [7.83].

- The coupling width w has a direct influence on the performance as well. In [7.83], a distinction between weakly coupled codes (small w, usually $w < 5$) and strongly coupled codes (large w, often $w > 5$) was noted. Both schemes may have their advantages and disadvantages.
- The replication factor L has an influence on the rate loss of the code and also an influence on the error probability of the windowed decoder as shown in [7.166, 180]. However, its influence is not as obvious as the influence of the other parameters.
- The parameters of the windowed decoder have a significant influence on the performance. The complexity of the windowed decoder equals $I\,(W_D + w - 1)$. However, when W_D is chosen that is too small, even a large number of iterations will not lead to good decoding performance. If W_D is too large, we may carry out too many unnecessary operations.

In the following, we will highlight some of these parameter trade-offs with simulation examples. In all simulations we fix $L = 100$. In a first example, we fix the (approximate) decoding complexity to $W_D\,I = 10$, corresponding to an equivalent of roughly 10 layered decoding iterations of conventional LDPC codes. We construct an SC-LDPC of rate $R = 4/5$ using the spreading technique explained in Sect. 7.4.2. We start from a QC-LDPC base matrix $\mathbf{A}$ of size 60×300 and fix the lifting parameter to $S = 32$. We apply the spreading construction to the base matrix. Hence, the sub-matrices $\mathbf{H}_i$ are of size 1920×9600. We show the simulation results in Fig. 7.25. For the configuration with $W_D = 5$ and $I = 2$ (Fig. 7.25a.), the code with $w = 2$ yields the best performance. When we switch the configuration to $W_D = 10$ and $I = 1$ (Fig. 7.25b), we can see that the code with $d_v = 4$ and $w = 4$ has the best performance and already outperforms the conventional LDPC code by about 0.1 dB. As the code with $d_v = 4$ gives superior performance, we will no longer consider the scheme with $d_v = 3$ in the following.

In Fig. 7.26, we consider the performance of codes with varying windowed decoder complexity. We consider complexities of $W_D\,I \in \{10, 20, 40, 80\}$ and compare the codes for different coupling factors w. We can see that the decoder performance does not saturate early, but keeps improving if we increase the complexity. We picked the window configuration among a few tested such that the coding gains are maximized. Furthermore, we see that, as observed already before, increasing w leads to significant performance improvements. In particular, the code with $w = 6$ improves about 0.5 dB with respect to the LDPC code of Fig. 7.24 at a BER of 10^{-6}, and approaches the achievable rate of the channel by about 0.3 dB. Considering that the slope of the BER curve is much steeper, the additional coding gain at a low BER of 10^{-15} is even larger. Hence, SC-LDPC codes are future-proof in the sense that we can use them today with low complexity, but with only small coding gains. In the future, when high-performing circuits are available, we can spend more complexity to achieve higher NCGs. In the next section, we look at the performance at low BERs.

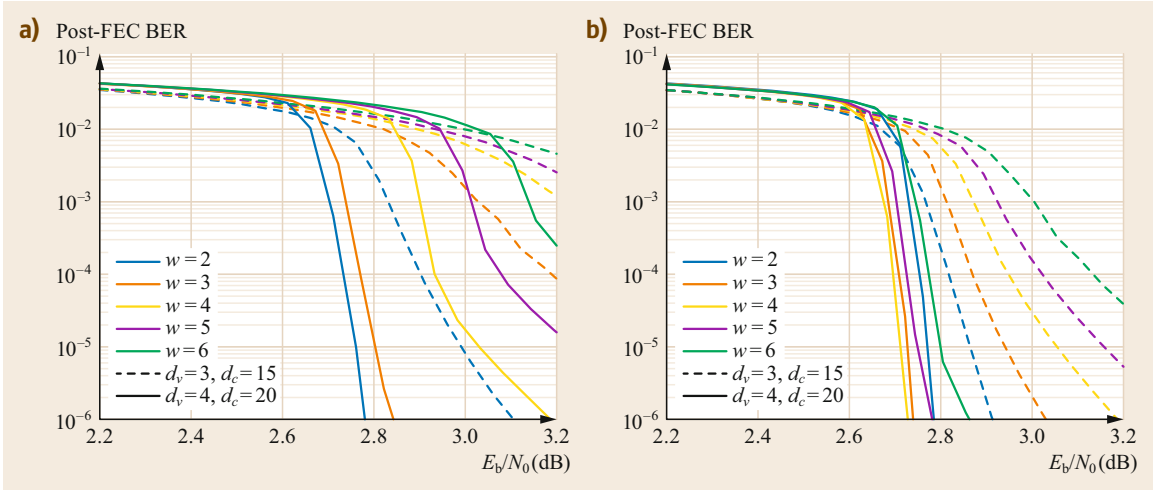


Fig. 7.25a,b Performance simulation of regular SC-LDPC codes of rate $R = 4/5$ with windowed decoding (full update rule) for two window configurations. **(a)** $W_D = 5$ and $I = 2$; **(b)** $W_D = 10$ and $I = 1$

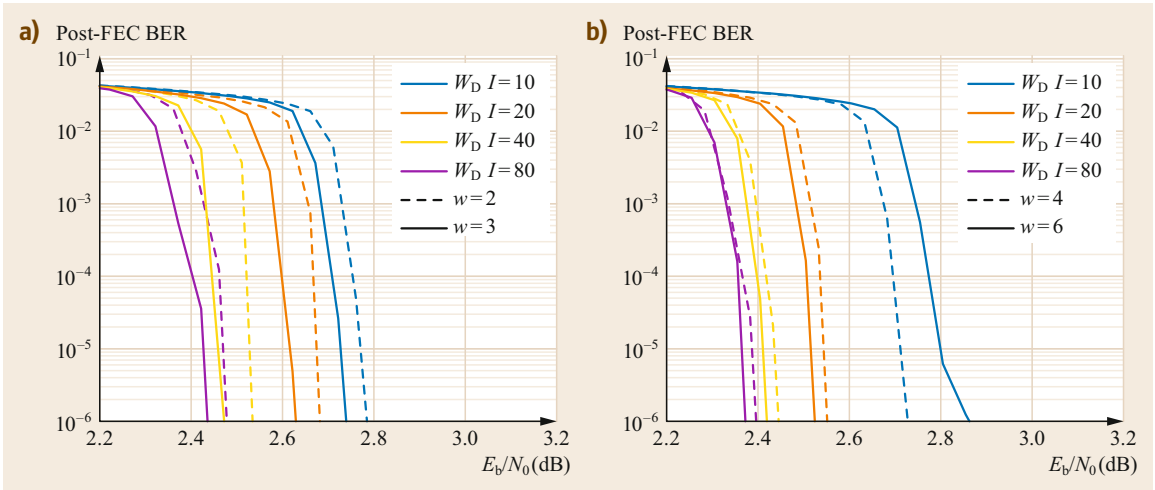


Fig. 7.26a,b Performance simulation of regular SC-LDPC codes of rate $R = 4/5$ with windowed decoding (full update rule) for four windowed decoding configurations of increasing complexity: $W_D = 10$ and $I = 1$, $W_D = 20$ and $I = 1$, $W_D = 20$ and $I = 2$, $W_D = 20$ and $I = 4$. **(a)** $w = 2$ and $w = 3$; **(b)** $w = 4$ and $w = 6$

FPGA-Based Evaluation of Decoding Performance of Spatially Coupled Low-Density Parity-Check Codes

A drawback of LDPC and SC-LDPC codes is that no analytic expression exists for the error performance, and the FEC designer must resort to Monte Carlo simulations. As optical communication systems usually require extremely low post-FEC error rates in the range of 10^{-15}, we require either extremely large compute clusters or FPGAs.

In order to verify the performance of some LDPC and SC-LDPC codes at low BERs, we use an FPGA platform, whose high-level diagram is illustrated in Fig. 7.27 and which is described in detail in [7.165]. This platform is similar to other platforms reported in the literature [7.84, 196, 197] and consists of three parts: a Gaussian noise generator, an FEC decoder, and an error-detecting circuit. The Gaussian noise generator is built upon a Tausworthe pseudo-random number generator with a sequence length of $\approx 10^{88}$, followed by a Box–Muller circuit transforming two uniformly distributed random numbers into two Gaussian-distributed numbers [7.198]. The output of the Gaussian noise generator is multiplied with a weighting factor in order to adjust N_0 and added to the all-zero codeword ($\boldsymbol{x} = (+1, \ldots, +1)$), yielding LLRs, which are then fed to

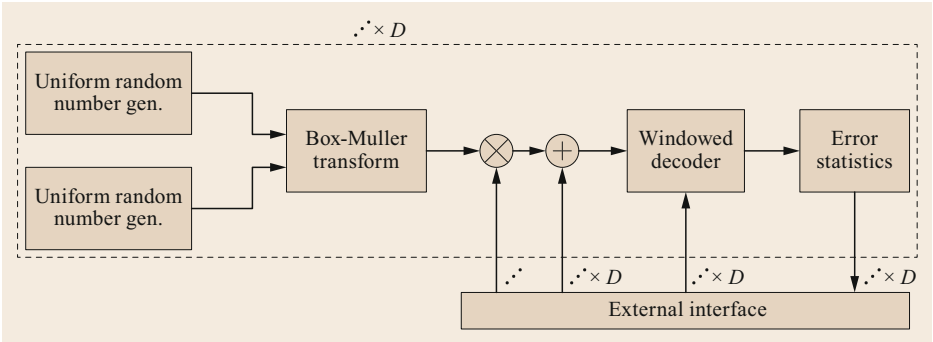


Fig. 7.27 High-level schematic of the FPGA evaluation platform

the LDPC decoder after quantization to 15 levels. The LDPC decoder is based on the layered decoding algorithm and uses a scaled min-sum CN computation rule.

The windowed decoder carries out I layered decoding iterations as described in Sect. 7.4.2. The effective number of iterations, i.e., the number of updates of each batch, thus corresponds to $I(W_\mathrm{D} + w - 1)$. For simplicity, we set $I = 1$ in most examples such that the window length corresponds roughly to the number of iterations in a conventional decoder. The output of the LDPC decoder is connected to the BER evaluation unit, which counts the bit errors and reports the error positions. A single Virtex-7 FPGA is able to emulate a block LDPC code with 12 decoding iterations at a rate of about 5 Gbit/s and an SC-LDPC code with $w = 3$, $W_\mathrm{D} = 13$, and a single iteration ($I = 1$) at approximately 1.5 Gbit/s. This lower rate is mainly due to increased memory requirements of the windowed decoder compared with the block decoder, which only requires a single a posteriori memory of size n, while the windowed decoder requires $(W_\mathrm{D} + w - 1)n_\mathrm{b}$ memory blocks. Note that in an ASIC implementation with an iteration-unrolled decoder, these numbers will be approximately equal.

The decoder we implemented uses a decoding window of size $W_\mathrm{D} = 13$, carries out a configurable number of iterations I, and can use either a single engine ($E = 1$) or two decoding engines ($E = 2$), in which case the offset of the second engine is $O_2 = 6m_\mathrm{b} + 1$. The decoder uses a scaled min-sum update rule with either $\alpha = 0.75$ or $\alpha = 0.875$ (depending on the code, different α will be better). We evaluate the performance of several coding schemes of rate $R = 4/5$ (OH = 25%). We consider a ($d_\mathrm{v} = 3, d_\mathrm{c} = 15$) regular LDPC code as reference. This code is a QC code constructed from cyclic permutation matrices of size 32×32 and has girth 10 and length $n = 31\,200$.

In addition to the block code, we also compare two SC-LDPC codes, where we select $L \to \infty$ to avoid any termination effects and to simplify the decoder data flow. The performance of such a code is an upper bound of the BER performance of a code with finite L, as decreasing L will only improve the BER of the code (note that it may not necessarily improve the NCG as the rate R of the code decreases and hence E_b/N_0 changes). We consider two codes in the simulation:

- □ SC-LDPC code A (marker □) is an irregular code with coupling width w, VN degree distribution $\boldsymbol{\delta}_\mathrm{v} = (0, 0, 4/5, 0, 0, 1/5)$, and regular $d_\mathrm{c} = 18$. This degree distribution has been optimized to maximize the decoding velocity at a target BER of 10^{-6}, as detailed in [7.165]. The batch size is $n_\mathrm{b} = 7500$ ($\dim(\mathbf{H}_i) = 1500 \times 7500$).
- ◊ SC-LDPC code B (marker ◊) is a regular code with $d_\mathrm{v} = 4$, $d_\mathrm{c} = 20$, coupling width $w = 2$, and batch size $n_\mathrm{b} = 7500$, with $m_\mathrm{b} = 1500$.

Both SC-LDPC codes are QC codes constructed from cyclic permutation matrices of size 32×32. Note that the batch size n_b of the spatially coupled codes is significantly smaller than the length n of the LDPC code, which may simplify the decoder implementation, especially the routing network.

The simulation results are shown in Fig. 7.28. As expected, the LDPC code performs quite well, albeit slightly worse than the code shown in Fig. 7.21. This is because here, the channel output and the messages inside the decoder have been quantized, which leads to a small performance penalty of ≈ 0.12 dB. First we look at the case $I = 1$ and $E = 1$. The LDPC code is decoded with $I = 13$ iterations, and hence the number of operations per bit is equivalent for both cases. We can see that both SC-LDPC codes outperform the LDPC code. SC-LDPC code A, which has been optimized for convergence speed, shows a very early start of the waterfall region. However, the slope of the BER curve starts to decrease slowly, such that the performance at low BERs is relatively poor and is expected to cross the BER curve of the block LDPC code. The situation is reversed for SC-LDPC code B, which exhibits a late start of the waterfall curve, which then drops exceptionally

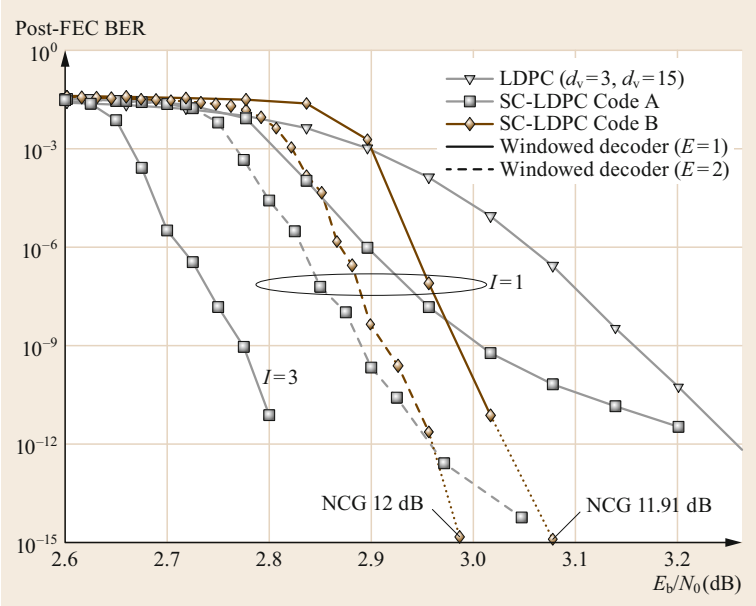


Fig. 7.28 Decoding results of the FPGA platform for two different codes and two setups of the decoder

quickly. We did not observe any error patterns for the next E_b/N_0 point (3.08 dB) of the curve after simulation of 7.9×10^{14} bits, so by extrapolating the BER curve, we may conjecture that SC-LDPC code B has an NCG of 11.91 dB at a post-FEC BER of 10^{-15}. This NCG is achievable with a modest number of $(W_D + w - 1) = 14$ equivalent iterations.

In order to show the potential of the proposed code, we increase the number of iterations I in step 2 of the windowed decoder of SC-LDPC code A from $I = 1$ to $I = 3$. This result is shown by the leftmost curve in Fig. 7.28. By extrapolating the curve, we conjecture an NCG of approximately 12.14 dB. If an error floor occurs at BERs lower than 10^{-13}–10^{-14}, it can usually be well compensated for by adequate post-processing techniques, as detailed in [7.196].

As increasing the number of iterations to $I = 3$ also increases the decoder latency, we keep $I = 1$ and increase the number of decoding engines to $E = 2$. We can see an improvement in the NCG by around 0.1 dB without an increase in latency (although a doubling of the number of iterations is required). We can also see that the slope of the BER curve of SC-LDPC code A becomes significantly steeper such that it becomes almost comparable in performance to SC-LDPC code B.

To reuse these codes in, for example, transmission experiments, we provide the most common thresholds in terms of MI/GMI and pre-FEC BER in Table 7.4. For a detailed discussion about threshold-based performance evaluation, we refer the reader to Sect. 7.7.

To summarize, we can say that SC-LDPC codes generally outperform conventional (block) LDPC codes. The performance can be further improved if the number of iterations is increased; hence such candidates can be used to build a future-proof communication standard. Although they are being used today with low decoding complexity, the systems could be upgraded in the future by a more complex decoder carrying out a higher number of iterations with increased NCG still guaranteeing backward compatibility.

In order to compare SC-LDPC and conventional LDPC codes, we show the potential of both codes by performing a simulation of regular codes with different rates and compute the extrapolated NCG at a target BER of 10^{-15}. For the SC-LDPC codes, we consider regular codes with $d_v = 4$ and we vary d_c to achieve different rates. Motivated by the results in Fig. 7.26, we select $w = 6$ and we carry out windowed decoding with different window configurations. We compare these codes with regular LDPC codes with $d_v = 3$,

Table 7.4 Common thresholds for the two SC-LDPC code B of rate $R = 4/5$ shown in Fig. 7.28

	GMI	Pre-FEC BER				
	$R_{\text{BW-SDD}}$	BPSK/QPSK	8-QAM	16-QAM	32-QAM	64-QAM
SC-LDPC Code B, $W_D = 13$, $I = 1$, $E = 1$	0.866	0.0358	0.0362	0.0363	0.0367	0.0372
SC-LDPC Code B, $W_D = 13$, $I = 1$, $E = 2$	0.862	0.0370	0.0375	0.0377	0.0379	0.0385
SC-LDPC Code A, $W_D = 13$, $I = 3$, $E = 1$	0.853	0.0387	0.0395	0.0399	0.0398	0.0405

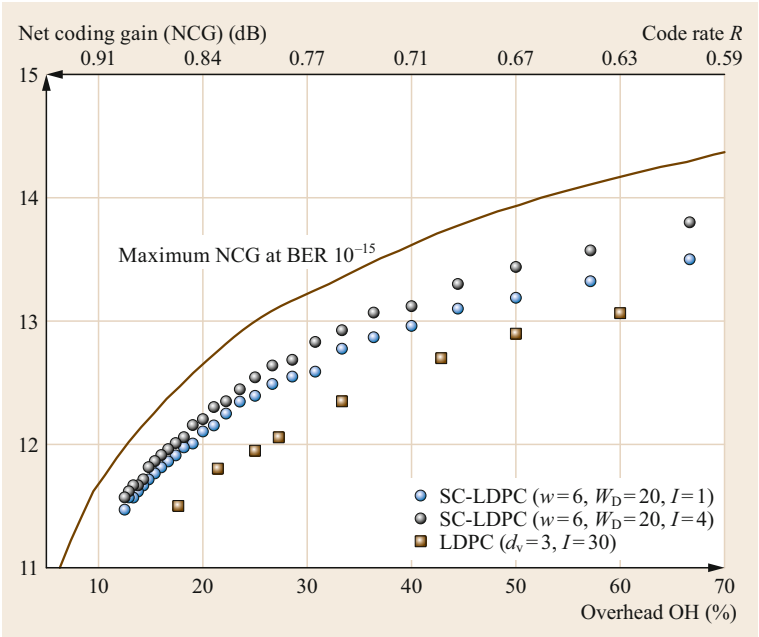


Fig. 7.29 Achievable NCG at a BER of 10^{-15} for several coding schemes

which are frequently used because of their good error floor properties. Figure 7.29 shows the achievable NCGs. We can clearly see that, if we can spend the complexity to run $I = 4$ iterations per windowed decoding step, we can nearly close the gap to the limit, and gains of around 0.8 dB are possible when compared with LDPC codes. However, already with $I = 1$, we can significantly outperform regular LDPC codes.

7.5 Hard-Decision Forward Error Correction

The adoption of SDD in fiber-optic communications represented a breakthrough with respect to the classical schemes based on algebraic codes (BCH and Reed–Solomon codes) and HDD. As we have discussed in the previous sections, LDPC codes and SDD provide very large net coding gains and achieve performance close to the Shannon limit. However, the implementation of BP decoding of LDPC codes still presents several challenges at very high data rates. This motivates the use of HDD for applications where complexity and throughput is a concern, since HDD can yield significantly reduced complexity and decoder data flow compared with SDD. HDD-based FEC is currently used in regional/metro optical transport networks [7.22] and for other applications such as optical data center interconnects [7.199]. Powerful code constructions for HDD date back to the 1950s, when *Elias* introduced PCs [7.18]. In recent years, the introduction of new code constructions such as staircase codes [7.49] and braided codes [7.50, 200], which can be seen as generalizations of PCs, and the link between these constructions and codes on graphs has fostered a revival of HDD for optical communications.

In this section, we focus on FEC codes for HDD that belong to the second and third generation of FEC, i.e., staircase codes and braided codes, as well as similar structures. These codes are built from algebraic block codes as component codes, usually BCH codes and Reed–Solomon codes for binary and nonbinary codes, respectively. They can be seen as concatenated codes and also fall within the category of codes on graphs or modern codes.

We start this section by briefly describing BCH and Reed–Solomon codes. We then review Elias' PCs and the underlying iterative decoding algorithm based on bounded distance decoding (BDD) of the component codes. We also discuss some interesting generalizations of PCs, including staircase codes and braided codes, and their connection to LDPC codes and SC-LDPC codes. Finally, we briefly discuss the link between iterative BDD (iBDD) of product-like codes and BP decoding and how to analyze the performance of product-like codes.

7.5.1 Classical Block Codes: Bose–Chaudhuri–Hocquenghem and Reed–Solomon Codes

BCH codes and Reed–Solomon codes are two of the most important classes of linear block codes. They are characterized by a rich algebraic structure that enables the use of efficient decoding algorithms that can be implemented with acceptable complexity.

Bose–Chaudhuri–Hocquenghem Codes

BCH codes, invented independently by *Alexis Hocquenghem* in 1959 [7.32] and by *Raj Chandra Bose* and *Dijen Ray-Chaudhuri* in 1960 [7.33], form a large class of cyclic codes. For short block lengths (up to a few hundred bits), BCH codes are among the best known codes of the same length and rate. For any pair of positive integers $\nu \geq 3$ and $t < 2^{\nu-1}$, there exists an $(n, k, d_{\min})$ binary BCH code with parameters

$$n = 2^{\nu} - 1, \quad n - k \leq \nu t, \quad d_{\min} \geq 2t + 1 \,, \tag{7.83}$$

where ν is the Galois field extension. This code can correct all error patterns of t or fewer errors, and it is usually referred to as a t-error-correcting BCH code.

BCH codes offer great flexibility in the choice of code length, rate, and error-correcting capability. To achieve further flexibility in terms of code length and rate, BCH codes may be shortened. An $(n', k', d_{\min})$ BCH code of parameters

$$(n', k', d'_{\min}) = (n - s, k - s, d'_{\min} \geq d_{\min}) \tag{7.84}$$

is obtained from an $(n, k, d_{\min})$ BCH code by setting s information bits to zero and not transmitting them. The parameter s is commonly referred to as the shortening parameter. At the receiver, the shortened positions can then be assumed to be transmitted without error. Since the parameters ν, t, and s completely determine (n, k), sometimes it is useful to define a BCH code through the triple (ν, t, s).

The rich structure of BCH codes allows for simple syndrome-based HDD. A very efficient algebraic HDD for BCH codes is the *Berlekamp–Massey* algorithm [7.38, 39], which finds the most probable error pattern introduced by the channel iteratively given the syndrome corresponding to the received binary vector. If the number of redundancy symbols $n-k$ is very large, however, (complete) syndrome decoding may be too complex. To lower the decoding complexity, *incomplete* syndrome decoding can be used. Incomplete syndrome decoding is usually referred to as BDD, which can also be implemented based on the Berlekamp–Massey algorithm. For a block code with error correction capability t, BDD corrects all error patterns with Hamming weight less than or equal to t and no others. Note that BDD is not ML; therefore, it incurs a penalty in performance.

For details on the construction of BCH codes and the Berlekamp–Massey algorithm, the interested reader is referred to [7.15, Chap. 9] and [7.101, Chap. 3]. A nice overview of syndrome decoding can be found in [7.201, Chap. 10].

Reed–Solomon Codes

Reed–Solomon codes are the most important and widely used class of nonbinary block codes. Reed–Solomon codes have made their way into multiple applications and standards, such as Digital Video Broadcasting (DVB) [7.202] and deep-space communications [7.203], as well as for data storage and distributed data storage systems such as RAID 6 [7.204].

Reed–Solomon codes were introduced in 1960 by *Reed* and *Solomon* [7.37] and can be seen as a subclass of BCH codes generalized to the nonbinary case. For Reed–Solomon codes, both the construction field and the symbol field are the same, i.e., the code symbols are over a nonbinary field. An $(n, k, d_{\min})$ Reed–Solomon code over the Galois field GF(2^{ν}) has parameters

$$n = 2^{\nu} - 1, \quad k = 2^{\nu} - 2t - 1, \quad d_{\min} = 2t + 1 \,, \tag{7.85}$$

where the code symbols are over GF(2^{ν}). The Galois field GF(2^{ν}) is an extension field of the binary numbers $\{0, 1\}$ and contains 2^{ν} nonbinary symbols, which can be represented using ν bits, together with well-defined addition and multiplication operations. A Reed–Solomon code with these parameters is capable of correcting all error patterns of t or fewer symbol errors.

The minimum Hamming distance of Reed–Solomon codes achieves the Singleton bound, $d_{\min}(C_{\mathrm{RS}}) = n - k + 1$ (this is readily seen from (7.85)), i.e., their minimum Hamming distance is the maximum possible for a linear code of parameters (n, k) (see Theorem 7.1 in Sect. 7.3).

Note that the code symbols of a Reed–Solomon code can also be interpreted as binary vectors by representing each symbol by a binary tuple of ν bits. In this case, Reed–Solomon codes can be used for binary transmission (e.g., using BPSK) and are well suited for burst-error correction.

Similar to BCH codes, Reed–Solomon codes can also be shortened to yield higher flexibility in terms of block length and code rate. A shortened Reed–Solomon code with shortening parameter s obtained from an

$(n, k, d_{\min})$ Reed–Solomon code over the Galois field GF(2^{ν}) has parameters

$$(n', k', d'_{\min}) = (n - s, k - s, d_{\min}) \,. \tag{7.86}$$

Reed–Solomon codes can also be efficiently decoded using the Berlekamp–Massey algorithm.

7.5.2 Product Codes

PCs, introduced by *Elias* in 1954 [7.18], are one of the first examples of a construction of long, powerful codes from short component codes. Let C_1 be an (n_1, k_1, d_1) linear block code of rate $R_1 = k_1/n_1$ and minimum Hamming distance d_1, and C_2 an (n_2, k_2, d_2) linear block code of rate $R = k_2/n_2$ and minimum Hamming distance d_2. A two-dimensional $(n, k, d_{\min})$ PC $C_{\mathrm{PC}} = C_1 \times C_2$ is defined by a rectangular array $\mathbf{C}$ such that each row of $\mathbf{C}$ is a codeword of C_1 and each column is a codeword of C_2. The codes C_1 and C_2 are commonly referred to as *component codes*. Both binary and nonbinary PCs can be constructed. In the first case, the component codes are typically BCH codes, although other (simpler) linear block codes such as single parity-check codes and Hamming codes can also be used. The component codes of nonbinary PCs, on the other hand, are usually Reed–Solomon codes. The code array $\mathbf{C}$ for a PC with component codes with parameters $(n_1, k_1) = (6, 4)$ and $(n_2, k_2) = (6, 3)$ is depicted in Fig. 7.30.

A PC can be conveniently described by the encoding procedure. The $k = k_1 k_2$ information bits are arranged in a $k_2 \times k_1$ array $\mathbf{U}$ and placed in the upper-left quadrant of the code array (in orange in Fig. 7.30). Each row of the array $\mathbf{U}$ is then encoded by the component code C_1 to generate $n_1 - k_1$ parity bits, marked in blue in the figure. Each column of the resulting array, of dimensions $k_2 \times n_1$, is then encoded by the code C_2 to generate $n_2 - k_2$ parity bits, marked in green. The resulting PC is a systematic code with parameters $k = k_1 k_2$, $n = n_1 n_2$, and rate $R = k_1 k_2/(n_1 n_2) = R_1 R_2$. It is also easy to show that the minimum Hamming distance of a PC is the product of the minimum Hamming distances of the component codes, i.e., $d_{\min} = d_1 d_2$. Thus, PCs are usually characterized by a large minimum Hamming distance. On the other hand, the minimum Hamming distance multiplicity, i.e., the number of codewords of weight $d_{\min}$, of a PC is equal to the product of the minimum Hamming distance multiplicities of the component codes [7.205]. As a result, the minimum Hamming distance multiplicities of PCs are typically high.

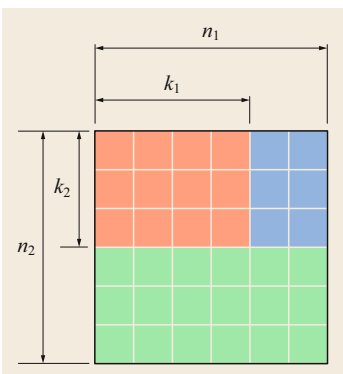


Fig. 7.30 Code array of a PC with component codes C_1 and C_2 of parameters $(n_1, k_1) = (6, 4)$ and $(n_2, k_2) = (5, 3)$, respectively. *Orange squares* correspond to data bits; *blue* and *green squares* correspond to parity bits

With the construction above, it is apparent that each row of $\mathbf{C}$ is a codeword of C_1 and each column of C is a codeword of C_2. Therefore, all code bits of a PC are protected by two component codes, or in other words, each code bit participates in two *code constraints*, a row code constraint and a column code constraint. For example, the second data bit is protected by the second row and second column constraints.

Irregular Product Codes and Multidimensional Product Codes

PCs, as described above, are based on a single row and column code, i.e., the same code is used for all rows and the same code is used for all columns. However, all row and column codes do not need to be the same, and codes of different rates and erasure-correcting capabilities can be used as row and column codes. The resulting PC is referred to as an *irregular* PC [7.206, 207]. Like irregular LDPC codes, irregular PCs based on code mixtures may yield performance improvements. It is also important to note that PCs may be built over arrays with more than two dimensions [7.18]. However, two-dimensional PCs are the most common, and therefore in this chapter we restrict ourselves to two-dimensional PCs.

7.5.3 Generalized Product Codes

The construction of PCs, with their row-column encoding and the underlying rectangular code array, is a natural construction for constructing long codes from simpler codes. However, more general code arrays, where each row and each column of the array are codewords of given linear block codes, can also be constructed. This gives rise to so-called *product-like codes*, which we will refer to here as generalized product codes (GPCs), since they can be seen as generalizations of PCs. Two of the most popular classes of GPCs are staircase codes and braided codes.

Staircase Codes

Staircase codes are an important class of error-correcting codes proposed by *Benjamin Smith* et al. [7.49] for optical transport networks. A staircase code is defined by a two-dimensional code array that has the form of a staircase. Formally, given a component code C of length n_c, a staircase code is defined as the set of all binary matrix sequences $\mathbf{B}_i$, $i = 1, 2, \ldots$, of dimensions $a \times a$, where $a = n_c/2$, such that each row of $[\mathbf{B}_i^\top, \mathbf{B}_{i+1}]$ is a valid codeword of C. The matrix $\mathbf{B}_1$ is assumed to be initialized to all zeros. The code array of a staircase code with component code C of code length $n_c = 12$ and dimension $k_c = 10$ is depicted in Fig. 7.31. Encoding is performed in batches, where each batch corresponds to one of the matrices $\mathbf{B}_i$. Specifically, batch i corresponds to $\mathbf{B}_i^\top$ for i odd and to $\mathbf{B}_i$ for i even. The number of bits per batch, denoted by B, is $B = a^2 = n_c^2/4$.

Like PCs, staircase codes are well defined through the encoding procedure. With reference to Fig. 7.31, the first batch (matrix $\mathbf{B}_1^\top$) is initialized to all zeros. Then, $(k_c - n_c/2)(n_c/2) = 4 \times 6$ data bits are placed in the left part of batch 2 (marked with a thick red frame in the figure), and the the rows of batches 1 and 2 are encoded by the component code C, thus generating $(n_c - k_c)(n_c/2) = 2 \times 6$ parity bits, i.e., $n_c = 2$ parity bits per row. $(k_c - n_c/2)(n_c/2)$ data bits are then placed in the upper part of batch 3, and the columns of the array formed by batches 2 and 3 are encoded using C, generating $n_c - k_c = 2$ parity bits per column. The row/column encoding process continues accordingly for the next batches. The code rate of the staircase code, R_{SC}, is the ratio between the number of information bits per batch and B, i.e.,

$$R_{SC} = \frac{(k_c - n_c/2)(n_c/2)}{n_c^2/4} = 2\frac{k_c}{n_c} - 1 \,. \tag{7.87}$$

Note that, as for PCs, each bit of a staircase code is protected by two code constraints, a row constraint and a column constraint. Also, it is worth mentioning that staircase codes are stream-oriented. Indeed, the code array in Fig. 7.31 can grow infinitely large.

Good (binary) staircase codes with BCH component codes and overhead ranging from 6% to 33% can be found in [7.208], where the parameters of the component BCH codes were optimized based on computer simulations. An optimization of the staircase code parameters based on density evolution (see Sect. 7.5.7 below) is addressed in [7.209]. An extension of the original code construction with larger staircase block sizes by allowing for multiple code constraints per row/column in the staircase array is also proposed in [7.209], leading to codes with steeper waterfall performance. Nonbinary staircase codes with Reed–Solomon component codes are discussed in [7.90, 210].

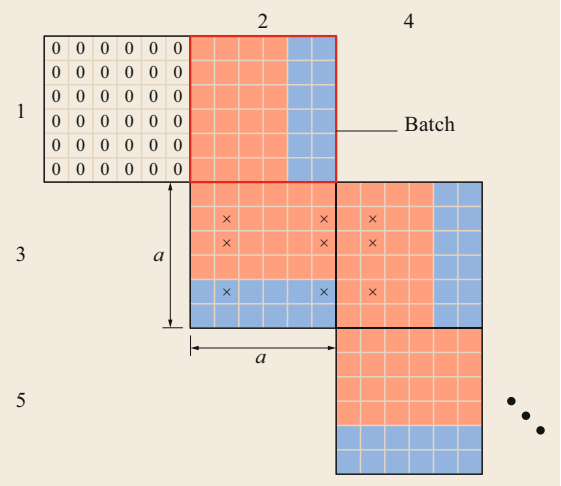


Fig. 7.31 Code array of a staircase code with component code of code length $n_c = 12$ and dimension $k_c = 10$. The *orange squares* correspond to data bits, while the *blue squares* correspond to parity bits

Braided Codes

Braided codes were originally proposed by *Jiménez Feltström* et al. [7.200]. Similar to PCs and staircase codes, the code construction is defined by a two-dimensional array where each bit is protected by a row and a column code. Braided codes can be constructed using both convolutional and block codes as component codes. Correspondingly, the resulting codes are referred to as braided convolutional codes and braided block codes, respectively. Here we are interested mainly in block braided codes, as they are more suitable for HDD, and for ease of exposition we will refer to them simply as braided codes.

Braided codes come in many flavors, depending on the structure of the code array. In Fig. 7.32, we depict the code array of a so-called block-oriented braided code with component code C of code length $n_c = 12$ and dimension $k_c = 10$. The encoding is similar to that of product and staircase codes: Data bits are placed in the code array in the parts marked in orange, and parity bits (in blue) are generated by row and column encoding. Encoding is better understood with reference to Fig. 7.32. The code array consists of blocks of dimensions $b \times b$, where $b = n_c/3$. The blocks are grouped in batches comprising three blocks. The number of bits per batch is $B = 3b^2 = n_c^2/3$. The first batch is shown by the thick red frame in the figure. The leftmost and uppermost blocks of the code array are initialized to all zeros. Then, $b^2 + 2(k_c - 2b)b = n_c^2/9 + 2(k_c - 2n_c/3)(n_c/3) =$

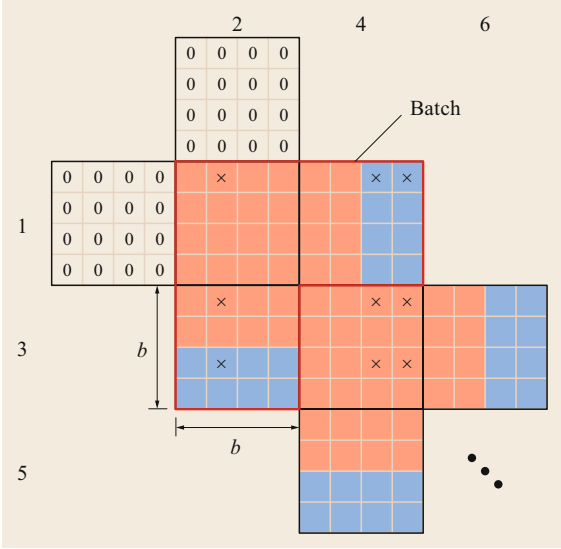


Fig. 7.32 Code array of a braided code with component code of code length $n_c = 12$ and dimension $k_c = 10$. The *orange squares* correspond to data bits, while the *blue squares* correspond to parity bits

32 data bits are placed in the orange squares of the first batch. The rows of the blocks at (horizontal) position 1 are then encoded by the (n_c, k_c) component code C, generating $(n_c - k_c)(n_c/3) = 2 \times 4 = 8$ parity bits, i.e., $n_c = 2$ parity bits per row (marked in blue in the figure). The columns of the array formed by the three blocks in (vertical) position 2 are encoded using C, generating $n_c - k_c = 2$ parity bits per column. The row/column encoding process continues accordingly for the next batches. The code rate of the braided code, R_{BC}, is the ratio between the number of information bits per batch and B, i.e.,

$$R_{BC} = \frac{(n_c^2/9) + 2(k_c - 2n_c/3)(n_c/3)}{n_c^2/3} = 2\frac{k_c}{n_c} - 1 \,. \tag{7.88}$$

Braided codes have been explicitly considered for use in fiber-optic communication systems in [7.50].

Half-Product Codes and Half-Braided Codes

Consider a PC based on component codes of lengths $n_1 = n_2 = n_c$ (Fig. 7.30). Based on this PC, a new code can be obtained by retaining only codeword arrays with zeros in the main diagonal that are symmetric, in the sense that the array is equal to its transpose. The code bits in the main diagonal of the resulting array do not carry any information (they are known), and hence they can be punctured, i.e., they are not transmitted. Also,

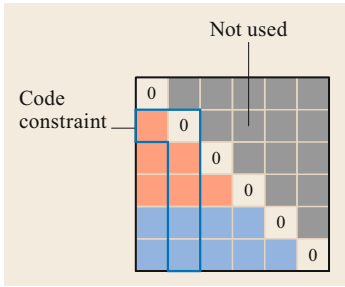


Fig. 7.33 Code array of an HPC with component code of code length $n_c = 6$ and dimension $k_c = 3$. The *orange squares* correspond to data bits, while the *blue squares* correspond to parity bits. The bits corresponding to the *gray squares* are not transmitted

the bits in the upper triangular part are identical to those in the lower triangular part and can also be punctured. The resulting code has length $n = \binom{n_c}{2}$ and is referred to as a half-product code (HPC) [7.211]. The code array of an HPC with component codes of length $n_c = 6$ is depicted in Fig. 7.33, where one particular code constraint is marked with the thick blue frame. In [7.212], it is shown that HPCs can achieve larger normalized minimum distance than PCs and superior performance in both the waterfall region and the error floor compared with PCs of similar length and rate.

The concept above can be extended to other code arrays. For example, the same symmetry constraint can be applied to braided codes, and the resulting codes are referred to as half-braided codes. The code array of a half-braided code with component codes with block length $n_c = 12$ and dimension $k_c = 10$ is depicted in Fig. 7.34. It is easy to see that the code rate of a half-braided code in the form of Fig. 7.34, R_{HBC}, is

$$R_{HBC} = 2\frac{k_c - 1}{n_c - 1} - 1 \,. \tag{7.89}$$

HPCs and half-braided codes belong to a larger class of symmetric GPCs [7.212]. Symmetric GPCs are in turn a subclass of GPCs that use symmetry to reduce the block length of a GPC while using the same component code. In [7.213], it was shown that half-braided codes can outperform both staircase codes and braided codes in waterfall performance, at a lower error floor and decoding delay.

Similar to staircase codes, braided codes, HPCs, and half-braided codes, other generalizations of PCs (leading to other code arrays) can be envisaged. In Sect. 7.5.5, we discuss a general construction of GPCs that encompasses staircase codes and braided codes, as well as other classes of GPCs.

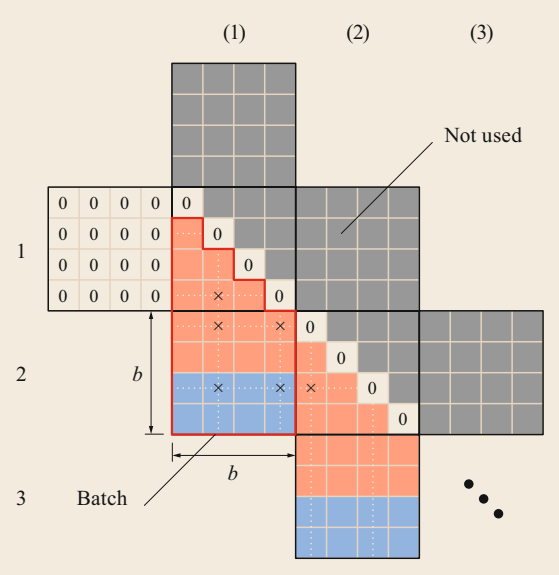


Fig. 7.34 Code array of a half-braided code with component code of code length $n_c = 12$ and dimension $k_c = 10$. The *orange squares* correspond to data bits, while the *blue squares* correspond to parity bits. The bits corresponding to the *gray squares* are not transmitted

7.5.4 Generalized Product Codes as Instances of Generalized Low-Density Parity-Check Codes

It is interesting to observe that GPCs can be interpreted as instances of generalized LDPC (G-LDPC) codes. The link between GPCs and G-LDPCs is relevant because it allows us to borrow tools for the analysis of LDPC codes (and codes on graphs in general) to analyze the behavior of GPCs. G-LDPC codes are a generalization of LDPC codes, where the constraint nodes are not simple single parity-check codes as for LDPC codes, but general linear block codes, e.g., Hamming codes or BCH codes [7.101]. A G-LDPC code is also defined by a bipartite graph where an edge between a VN and a CN indicates that the code bit corresponding to the VN participates in the code constraint enforced by the component code represented by the CN.

To clarify the link between GPCs and G-LDPC codes, let us first consider the Tanner graph representation of a PC. As any block code, PCs can be represented by a Tanner graph where VNs represent code bits and CNs represent row and column codes. For a PC with component codes of code lengths n_1 and n_2, the corresponding Tanner graph has $n_1 + n_2$ CNs (n_1 CNs corresponding to the column codes and n_2 CNs corresponding to the row codes, Fig. 7.30) and $n_1 n_2$ VNs corresponding to the $n_1 n_2$ code bits. The (bipartite) Tanner graph of the PC with component codes with parameters $n_1 = n_2 = 6$ in Fig. 7.30 is depicted in Fig. 7.35. Note that each CN has degree 6, since the component codes have code length $n_1 = n_2 = 6$, and each VN has degree 2, since each code bit participates in two code constraints. A simplified version of the Tanner graph of Fig. 7.35 is illustrated in Fig. 7.36. In the simplified Tanner graph, we distinguish between two types of CNs, corresponding to row and column constraints. The VNs (corresponding to code bits) are represented by edges. Since each code bit is protected by a row and a column code, each edge is connected to one CN of each type. The resulting Tanner graph is a complete graph where each CN of one type is connected to all CNs of the other type through an edge. There are $n_1 n_2$ edges, each corresponding to one code bit. Note that the code array in Fig. 7.30, the *standard* Tanner graph in Fig. 7.35, and the simplified Tanner graph in Fig. 7.36 are equivalent representations of the PC. For compactness reasons, it is more convenient to represent PCs with the simplified Tanner graph than with the conventional Tanner graph of Fig. 7.35.

Similar to PCs, GPCs can be represented by a Tanner graph. The standard Tanner graph and the simplified Tanner graph of a staircase code (Fig. 7.31) with component code of code length $n_c = 6$ are depicted in Fig. 7.37 and Fig. 7.38, respectively. The standard Tanner graph of a staircase code is defined by a number of positions, L, corresponding to the the total number

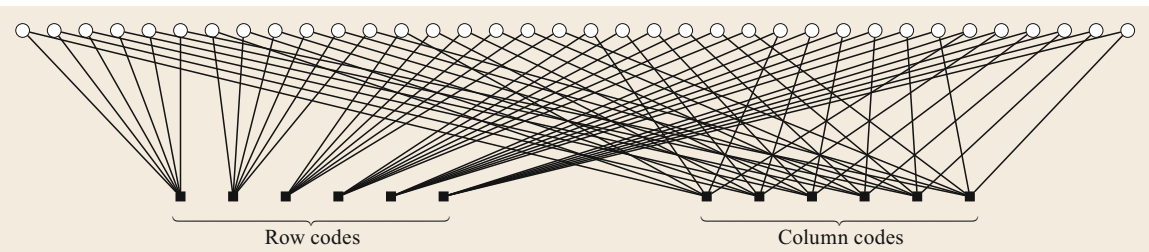


Fig. 7.35 Bipartite (Tanner) graph of a PC with component codes of lengths $n_1 = n_2 = 6$ of Fig. 7.30

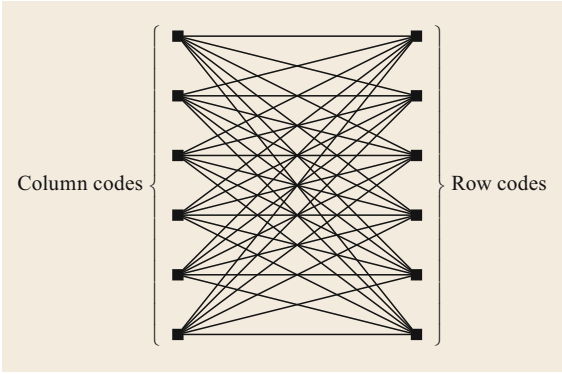


Fig. 7.36 Simplified Tanner graph of a PC with component codes of lengths $n_1 = n_2 = 6$ of Fig. 7.30

of batches minus one. Each position corresponds to one batch of code bits and correspondingly has $a^2 = (n_c/2)^2$ VNs, corresponding to the a^2 code bits in the batch, and $a = n_c/2$ CNs, corresponding to the row (or column) codes defined by the batches $[\mathbf{B}_i^\top, \mathbf{B}_{i+1}]$ for position i, $i = 1, 2, \ldots, L$. Accordingly, the simplified Tanner graph is also defined by L positions, with a CNs per position. Each CN in position i is then connected to all CNs in spatial position $i+1$ by an edge (a VN, corresponding to a code bit), Fig. 7.38. Note that all VNs have degree 2, since a code bit participates in two code constraints, and all CNs have degree 6, since $n_c = 6$.

Figures 7.35–7.38 make apparent the link between GPCs and G-LDPC codes. In particular, the bipartite Tanner graph of a staircase code resembles that of an SC-LDPC code (Fig. 7.37 and Fig. 7.19 in Sect. 7.4.2). Indeed, staircase codes can be seen as spatially coupled versions of PCs, and as such, as a subclass of spatially coupled G-LDPC codes with coupling width $w = 2$. Similar to staircase codes, other GPCs such as braided codes and half-braided codes can be classified as SC-GLDPC codes.

The connection of GPCs to G-LDPC codes enables the use of tools for the analysis of codes on graphs, such as density evolution, to analyze GPCs. This is discussed in Sect. 7.5.7.

7.5.5 A General Code Construction of Generalized Product Codes

In the previous sections, we described some of the most popular classes of GPCs. In this section, we briefly review a deterministic construction of GPCs proposed in [7.214, 215] that is sufficiently general to encompass several classes of GPCs, including irregular PCs, HPCs, staircase codes, block-wise braided codes, and half-braided codes.

The code construction in [7.214, 215] is defined in terms of three parameters, $\boldsymbol{\eta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\tau}$, and is specified over the simplified Tanner graph. We denote the corresponding GPC by $C_m(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\tau})$, where m denotes the total number of CNs in the underlying Tanner graph; $\boldsymbol{\eta}$ is a binary, symmetric $L \times L$ matrix, where L is the number of positions of the Tanner graph, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)^\top$ is a probability vector of length L, i.e., $\sum_{i=1}^{L} \gamma_i = 1$ and $\gamma_i \geq 0$. The parameters $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ determine the graph connectivity. Considering the representation of a GPC in terms of the two-dimensional code array (Figs. 7.30–7.34), one may alternatively think about $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ as specifying the array shape. Different choices for $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ recover well-known code classes. The parameter $\boldsymbol{\tau}$ is used to specify GPCs employing component codes with different erasure-correcting capabilities.

The simplified Tanner graph describing the GPC $C_m(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\tau})$ is constructed as follows. Assume that the Tanner graph spans L positions. Then, place $m_i \triangleq \gamma_i m$ CNs at each position $i = 1, \ldots L$, where it is assumed that m_i is an integer for all i. Then, connect each CN at position i to each CN at position j through a VN (i.e., an edge) if and only if $\eta_{i,j} = 1$.

Example 7.5

A PC is obtained by choosing $L = 2$ and $\boldsymbol{\eta} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. The two positions correspond to the row and column codes. Choosing $\boldsymbol{\gamma} = (1/2, 1/2)$ leads to a standard PC (with a square code array) with component codes C_1 and C_2 of length $n_1 = n_2 = m/2$. Note that by selecting $m = 12$, we recover the Tanner graph of the PC in Fig. 7.36.

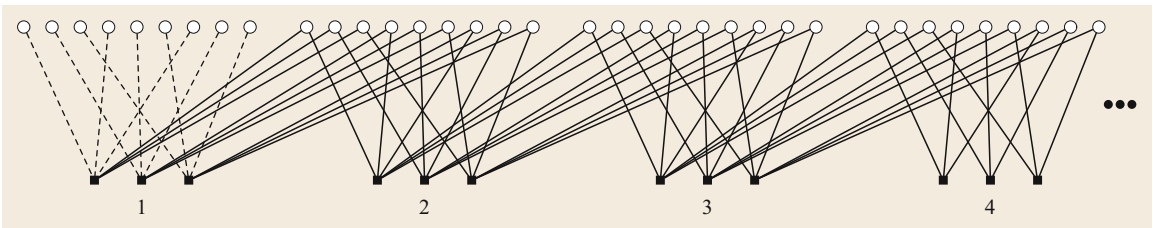


Fig. 7.37 Bipartite (Tanner) graph of a staircase code (Fig. 7.31) with component code of code length $n_c = 6$

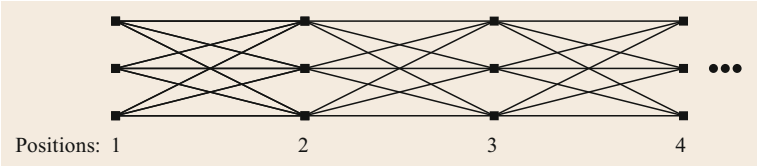


Fig. 7.38 Simplified Tanner graph of a staircase code (Fig. 7.31) with component code of code length $n_c = 6$

Example 7.6

For a fixed $L \geq 2$, the matrix $\boldsymbol{\eta}$ describing a staircase code has entries $\eta_{i,i+1} = \eta_{i+1,i} = 1$ for $i = 1, \ldots, L-1$ and zeros elsewhere. The distribution $\boldsymbol{\gamma}$ is uniform, i.e., $\gamma_i = 1/L$ for all $i \in [L]$. For example, the staircase code corresponding to the code array shown in Fig. 7.31, where $L = 4$ and $m = 24$ (i.e., $m_i = 6$), is defined by

$$\boldsymbol{\eta} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{7.90}$$

and $\gamma_i = 1/6$. The CNs at all positions have the same degree $2n\gamma_i = 12$, except for positions 1 and L, where the degrees are $n\gamma_i = 6$. With this choice of $\boldsymbol{\eta}$ we recover the simplified Tanner graph in Fig. 7.38.

For more details on this general code construction, the interested reader is referred to [7.215].

7.5.6 Decoding of Product-Like Codes: Iterative Bounded Distance Decoding

PCs may be decoded with SDD using an iterative decoding algorithm based on the iterative exchange of soft information between two soft-input soft-output (SISO) decoders matched to the two component codes. The decoding is similar to that of turbo codes, and for this reason, PCs with iterative SDD are often referred to as turbo product codes (TPCs) or block turbo codes (BTCs). Similarly, GPCs can also be decoded by exchanging soft information between the SISO decoders of the component codes in an iterative fashion. The SISO decoders of the component codes may implement the MAP decoding rule, which can be done based on a trellis representation of the block code. However, MAP decoding of algebraic block codes is complex, and therefore iterative SDD of GPCs based on MAP decoding of the component codes is impractical. An efficient, lower-complexity iterative SDD algorithm for PCs with near-optimal performance was proposed by *Pyndiah* [7.58]. In this algorithm, the SISO decoders of the component codes are based on a modification of the Chase decoding algorithm, a low-complexity suboptimal algorithm for near-ML decoding of linear block codes proposed by *David Chase* in 1972 [7.216]. Since the Chase decoder is a vital component of the iterative SDD algorithm proposed in [7.58], the algorithm used to decode the component codes is usually referred to as the Chase–Pyndiah decoding algorithm. The advantage of this decoding algorithm is that it is highly parallelizable.

The iterative decoding of GPCs employing SISO component decoding is usually referred to as turbo product decoding (TPD) and is typically implemented in practice via the Chase–Pyndiah algorithm. However, it is worth mentioning that TPD can also be based on other component decoders, such as the forward-backward algorithm applied to the component code trellis.

The complexity of the Chase–Pyndiah decoder may still be unacceptable for some applications. Alternatively, one can decode PCs and GPCs using HDD. Indeed, PCs and GPCs are very well suited for HDD thanks to the fact that the component codes (typically BCH or Reed–Solomon codes) have a rich algebraic structure, which enables efficient implementation of BDD.

Iterative Bounded Distance Decoding of Product Codes

For large block codes, optimal (ML) HDD is a formidable task. As mentioned in Sect. 7.5.1, to reduce the decoding complexity, incomplete BDD can be used. However, even BDD of powerful codes such as PCs and GPCs is a challenging task. To achieve acceptable complexity, an iterative (suboptimal) decoding algorithm, based on BDD of the component codes, can then be used. Let us first consider HDD of PCs. The decoding of a PC can be accomplished by iteratively applying BDD to the row and column codes, described as follows. The received bits are arranged in an $n_2 \times n_1$ array corresponding to the code array (Fig. 7.30). Decoding is then carried out on the columns of the array, i.e., BDD decoding of the column component codes C_2 is performed, and subsequently on the rows of the array, i.e., BDD decoding of the row component codes C_1 is performed. Errors remaining after the first column decoding may be corrected by the bounded distance decoders of the row codes. The column-row decoding is then iterated until a maximum number of iterations is reached. We refer to this algorithm as iBDD.

iBDD achieves excellent performance-complexity trade-off. The decoding complexity of iBDD is roughly the sum of the decoding complexity of the BDD of the component codes times the number of iterations. Hence, the complexity of iBDD is much lower than that of ML decoding of the PC. On the other hand, since iBDD is suboptimal, the lower decoding complexity comes at a cost of lower performance. The complexity of iBDD of PCs and GPCs in general is also much lower than that of (suboptimal) BP decoding of LDPC codes. Moreover, the decoder data flow requirements are estimated to be more than one order of magnitude smaller than the requirements for a comparable LDPC code with BP decoding [7.49].

Iterative Bounded Distance Decoding of Generalized Product Codes

Like PCs, GPCs can be decoded using iBDD by iterating between the row and column decoders. For simplicity, in this section we describe the decoding of staircase codes. However, the decoding of other GPCs such as braided codes and half-braided codes follows the same principle.

Since staircase codes may be very long (as mentioned earlier, they are stream-oriented in nature), waiting to receive the entire code array prior to the start of decoding would entail a very long delay, which is not acceptable in fiber-optic communications. Fortunately, the batch-oriented structure of staircase codes allows us to perform decoding in a sliding-window fashion, similar to windowed decoding of SC-LDPC codes (Sect. 7.4.2). The windowed decoder operates on part of the array consisting of a window comprising a subset of W received batches. The decoder then iterates between the BDD decoders for all rows and all columns within the window. After a predetermined number of iterations, the window slides to the next position, and iterative decoding is performed within the new window. The decoding delay (in bits) is given by $D = WB$, where $B = a^2$ is the number of code bits per batch (Sect. 7.5.3). The window size provides a trade-off between performance and decoding latency, i.e., one expects a performance improvement by increasing the window size at a cost of higher decoding latency. In general, small values of W (6 to 8) are typically sufficient.

Despite the suboptimality of the iBDD algorithm, staircase codes—and GPCs in general—provide excellent performance for high code rates. For example, the $R = 0.937$ staircase code designed in [7.49] performs only about 0.56 dB from the channel capacity of the BSC.

It is worth mentioning that BDD of the component codes may lead to *miscorrections*, i.e., the component decoders may declare successful decoding, but the decoded codeword is not the correct one. Miscorrections introduce additional bit errors into the iterative decoding process, which result in performance degradation. The impact of miscorrections becomes less severe as the error-correcting capability t increases. In [7.217], a decoding algorithm for staircase codes was proposed, called *anchor decoding*, which reduces the effect of undetected component code miscorrections. The algorithm significantly improves performance, in particular when t is small, while retaining low-complexity implementation.

Iterative Bounded Distance Decoding as a Message-Passing Algorithm

With reference to the Tanner graph representation of GPCs, iBDD of GPCs can be interpreted as an iterative message-passing decoding algorithm similar to that of LDPC codes, where *hard* (binary) messages are exchanged between VNs and CNs in the Tanner graph describing the code. This opens the door to the analysis of the performance of iBDD of GPCs using codes-on-graphs tools.

It is worth pointing out that the classical message-passing rule typically used to decode PCs and GPCs violates the principle that only *extrinsic* information should be exchanged during the iterative decoding process [7.218], since the computation of the messages passed from CNs to VNs use the input messages at the CNs. For this reason, we will use the same terminology introduced in [7.50] and refer to the decoding algorithm described above as iterative HDD with intrinsic message passing (IMP). In [7.219], a modified message-passing algorithm that exchanges only extrinsic information, dubbed extrinsic message passing (EMP), was proposed. EMP yields better performance than IMP at a cost of some increase in complexity (mostly memory complexity, which is doubled). A low-complexity EMP HDD algorithm is given in [7.50].

7.5.7 Analysis and Optimization of Generalized Product Codes

Similar to LDPC codes and codes on graphs in general, the BER curve of GPCs is characterized by two well-defined regions. In the first one, called the waterfall region, the BER decreases sharply with E_b/N_0. The curve then flattens out in the so-called error floor region. The error floor is usually caused by combinations of errors that cannot be corrected by the decoder. For GPCs with iBDD, these patterns are usually referred to as *stall patterns* [7.49]. A stall pattern for the staircase code, the braided code, and the half-braided code in Figs. 7.31, 7.32, and 7.34, respectively, assum-

ing $t = 2$ error-correcting component codes, is shown by the crosses in the figures. Since every component codeword involved has three errors, decoding will not be able to proceed, and stalls.

Let $s_{\min}$ be the size of the minimum stall pattern, defined as the minimum number of array positions that, when all received in error, cause the decoder to stall. A stall pattern is said to be assigned to a batch if at least one of its array positions belongs to the batch and no positions belong to previous batches. The error floor of staircase codes, braided codes, and half-braided codes can be approximated by

$$\mathrm{BER} \approx \frac{s_{\min} M p^{s_{\min}}}{B}, \tag{7.91}$$

where M denotes the number of minimum stall patterns that can be assigned to a batch. For staircase codes, M is derived in [7.49, Sect. V-B]. Using similar arguments, the values of M for braided codes and half-braided codes were obtained in [7.213].

The performance of GPCs in the waterfall region is more difficult to analyze. However, the asymptotic performance of GPCs in the limit of infinite block lengths can be analyzed exploiting the connection with G-LDPC codes. The analysis of codes on graphs such as LDPC codes and G-LDPC codes is commonly based on an ensemble argument, i.e., rather than analyzing a particular code, which is very cumbersome, it is customary to analyze the average behavior of a code ensemble. A code ensemble can be regarded as a family of codes that share some common characteristics, such as the degree distribution. The asymptotic behavior of a code ensemble can be analyzed via a tool called density evolution [7.117], which can be used to find the so-called decoding threshold. The decoding threshold divides the channel parameter range (e.g., SNR) into a region where reliable decoding is possible and another region where it is not, and accurately predicts the region where the code performance curve *bends* into the characteristic waterfall behavior. There exists a concentration phenomenon that ensures with high probability that a particular code taken uniformly at random from the ensemble will have actual performance close to th predicted by density evolution.

The parameters of GPCs, i.e., the parameters of the component codes, can be optimized to achieve good performance in the waterfall region based on extensive software simulations for predicting the code performance, as was done in [7.208] for staircase codes. Alternatively, one can use the fact that a specific GPC, e.g., a staircase code, is contained in a given ensemble of G-LDPC codes, and can optimize the code parameters based on density evolution to optimize the decoding threshold. The optimization approach based on density evolution offers significantly reduced optimization time (or the possibility of exploring a larger parameter space) with respect to a simulation-based approach. The optimization of staircase code parameters based on density evolution was addressed in [7.209].

It is very important to emphasize that, contrary to LDPC codes, which are drawn from an ensemble, GPCs are deterministic codes, i.e., they have a very regular structure in terms of their Tanner graph and are not at all random-like. Thus, while the optimization of GPCs via ensemble-based density evolution yields good codes, formally speaking, the asymptotic performance of GPCs is not necessarily characterized by density evolution of the corresponding spatially coupled G-LDPC code ensemble. For transmission over the binary erasure channel (BEC), the density evolution equations that characterize the asymptotic decoding performance of a broad class of GPCs with a fixed Tanner graph were derived in [7.213–215], without relying on the definition of a code ensemble. The results obtained can also be used to analyze the code performance over the BSC assuming idealized BDD, i.e., completely ignoring miscorrections (idealized BDD over the BSC is conceptually equivalent to transmission over the BEC). However, the rigorous characterization of the asymptotic performance of GPCs over the BSC, including the effect of decoder miscorrections, is still an open problem. An asymptotic analysis taking into account the impact of miscorrections for a code ensemble related to staircase codes is given in [7.220].

7.5.8 Soft-Aided Decoding of Product-Like Codes

In Sect. 7.4 and in this section, we have provided an extensive review of soft-decision FEC and hard-decision FEC, respectively. Roughly speaking, soft-decision FEC yields higher NCGs at a cost of higher decoding complexity and data flow compared with hard-decision FEC. Thus, despite its smaller NCG, for applications where very high throughput and low power consumption are required, hard-decision FEC is still an appealing alternative. An interesting question is whether it is possible to close the gap between the performance of hard-decision and soft-decision FEC while keeping the decoding complexity low. One line of research recently pursued in this direction is to concatenate an inner soft-decision FEC code, e.g., an LDPC code decoded via BP, with an outer staircase code [7.221, 222]. Another alternative, recently investigated in [7.223, 224], is to improve the performance of iBDD of product-like codes by exploiting some level of soft information while keeping the core

algebraic decoding of the component codes, an approach that we refer to here as *soft-aided* decoding of product-like codes. Without going into detail, the algorithm in [7.223], dubbed iBDD-SR, exploits the channel reliabilities while still only exchanging binary (hard-decision) messages between component codes. iBDD improves performance over conventional iBDD, with only a minor increase in complexity. In [7.224], another soft-aided decoding algorithm was proposed, based on generalized minimum distance (GMD) decoding of the component codes. The algorithm, referred to as iterative GMD decoding with scaled reliability (iGMDD-SR) yields more significant coding gains while maintaining significantly lower complexity than SDD. In particular, for a PC, it was shown that iGMDD-SR closes over 50% of the performance gap relative to turbo product decoding, while maintaining significantly lower complexity. We note that while the algorithms in [7.223, 224] were demonstrated for PCs, their extension to staircase codes and other product-like codes is straightforward. The implementation of iBDD for a PC has been considered in [7.225], where it is shown that, using a single bit of soft information, the NCG can be improved by 0.2 dB with respect to iBDD, reaching 10.3−10.4 dB, which is similar to that of more complex hard-decision staircase decoders, but with significantly lower circuit area and energy dissipation.

7.5.9 Performance Curves

In Fig. 7.39, we plot the BER performance of two staircase codes of rate $R = 0.75$ (corresponding to a code overhead of 33.33%) for transmission over the BSC. The net coding gain is also shown. In particular, the figure shows the performance of the staircase code with BCH component codes with parameters $(\nu, t, s) = (9, 5, 151)$, i.e., $(n_c, k_c) = (360, 315)$ [7.208, Table II], referred to as C_1 in the figure. The parameters of the BCH component code were optimized in [7.208] based on extensive simulations. The performance of the code is shown for both IMP decoding, as originally proposed in [7.49], and EMP decoding, as proposed in [7.50]. The decoding is performed in a sliding-window fashion as explained in Sect. 7.5.6. A window size $W = 7$ and $\ell = 8$ iterations within a window are considered. It is observed that EMP yields better performance than IMP in the waterfall region. In the figure, we also plot density evolution results [7.209]. The density evolution curve predicts the pre-FEC BER region where the simulated curve (with EMP) bends into the characteristic waterfall.

In the figure, we also plot the performance of a staircase code of the same code rate whose parameters have been optimized based on density evolution [7.209], and

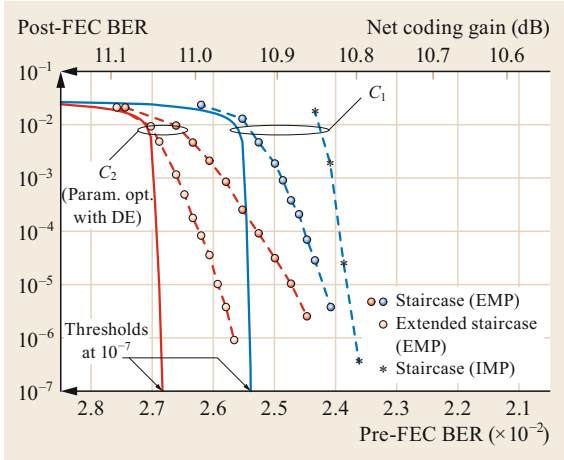


Fig. 7.39 Simulation results (*dashed curves with markers*) and density evolution results (*solid curves*) for staircase codes and extended staircase codes

which is referred to as C_2 in the figure. The component BCH code has parameters $(\nu, t, s) = (8, 3, 63)$, i.e., $(n_c, k_c) = (192, 168)$. As can be seen, the optimization of the code parameters based on density evolution yields better performance in the waterfall region. The poorer slope of the staircase code with component code C_2 than with component code C_1 is due to the fact that for the former, the size of one staircase block, a (Fig. 7.31), is smaller, namely $a = 96$ versus $a = 180$, corresponding to a decoding delay of $D = 64\,512$ bits and $D = 226\,800$ bits, respectively [7.209, Table 1]. Finally, in the figure we also plot the performance of an extended staircase code with block size $a = 192$ [7.209], which has a decoding delay comparable to that of the staircase code based on C_1. The extended staircase code has a steeper curve in the waterfall region.

In Fig. 7.40, we plot simulation results for a staircase code (red curve with markers), a braided code (blue curve with markers), and a half braided code (green curve with markers) with BCH component codes with parameters $(n_c, k_c) = (720, 690)$ and $t = 3$. All three codes have roughly the same code rate, $R \approx 0.917$, corresponding to an FEC overhead of 9.1%. Transmission over a BSC is assumed, and we plot the post-FEC BER as a function of the channel transition probability p. The decoding is performed using a windowed decoder with window size $W = 8$ for the staircase code and $W = 6$ for the braided code, such that the two codes have the same delay, $D = 1\,036\,800$ bits, while the decoding delay of the half-braided code with $W = 6$ is roughly half, $D = 517\,680$ bits. The staircase and braided codes are decoded by iterating $\ell = 8$ times between rows and columns within each window.

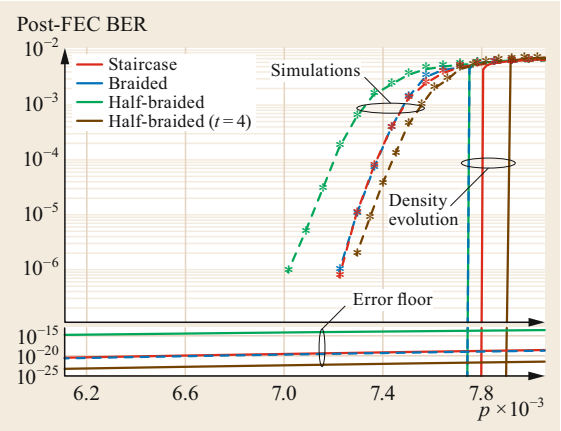


Fig. 7.40 Simulation results (*dashed curves with markers*), density evolution results, and error floor curves for staircase, braided, and half-braided codes

For the half-braided code, all component codes within each window are decoded simultaneously and ℓ is increased to 16 to maintain the same decoding complexity (note that the number of component codes per window is reduced by half for half-braided codes).

In the figure, we also plot density evolution results and the error floor approximation given in (7.91). The density evolution results show a slight performance advantage for the staircase code in the asymptotic regime of large block lengths. However, for the chosen parameters at finite lengths, the simulation curves for the staircase code and the braided code are virtually on top of each other. The density evolution results for the braided code and the half-braided code are roughly the same. However, the simulated half-braided code has worse performance than the staircase and braided codes, caused by different scaling behavior at finite length due to the reduced number of bits within the decoding window. Furthermore, the error floor is increased from $\approx 10^{-20}$ for the staircase and braided codes to $\approx 10^{-14}$ because of the reduction of $s_{\min}$. On the other hand, the half-braided code operates at only half the decoding delay. In the figure, we also plot the performance of a half-braided code with a BCH component code of parameters $(960, 920, 4)$. The code rate is the same, but the decoding delay is now $D = 920\,640$ bits, which is slightly less than that of the staircase and braided codes. The code shows a reduced error floor ($\approx 10^{-23}$) and also improves the waterfall performance, as predicted by density evolution and confirmed by the simulations.

In Fig. 7.41, we plot the BER performance of a PC with double-error-correcting extended BCH (eBCH) codes with parameters $(256, 239, 6)$ as component codes for transmission over the AWGN channel. The code rate of the resulting PC is $R = 239^2/256^2 \approx 0.8716$, which corresponds to overhead OH $\approx 15\%$. In the figure, we plot the performance of both iBDD (red curve with triangle markers) and TPD based on the Chase–Pyndiah algorithm (blue curve with blue markers). Both iBDD and TPD perform about 1 dB from the respective capacity limits at a BER of 10^{-5}. TPD yields a coding gain of about 1.5 dB with respect to iBDD, at a cost of significantly higher decoding complexity. As a reference, we also show the performance of ideal iBDD, where a genie prevents all miscorrections. Additionally, we plot the performance of the

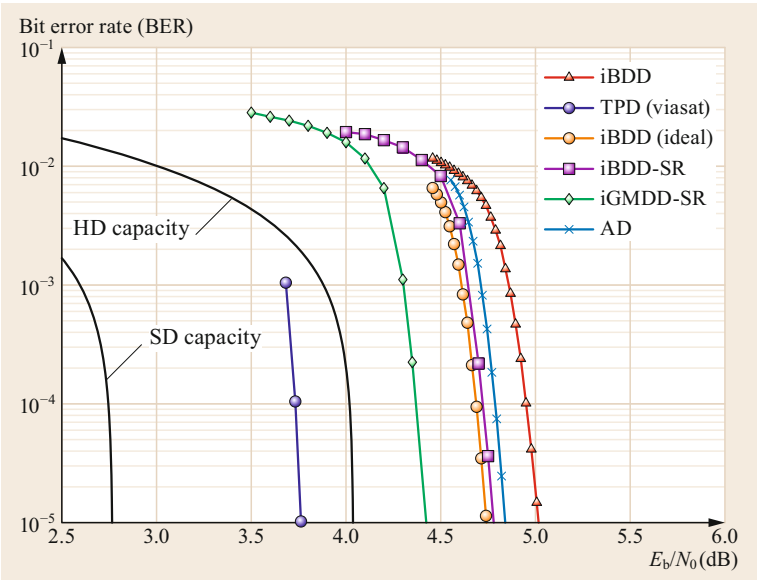


Fig. 7.41 Simulation results for a PC with $(256, 239, 6)$ double-error-correcting eBCH codes as component codes and different decoding algorithms

anchor decoding algorithm proposed in [7.217] and of iBDD-SR [7.223]. We can observe that both anchor decoding and iBDD-SR are effective algorithms for combating miscorrections. The performance degradation of iBDD-SR compared with ideal iBDD is very small (<0.01 dB). iBDD-SR outperforms the conventional iBDD by 0.25 dB at a BER of 10^{-5}, with only a very small increase in complexity [7.223, 225]. Finally, we also plot the performance of iGMDD-SR, which outperforms iBDD, anchor decoding, and iBDD-SR. In particular, the performance gain for iGMDD-SR over iBDD is 0.60 dB. Furthermore, iGMDD-SR performs 0.52 dB from TPD, i.e., it closes over 50% of the performance gap between iBDD and TPD [7.224].

7.6 Coded Modulation—An Introduction

In the previous sections we have focused mostly on binary codes and binary (BPSK) transmission. As we have seen, to operate close to the Shannon limit, (powerful) error-correcting codes are required. Forward error correction (FEC) increases the power efficiency of the system but introduces redundancy to the transmitted sequence. The added redundancy requires the modulator to operate at a higher data rate and hence requires a larger bandwidth, i.e., the spectral efficiency is decreased. Achieving higher spectral efficiency can be realized by using a higher-order signal constellation (i.e., constellations with cardinality larger than 2). Therefore, to transmit reliably at high spectral efficiencies, as required in fiber-optic communications, error-correcting codes must be combined with the use of a higher-order constellation. The combination of error-correcting coding and higher-order constellations is usually referred to as coded modulation (CM).

CM has been studied since the 1970s and comes in different flavors, depending on the code used (binary or nonbinary) and the way the code is coupled with the higher-order constellation. In this section, we briefly review the most important CM schemes, namely trellis-coded modulation (TCM), multilevel coding (MLC), bit-interleaved coded modulation (BICM), and CM with nonbinary codes. We also briefly discuss constellation shaping as a means to close the fundamental gap relative to capacity that the use of conventional signal constellations with equally spaced signal points and uniform signaling entails.

7.6.1 Trellis-Coded Modulation

One of the first CM schemes to appear was TCM, a CM scheme proposed by *Gottfried Ungerboeck* in the late 1970s/early 1980s [7.226, 227] as a way to increase the spectral efficiency of communication systems. It consists of the concatenation of a binary trellis code (a convolutional code) and a higher-order constellation through a careful mapping of the code bits to the constellation symbols based on the set partitioning principle. Due to the correspondence of signal sequences and paths along the trellis of the convolutional code, TCM enables efficient ML decoding using Viterbi decoding.

We have seen in Sect. 7.3.3 that, for transmission over the AWGN channel, the ML rule for SDD is to select from among all possible (modulated) codewords the one at the minimum Euclidean distance to the received vector. However, the mapping of the code bits of a (binary) code optimized in terms of its minimum Hamming distance into constellation points does not guarantee that a good Euclidean distance is obtained. The basic notion of TCM is to optimize the code (now seen as the combination of a binary code and the modulation) in the Euclidean space instead. The signal set (i.e., the constellation) $\mathcal{X}$, of cardinality $M = 2^m$ symbols, is successively partitioned into smaller disjoint subsets of sizes $M/2, M/4, M/8, \ldots$, such that the minimum intra-subset Euclidean distance (the minimum Euclidean distance between signal points within each subset) is maximized. The number of subsets into which the constellation is partitioned depends on the binary code used. The partitioning defines a mapping of binary labels $\mathsf{L}(x) = (b^{(1)}(x), \ldots, b^{(m)}(x))$ to signal points $x \in \mathcal{X}$.

The next step is to assign constellation points to branches of the trellis code. In particular, assume the use of a rate $R = k_1/n_1$, $n_1 \leq m$, convolutional code. In this case, the constellation is successively partitioned into 2^{n_1} subsets, each containing 2^{m-n_1} constellation points. Each block of k data bits is then split into two sub-blocks of lengths k_1 and $k_2 = m - n_1$. The block of k_1 data bits is encoded by the convolutional encoder, generating n_1 code bits, which are used to select one of the 2^{n_1} subsets in which the constellation has been partitioned. The remaining k_2 data bits are used to select one of the points within the selected subset. Note that the presence of uncoded bits introduces parallel transitions in the trellis. Ungerboeck showed that by choosing $n_1 = k_1 + 1$ and $k_2 = 1$ (a single uncoded bit), it is possible to design CM schemes with coding gains of between

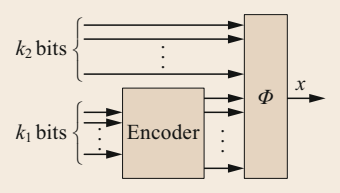


Fig. 7.42 TCM encoder

3 and 6 dB. A block diagram of the TCM encoder is shown in Fig. 7.42.

Finally, the assignment of constellation points to the branches of the code trellis can be done by means of an exhaustive computer search in order to maximize the minimum Euclidean distance of the coded sequences. Ungerboeck introduced a set of rules that are conjectured to give rise to the best TCM schemes:

1. Members of the same partition are assigned to parallel transitions.
2. Members of the next larger partition are assigned to transitions originating from the same state or converging to the same state.
3. All signal points are used equally often.

After its introduction, TCM became very popular and was the object of intensive research. It was also quickly adopted for modem standards in the early 1990s. A thorough description and analysis of TCM can be found in [7.228]. The concept of TCM was later extended to turbo TCM [7.229] by replacing the convolutional code with a turbo code to decrease the gap relative to the Shannon limit for AWGN channels. TCM with multidimensional constellations was proposed in [7.230]. TCM was first proposed for fiber-optic systems in [7.231]. The concatenation of TCM with an outer Reed–Solomon and BCH code was studied in [7.232], yielding NCGs of 8.4 and 9.7 dB, respectively, at a BER of 10^{-13} for the AWGN channel.

7.6.2 Multilevel Coding

In parallel to TCM, an alternative CM modulation scheme, MLC, was proposed by *Hideki Imai* and *Shuji Hirakawa* [7.233] in 1977. The key idea underlying MLC is to protect each bit of the constellation symbol by an individual binary error-correcting code C_i. In other words, MLC transforms the nonbinary channel into a set of m parallel binary sub-channels, and then uses individual binary error-correcting codes for each sub-channel. Specifically, MLC works as follows: A block $\boldsymbol{u} = (u_1 \ldots, u_k)$ of k data bits is partitioned into m blocks $\boldsymbol{u}^{(i)} = (u_1^{(i)}, \ldots, u_{k_i}^{(i)})$ of length k_i bits, $i = 1, \ldots, m$, with $\sum_{i=1}^{m} k_i = k$. Each data block $\boldsymbol{u}^{(i)}$ is then encoded by an individual binary encoder of rate $R_i = k_i/n$, generating codewords $\boldsymbol{c}^{(i)} = (c_1^{(i)}, \ldots, c_n^{(i)})$, of length n bits, where we assume for simplicity and ease of exposition that all codes have equal code length n (however, in principle, the choice of the component codes is arbitrary). The code rate of the overall coding scheme is equal to the normalized sum of the individual code rates, i.e.,

$$R = \frac{1}{m}\sum_{i=1}^{m} R_i = \frac{1}{m}\sum_{i=1}^{m} \frac{k_i}{n} = \frac{k}{nm} \,. \tag{7.92}$$

At each time instant t, the string of m code bits $c_t^{(1)}, \ldots, c_t^{(m)}$ at the output of the m encoders is mapped to a constellation symbol. As for TCM, the mapping is also derived by successively partitioning the signal set into subsets. Note that each component code has a different rate, i.e., each constellation bit is protected differently. A block level of the multilevel encoder is depicted in Fig. 7.43.

ML decoding of MLC requires joint decoding of all component codes, which is unfeasible. Fortunately, decoding of MLC can be efficiently performed using multistage decoding. The multistage decoder operates by decoding the m component codes separately. More precisely, the component codes C_i are decoded sequentially, starting with code C_1. We denote by $\mathcal{D}_i$ the decoder corresponding to component code C_i. At stage i, $i = 1, \ldots, m$, the decoder $\mathcal{D}_i$ is fed the received sequence and also the decisions of the previous decoding stages, i.e., $\hat{\boldsymbol{u}}^{(1)}, \ldots, \hat{\boldsymbol{u}}^{(i-1)}$ (or equivalently $\hat{\boldsymbol{c}}^{(1)}, \ldots, \hat{\boldsymbol{c}}^{(i-1)}$). The block diagram of the multistage decoder is shown in Fig. 7.44. Multistage decoding achieves good performance in practice, with limited complexity, and it has been shown to achieve the channel capacity [7.234].

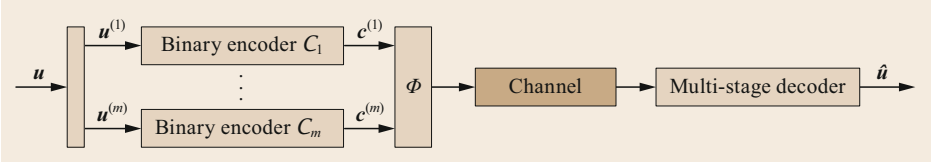


Fig. 7.43 MLC encoder

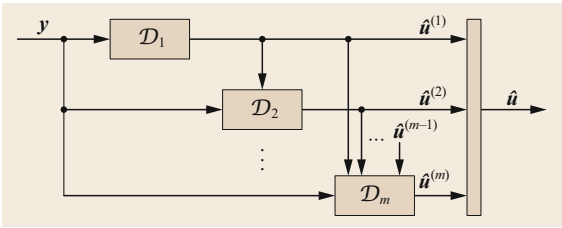


Fig. 7.44 Multistage decoder for MLC

7.6.3 Bit-Interleaved Coded Modulation

BICM was introduced by *Ephraim Zehavi* et al. in 1992 [7.235]. Given its simplicity, flexibility, and capacity-approaching performance, it is nowadays perhaps the most popular CM scheme and has become the de facto standard in many communication standards and modern communication systems.

BICM is a *pragmatic* approach to CM consisting in the concatenation of a (single) binary code with a higher-order constellation through a binary *interleaver* π. The effect of the interleaver is to uniformly distribute the code bits among the different sub-channels corresponding to the different constellation bits. Therefore, in contrast to TCM and MLC, BICM decouples the code from the modulation, which greatly simplifies the CM scheme design (the code is optimized independently of the modulation).

The block diagram of a BICM encoder and decoder is shown in Fig. 7.45. The information sequence $\boldsymbol{u}$ is first encoded by a binary encoder into codeword $\boldsymbol{c}$, which is interleaved by the interleaver π. The interleaver permutes the sequence $\boldsymbol{c}$ into another sequence $\tilde{\boldsymbol{c}}$. The interleaved sequence $\tilde{\boldsymbol{c}}$ is then parsed in blocks of m bits each, thus generating m parallel bit streams, denoted by $\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(m)}$. At time instant i, the modulator Φ takes m bits $(b_i^{(1)}, \ldots, b_i^{(m)})$, and maps them onto symbols of a constellation $\mathcal{X}$ of cardinality $M = 2^m$ according to the binary labeling of the constellation (usually Gray labeling). The resulting (modulated) sequence $\boldsymbol{x}$ is transmitted over the channel. The sequence at the output of the channel is denoted by $\boldsymbol{y}$.

At the receiver side, for SDD, the demodulator Φ^{-1} computes soft reliability information about the bits $(b_i^{(1)}, \ldots, b_i^{(m)})$ of the bitstreams $\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(m)}$ in the form of LLRs

$$l_i^{(j)} \triangleq \log\left(\frac{p_{Y_i|B_i^{(j)}}(y_i|0)}{p_{Y_i|B_i^{(j)}}(y_i|1)}\right) = \log\left(\frac{\sum_{x\in\mathcal{X}_0^{(j)}} p_{Y_i|X_i}(y_i|x)}{\sum_{x\in\mathcal{X}_1^{(j)}} p_{Y_i|X_i}(y_i|x)}\right), \tag{7.93}$$

where $l_i^{(j)}$ is the LLR for the j-th bit at time instant i; X_i, Y_i, and $B_i^{(j)}$ are the RVs corresponding to the channel input x_i, channel output y_i, and bit $b_i^{(j)}$, respectively; and $\mathcal{X}_0^{(j)} \subset \mathcal{X}$ and $\mathcal{X}_1^{(j)} \subset \mathcal{X}$ are the sub-constellations consisting of all constellation points with binary labeling with a 0 or a 1 in the j-th position, respectively, i.e., $\mathcal{X}_z^{(j)} = \{x \in \mathcal{X} : b^{(j)}(x) = z\}$, $z = \{0, 1\}$. The LLRs are then multiplexed, de-interleaved, and fed to a bit-wise soft-decision decoder.

It is also possible to use iterative decoding/demodulation at the receiver. In this case, the decoder and the demodulator exchange extrinsic information in an iterative fashion, and the scheme is referred to as BICM with iterative decoding (BICM-ID) [7.236–238]. Note that, here, *iterative decoding* refers to the iterations between the decoder and the demodulator, and not to the iterative decoding of the binary code. BICM-ID brings performance gains with respect to plain BICM for mappings other than Gray and non-square constellations, at a cost of higher complexity. With Gray labeling and square constellations, there is only a negligible advantage in iterating between the decoder and the demodulator.

For HDD, the demodulator performs minimum distance symbol-by-symbol detection of the received symbols (assuming uniform distribution at the input of the channel) and outputs the binary labeling associated with the detected symbol. The resulting stream of bits (after multiplexing) is de-interleaved and fed to a bit-wise hard-decision decoder.

7.6.4 Coded Modulation with Nonbinary Codes

As discussed in Sect. 7.2.1, the MI of nonbinary transmission over the AWGN channel is achievable with

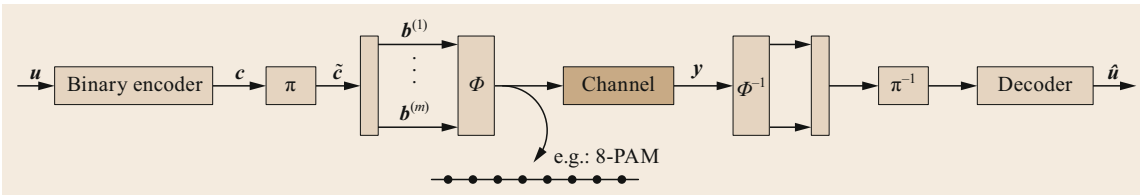


Fig. 7.45 BICM encoder and decoder

a symbol-wise soft-decision decoder. A natural way to couple an error-correcting code with a higher-order constellation is therefore to consider a nonbinary code that is matched to the constellation size. One can then assign each of the 2^m different code symbols to one of the constellation points. Note that in this case, the mapping between code symbols to constellation points becomes trivial. CM with nonbinary LDPC codes is discussed in [7.239].

In practice, the loss with respect to the MI due to the use of a BICM system is very small for most constellations, labeling, and channels in use today. Therefore, since the decoding of nonbinary codes is generally significantly more complex than the decoding of their binary counterparts, it is difficult to justify the use nonbinary codes, and BICM is perhaps the most reasonable CM scheme to consider.

7.6.5 Signal Shaping

So far we have assumed nonbinary transmission using a conventional signal constellation with equidistant signal points and uniform signaling, i.e., each signal point is transmitted with the same probability. This is the case for conventional PAM and quadrature amplitude modulation (QAM) constellations used in current fiber-optic systems. However, such constellations exhibit a gap relative to the Shannon limit. The reason is that conventional constellations are not capacity-achieving. For instance, for the AWGN channel, the capacity-achieving distribution is the Gaussian distribution. The use of a discrete signal constellation with equidistant signal points and uniform signaling instead leads to an asymptotic loss of 1.53 dB (asymptotic in the sense of high spectral efficiencies and number of signal points). In Fig. 7.46, the MI curves (see (7.3)) for 4, 8, 16, 32, 64, 128, and 256-QAM constellations are depicted for transmission over the AWGN channel and compared with the channel capacity curve. A gap with respect to the capacity curve is observed.

Thus, to achieve capacity, the use of powerful error-correcting codes is not enough. To reduce the fundamental 1.53 dB gap, constellation-shaping techniques that produce a Gaussian-like distribution, i.e., *shape* the constellation such that it mimics a Gaussian distribution (or the capacity-achieving distribution for a given channel) are required [7.240]. In other words, to achieve capacity, coding and CM techniques must be complemented with signal shaping.

There are two main classes of signal shaping, geometric shaping and probabilistic shaping [7.241, 242]. In geometric shaping, the constellation points are arranged in a non-equally spaced manner to mimic the capacity-achieving distribution. The best constellation

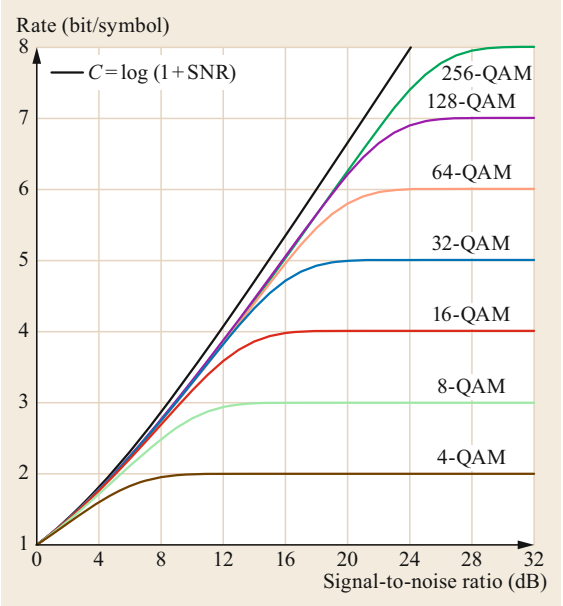


Fig. 7.46 Mutual information curves for several QAM constellations over the AWGN channel. There is a gap between the MI and the capacity of the channel

depends on the SNR and the decoding metric. Probabilistic shaping, on the other hand, starts with a conventional constellation with equidistant signal points, e.g., PAM or QAM, and assigns different probabilities to different constellation points to induce a Gaussian-like probability distribution. In Fig. 7.47, we plot a probabilistically shaped 8-PAM constellation. The inner signal points have higher probability than the outer signal points, mimicking a Gaussian distribution. A significant advantage of probabilistic shaping is that it builds upon off-the-shelf constellations. In general, geometric shaping requires more complex hardware, such as higher-resolution ADCs, than probabilistic shaping. Furthermore, in [7.243] it was shown that for the AWGN channel and bit-wise decoding, geometric

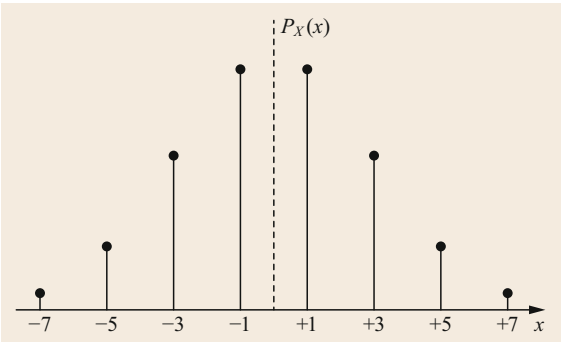


Fig. 7.47 Probabilistically shaped 8-PAM constellation

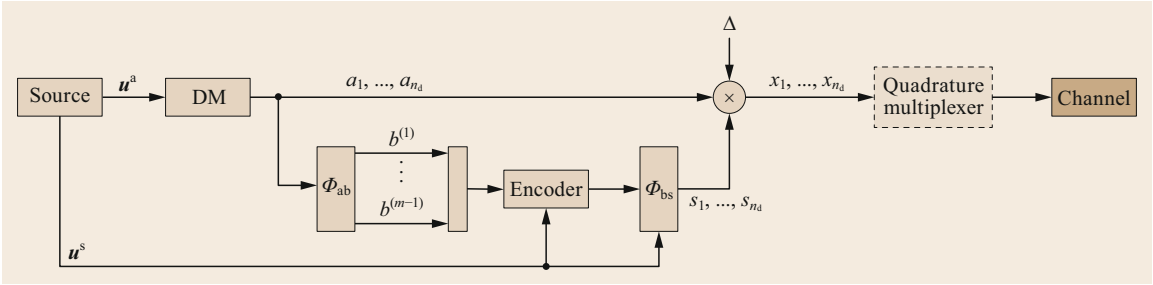


Fig. 7.48 Block diagram of the PAS scheme

shaping suffers from performance loss compared with probabilistic shaping.

Both geometric and probabilistic shaping have been considered for fiber-optic communications as a means of increasing the spectral efficiency, showing significant gains with respect to classical constellations [7.244–250].

Probabilistic Amplitude Shaping

A particularly appealing probabilistic shaping scheme, dubbed probabilistic amplitude shaping (PAS), was recently proposed in [7.251]. It achieves performance within 1.1 dB of the capacity of the AWGN channel for a wide range of spectral efficiencies with off-the-shelf LDPC codes [7.251]. More recently, this scheme was considered for fiber-optic communications in [7.246, 248] and is currently receiving a great deal of attention in the optical communications research community. In [7.249, 250], PAS was applied to HDD and adapted for the use of staircase codes.

PAS exploits the fact that, for the AWGN channel (and symmetric channels in general), the capacity-achieving distribution is symmetric. For one-dimensional and two-dimensional constellations, the capacity-achieving distribution is symmetric around zero. This means that the signs of the constellation points are uniformly distributed. A sketch of the capacity-achieving distribution based on a 8-PAM underlying constellation is depicted in Fig. 7.47. The optimal probability mass function is symmetric around zero, i.e., $\Pr(|x|) = \Pr(-|x|)$, and $\Pr(\mathrm{sign}(X) = 1) = \Pr(\mathrm{sign}(X) = -1) = 1/2$. The key idea in PAS is then to move the shaping of the amplitudes *before* the encoding, as opposed to geometric shaping and conventional probabilistic shaping, and use the (uniformly distributed) parity bits at the output of a systematic encoder and some information bits to generate the signs with the desired uniform distribution.

The PAS scheme proposed in [7.251] is depicted in Fig. 7.48. For simplicity, we consider PAM modulation as the underlying modulation, i.e., the channel input alphabet is given by $\mathcal{X} \triangleq \{-2^m + 1, \ldots, -1, 1, \ldots, 2^m - 1\}$, where m is the number of bits per symbol. The information sequence at the output of the source, $\boldsymbol{u} = (u_1, \ldots, u_k)$, where the information bits are uniformly distributed, i.e., $\Pr(U = 0) = \Pr(U = 1) = 1/2$, is split into two sequences $\boldsymbol{u}^s$ and $\boldsymbol{u}^a$. The sequence $\boldsymbol{u}^a$ is used to generate a sequence of amplitudes $\boldsymbol{a} = (a_1, \ldots, a_{n_d})$ with the desired distribution through a so-called distribution matcher (DM) [7.251]. The distribution matcher transforms blocks of uniformly distributed bits into blocks of amplitudes $a_i \in \{1, \ldots, 2^{m-1}\}$ with the desired probability mass function $P_A(a)$. The binary image of the amplitudes $\boldsymbol{b}$ and the remaining information bits, i.e., $\boldsymbol{u}^s$, are then multiplexed and encoded using a binary code (an LDPC code in [7.251]) with a systematic encoder. The parity bits generated by the systematic encoder and $\boldsymbol{u}^s$ are used to generate n_d sign labels $s_1, \ldots, s_{n_d}$. Since the parity bits at the output of the encoder tend to be approximately uniformly distributed, the distribution of the signs is the desired uniform distribution. Finally, the element-wise multiplication of the sequence of amplitudes $(a_1, \ldots, a_{n_d})$ with the sequence of signs $(s_1, \ldots, s_{n_d})$ and with a scaling factor Δ generates a sequence of symbols $\boldsymbol{x} = (x_1, \ldots, x_{n_d})$ with the desired distribution that is transmitted over the channel. The scaling factor Δ is chosen to meet the power constraint.

Note that since square QAM constellations can be obtained as the Cartesian product of two PAM constellations, the design of the PAS scheme described above readily extends to QAM constellations. In Fig. 7.48, the dashed block multiplexes two PAM modulated sequences, corresponding to the real and imaginary components of the QAM constellation.

7.6.6 Performance Curves

In Fig. 7.49, we give BER results for terminated and tail-biting SC-LDPC codes with SDD and for staircase codes and extended staircase codes with HDD, for an AWGN channel and 64-QAM with BICM. The code rate is $R = 0.75$ for all codes, corresponding to overhead OH = 33%, except for the terminated SC-LDPC

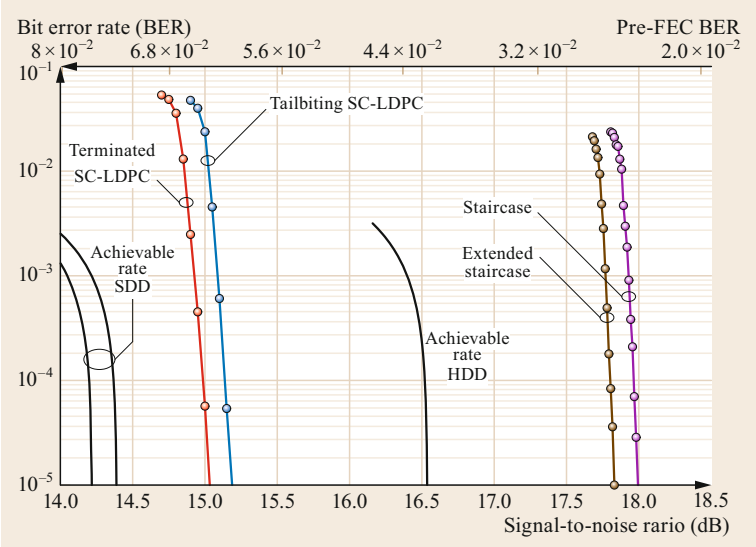


Fig. 7.49 Simulation results (*solid curves with markers*) and achievable rates (*solid curves*) for a BICM system with SC-LDPC codes, a staircase code, and an extended staircase code for transmission over the AWGN channel with 64-QAM

code, for which the code rate is $R = 0.741$. Staircase codes and extended staircase codes perform at 1.5 dB and 1.3 dB from the achievable rate at a BER of 10^{-5}. An extra coding gain of 2.6–2.8 dB can be achieved by using SC-LDPC codes and SDD, at a cost of a higher complexity and power consumption. SC-LDPC codes perform around 0.8 dB from capacity. The fact that the staircase code and the extended staircase code perform further from capacity is due to the moderate code rate. As explained in Sect. 7.5, GPCs perform very close to capacity for higher rates.

In Fig. 7.50, we plot the achievable rates for SDD and HDD with a (conventional) uniform 256-QAM constellation and with PAS assuming an underlying 256-QAM for transmission over the AWGN channel, and compare them with the capacity curve $C = \log(1 + \mathrm{SNR})$. Here, we consider a BICM scheme, and the achievable rates are computed as explained in Sects. 7.2.1 and 7.2.2. It is observed that for SDD, transmitting with uniform 256-QAM (dashed blue curve) entails a shaping loss. The solid blue curve is the achievable rate with PAS, which closes the gap relative to the capacity curve significantly, except for the spectral efficiency range, for which the curve bends to the maximum spectral efficiency with 256 signal points, i.e., 8 bits/symbol. The dashed red curve is the achievable rate with HDD and uniform signaling. HDD entails a loss with respect to SDD, and the loss increases for lower spectral efficiencies, as can be seen from the two different slopes of the dashed curves. As for SDD, PAS yields a performance improvement (cf. red dashed and solid curves). Interestingly, the gap with respect to the corresponding achievable rate for SDD (cf. solid red and solid blue curves) is now reduced, and an almost constant gap of about 1.5 dB is observed. Therefore, shaping seems to help in reducing the gap relative to SDD. In the figure, we also plot the performance of practical staircase codes with parameters in [7.249, 250], which are in agreement with the behavior predicted by the achievable rates.

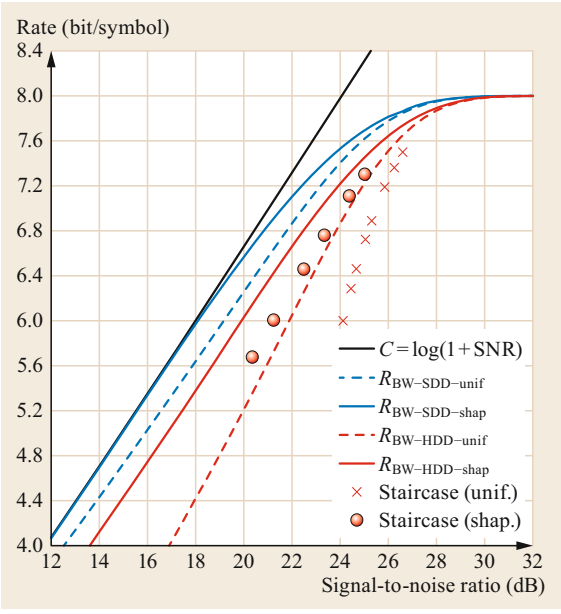


Fig. 7.50 Achievable rates for SDD and HDD with uniform signaling and shaping and performance of a probabilistically shaped CM scheme using staircase codes and comparison with a BICM system using a staircase code and conventional, uniform signaling. 256-QAM, AWGN channel

7.7 Evaluating Forward Error Correction Performance in Transmission Experiments

The design of many optical communication systems often requires the heavy use of transmission experiments to verify models, assumptions, and complete systems. This is largely due to the absence of a rigorous and widely accepted channel model that includes all the effects and impairments of transceivers and optical fibers. Furthermore, in contrast to wireless communications, for example, the fiber-optic communication channel is rather static, enabling relatively easy reproducibility of the experimental results.

Early transmission experiments required specialized hardware such as pattern generators, leading to some constraints on the data to be transmitted. With the absence of ADCs in early optical communication systems, HDD was the dominant decoding method until about 2012 and the advent of commercial systems with SDD [7.252]. Focusing on HDD, a standard solution in system experiments in the 10 and 40 Gbit/s era was to count the bit errors at the receiver (e.g., using a specialized hardware error counter) and to compare this so-called *pre-FEC BER* with the average error-correcting performance of an FEC code with HDD over the BSC. The receiver often consisted of a simple photodiode followed by a high-speed flipflop to sample the photodiode output. This approach assumes that the transmission channel is more or less stationary or is made stationary by a sufficiently large interleaver (to remove, e.g., burst errors). The pre-FEC BER at which the target post-FEC BER was obtained is commonly referred to as the *pre-FEC BER threshold*.

With the advent of ADCs and coherent optical communications, SDD became feasible and quickly became the state-of-the-art decoding scheme in many optical communication systems, especially long-haul and submarine communications. Despite the use of SDD and the absence of a BSC model, the use of a pre-FEC BER threshold was still prevalent in the field for many years. It was first recognized in [7.253] that the pre-FEC BER threshold was not a good performance predictor in systems with differentially detected QPSK followed by an FEC code with SDD. Such systems were dominant in early coherent systems due to a large amount of phase slip from neighboring OOK WDM channels. The authors of [7.253] suggested using a performance prediction threshold that had a better information theoretical foundation and is related to the *achievable rate* of the system (see Sect. 7.2), in that case the MI between code bits and differential detector output.

As coherent optical communication systems matured, and with the introduction of new high-speed digital-to-analog converters (DACs), higher-order modulation formats [7.254], sometimes in combination with probabilistic shaping [7.255], became the state-of-the-art technology for increasing the spectral efficiency of such systems. In transmission experiments, the pre-FEC BER was still often used as threshold despite the fact that the pair $(\mathrm{BER}_{\mathrm{in}}, \mathrm{BER}_{\mathrm{out}})$ does not necessarily fully characterize an FEC code with SDD when the channel, including modulation format, DSP, and quantization, is subject to change. Thus, as previously advocated in [7.253], the authors in [7.256, 257] suggested the use of an information-theoretic measure closely related to the channel and CM scheme utilized. With BICM being the pragmatic choice of CM, this information-theoretic measure is the achievable rate $R_{\mathrm{SDD\text{-}BW}}$ of bit-wise SDD given in (7.8). This quantity is often called GMI, which originates from the theory of mismatched decoding, i.e., the use of a suboptimal decoder [7.87, 88], and is a lower bound of the achievable rate. For BICM, it was shown in [7.97] that the GMI equals the achievable rate $R_{\mathrm{SDD\text{-}BW}}$ ((7.7)–(7.8)). This motivated the authors of [7.256] to refer to $R_{\mathrm{SDD\text{-}BW}}$ as GMI. The term GMI is now ubiquitously used in the field of fiber-optic communications as a proxy for $R_{\mathrm{SDD\text{-}BW}}$. We would like to emphasize here that the GMI is a much broader concept that can be used in a much more general way, and hence we prefer to use the term $R_{\mathrm{SDD\text{-}BW}}$ to denote the achievable rate of BICM.

The authors of [7.256] suggested characterizing a certain FEC code by a pair of achievable rate and output BER $(R_{\mathrm{in}}, \mathrm{BER}_{\mathrm{out}})$ (e.g., $(R_{\mathrm{SDD\text{-}BW}}, \mathrm{BER}_{\mathrm{out}})$), and showed by means of an example based on LDPC codes that this is a fairly good assumption when changing the modulation format but leaving the channel fixed. In [7.258, 259], these results were extended to the case of nonbinary FEC with SDD (i.e., using the pair $(R_{\mathrm{SDD\text{-}SW}}, \mathrm{BER}_{\mathrm{out}})$), another CM scheme that has been promoted for use in fiber-optic communications. At the same time, in [7.259, 260], the use of thresholds as performance predictor in general was questioned, as all threshold-based methods rely on the *universality* of FEC codes, which should be closely checked before employing a threshold-based method.

In this section, we discuss the universality of FEC schemes, introduce the threshold-based performance estimation approach, and discuss the most common thresholds and how to use them in practice. Finally, we show how to best avoid pitfalls of thresholds and how to include FEC in transmission experiments, which gives the most accurate estimates.

7.7.1 Threshold-Based Forward Error Correction Performance Prediction

While thresholds are a perfectly fine tool for predicting the performance of some FEC schemes with HDD (under some stationarity assumptions), the use of FEC schemes with SDD together with varying modulation formats and transmission links requires more caution. This is because SDD relies on knowledge of the probabilistic channel model (Fig. 7.1), which is subject to change in the latter scenarios.

Forward Error Correction Universality

When assessing and comparing the performance of different modulation formats and transmission scenarios (e.g., fiber types, modulators, converters) based on *thresholds*, it is important to understand the concept of FEC *universality*. An FEC code and decoder pair is said to be *universal* if its performance of the code (measured in terms of post-FEC BER or symbol error rate (SER)) does not depend on the channel, provided that the CM scheme is fixed and the achievable rate of the channel is fixed.

When we refer to *the channel*, we consider the whole transmission chain between the FEC encoder output $\boldsymbol{x}$ and the decoder LLR input $\boldsymbol{l}$, including modulation and demodulation, LLR computation, DSP, ADCs and DACs, optical transmission, filtering, and amplification including noise. We say that the channel changes if any of the components in the chain between $\boldsymbol{x}$ and $\boldsymbol{l}$ changes. This can be for instance the noise spectrum or the OSNR, but also the modulation format or the DSP algorithms. However, we assume that the CM scheme (e.g., BICM) is fixed, as it determines the type of threshold to be used.

Unfortunately, not much is known about the universality of practical coding schemes. It is conjectured that many LDPC codes used in practice are approximately universal [7.261], and some LDPC code ensembles (under some relatively mild conditions) have been shown to be universal in the limit of asymptotically large block lengths [7.262]. Guidelines for designing LDPC codes that show good universality properties are highlighted for instance in [7.263]. The class of SC-LDPC codes has also been shown to be asymptotically universal [7.76]. However, the conditions for achieving this asymptotic universality are relatively stringent, and do not allow much to be said about the universality in the non-asymptotic regime, i.e., for finite batch size n_b, finite coupling width w, and finite replication factor L. Polar codes (see Sect. 7.4.3) are examples of *nonuniversal* codes: a polar code needs to be redesigned for every different channel. It is worth pointing out, however, that there exists a modification that renders them universal [7.264]. Unfortunately, the required block length to achieve the same performance becomes considerably larger, limiting the practical application of this scheme.

Although most LDPC codes used in practice are asymptotically universal, we wish to emphasize that practical, finite-length realizations of codes may only be approximately universal. For instance, [7.261, Fig. 3] reveals that the performance of some LDPC codes at a BER of 10^{-4} differs significantly for different channels. This difference is expected to be even larger at very low BERs because of the different slopes of the curves in the waterfall region.

In the following, we define universality of FEC schemes as in [7.263], with the help of Fig. 7.51.

Definition 7.11 FEC Universality

As BICM is ubiquitously used as a CM scheme in fiber-optic communications, we assume BICM and bit-wise SDD. Consider an FEC encoder that generates a codeword consisting of n bits. We transmit these bits over two different communication channels with different (memoryless) channel transition PDFs: channel 1 with PDF $p_{Y_1|X_1}(y_1|x_1)$ and channel 2 with PDF $p_{Y_2|X_2}(y_2|x_2)$. Both channels have *identical* achievable rate $R_{\text{SDD-BW}}$. A fraction γn of the bits is transmitted over channel 1, while the remaining $(1-\gamma)n$ bits are transmitted over channel 2, where $\gamma \in [0, 1]$ such that γn is an integer. We say that a code is *universal* for channels 1 and 2 if the post-FEC BER is independent of γ.

We can extend this definition to i channels (having the same achievable rate) and say that a code is universal if the post-FEC BER is independent of the fraction of bits transmitted per channel and the channels.

In many practical cases, we generally do not experience any issue with universality, as most often the only change made to the channel is a change in the modulation format, while the underlying fiber and noise model

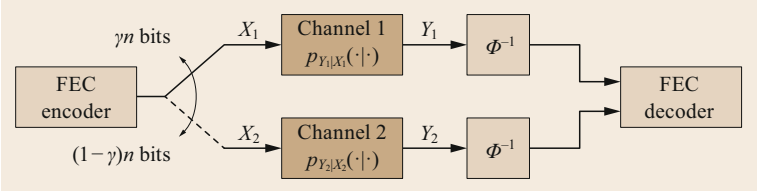


Fig. 7.51 Definition of universality of FEC schemes according to [7.263]. Channels 1 and 2 have the same achievable rate

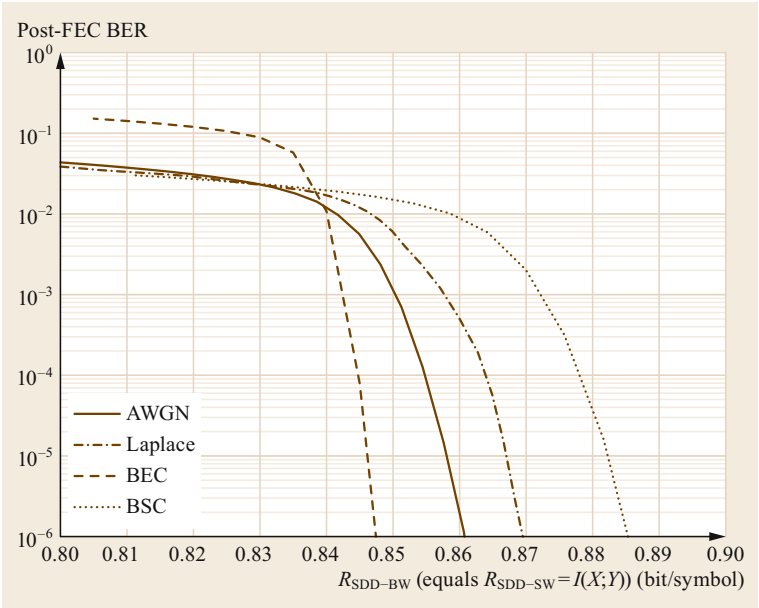


Fig. 7.52 Illustration of the non-universality of an LDPC code with scaled min-sum decoding

remain relatively constant. The common fiber-optic transmission with coherent reception and no in-line dispersion compensation can be accurately modeled as an AWGN channel. In this case, the codes are approximately universal [7.256, 257, 259], and the achievable rate threshold can serve as an accurate prediction. In some cases, however, the changes to the channel can be more drastic. For example, instead of coherent detection, we could use a simple direct detection (DD) scheme that can lead to significantly different noise distributions. On the other hand, the presence of in-line dispersion compensation might lead to significantly different channel statistics that cannot be easily modeled by AWGN.

In [7.259], the impact of a more drastic change in the channel on code universality was shown for the case of CM with nonbinary LDPC codes: a severe quantization was added at the channel output, and the performance (in terms of gap relative to the achievable rate) of the nonbinary LDPC decoder changed significantly in that case. In the following, we illustrate the concept of non-universality of a common FEC code/decoder by means of two different examples.

In the first example, we consider a regular QC-LDPC code of rate $R = 4/5$ with parameters $d_\mathrm{v} = 3$, $d_\mathrm{c} = 15$, $S = 128$, and $n = 38\,400$. We use this code to transmit codewords over four different channels: the previously introduced binary-input AWGN channel, the BSC, the BEC, and the binary-input Laplace channel. The BEC is a channel with binary input $\mathcal{X} = \{0, 1\}$ and ternary output alphabet $\mathcal{Y} = \{0, ?, 1\}$. The channel transition probabilities are $P_{Y|X}(0|0) = 1-\epsilon$, $P_{Y|X}(?|0) = \epsilon$, $P_{Y|X}(1|0) = 0$, and by symmetry, $P_{Y|X}(1|1) = 1-\epsilon$ and $P_{Y|X}(?|1) = \epsilon$. The channel output is either erased (?) with probability ϵ, or equal to the transmitted bit with probability $1-\epsilon$. The binary-input Laplace channel adds noise (per real dimension) according to the Laplacian distribution

$$p_{Y|X}(y|x) = \frac{1}{2b}\exp\left(-\frac{|y-x|}{b}\right) \quad \text{with } b = \frac{1}{2\sqrt{E_\mathrm{s}/N_0}}\,. \tag{7.94}$$

The channel parameters are configured such that all four channels have the same achievable rate $R_{\mathrm{SDD\text{-}BW}}$ (which for this example is equivalent to $R_{\mathrm{SDD\text{-}SW}}$, as the channel input is binary). After transmission over the channel, we compute LLRs using the true respective channel PDF, i.e., the receiver is matched to the channel. We consider a layered decoder employing the scaled min-sum decoding rule ($\alpha = 0.75$) as described in Sect. 7.4.1 with 10 decoding iterations. The simulation results are shown in Fig. 7.52. We observe not only different behavior of the codes, but also different slopes. Because of the different slopes, the difference will be even larger at low BERs. This offset can be attributed to the fact that the LDPC code utilized is not exactly universal and the code length is relatively small, which is an effect that was also observed in [7.261]. Note that if we are allowed to increase the code length and optimize the degree distribution, as highlighted for instance in [7.263], and use the non-simplified CN update rule (instead of the min-sum simplification),

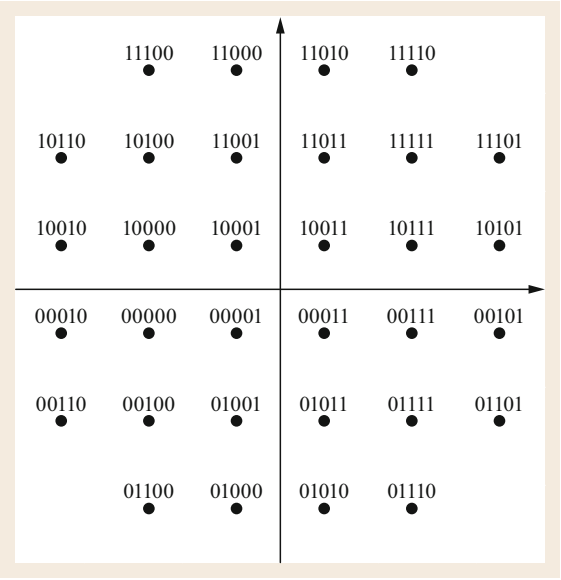


Fig. 7.53 32-QAM constellation with bit mapping maximizing $R_{\text{SDD-BW}}$

the performance prediction becomes more accurate again.

Another example is highlighted in [7.253, Fig. 3], where the authors compare a wide range of different WDM systems consisting of either only coherent 100 Gbit/s channels, or of a mix between coherent and OOK channels. Different fiber types are also used. It is shown that the achievable rate (in that case $R_{\text{SDD-SW}}$) serves as a better predictor than the pre-FEC BER, but itself also fails to accurately predict the performance due to changing channel characteristics. We illustrate the lack of universality with another example [7.260]. In this example, we use BICM together with five different constellations: QPSK, 8-QAM (as in [7.259, Fig. 3, C_3]), 16-QAM, 32-QAM (with the optimized bit labeling shown in Fig. 7.53), and 64-QAM. We consider transmission over both the AWGN and the Laplace channels using the same code as for the example in Fig. 7.52, and at the receiver employ scaled min-sum decoding ($\alpha = 0.75$) with 10 decoding iterations. The results in terms of normalized $R_{\text{SDD-BW}}/m$ are shown in Fig. 7.54. We can clearly see that even in the case of an AWGN channel, the GMI is only an approximately good threshold of the performance, but if the channel law changes (e.g., a Laplace channel is used), the thresholding effect is compromised, and a significantly higher value of $R_{\text{SDD-BW}}$ is required for decoding. This could lead to significantly misleading conclusions, e.g., in terms of reach prediction.

We hence conclude that performance predictors based on the achievable rate should be used with caution. They can still give rough first-order estimates of the decoding performance, even if we introduce drastic changes into the channel (e.g., a strong quantization, moving from dispersion-uncompensated to dispersion-compensated links, or even from coherent transmission to direct detection systems). Thresholds should therefore be used only to quickly assess the performance and to determine the range of fine measurements. We can improve the accuracy if the channel used to compute the threshold is fairly close to the actual channel of the

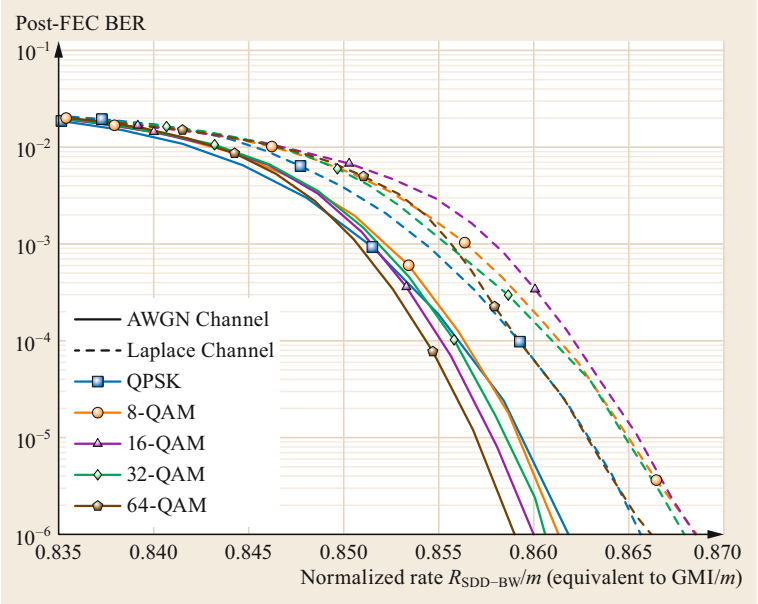


Fig. 7.54 Illustration of the non-universality with different modulation formats and BICM

system. However, in all cases, actual decoding mimicking as closely as possible the true FEC should be used.

Summary of Performance Prediction Thresholds and Usage Guidelines

In this section, we present some of the thresholds that are commonly used today, provide guidelines for their usage, and discuss some potential pitfalls. We assume that a transmission experiment has been carried out and that for a particular setup, we have a measurement consisting of N_M data points (x_κ, y_κ) that are given as the original transmit sequence $x_\kappa \in \mathbb{C}$ and the corresponding received sequence $y_\kappa \in \mathbb{C}$, both consisting of complex modulation symbols. To keep exposition simple, we do not consider 4-D constellation symbols (e.g., dual-polarization modulation formats), but assume an independent treatment of both polarizations. The computation of the achievable rate for 4-D constellations has been addressed in [7.265]. We further assume that both sequences are aligned and that eventual phase jumps have been taken care of (e.g., using pilot symbols). The sequence recovery can be performed either by cross-correlating some or all of the received symbols with the transmit symbols or, if a pseudo-random binary sequence (PRBS) is transmitted, by utilizing a simple PRBS synchronization algorithm.

Pre-FEC SER. The pre-FEC SER is perhaps the easiest to compute directly from the experimental measurement database. First, we require a demodulation function. Often, when the channel is Gaussian-like, the nearest-neighbor demodulator (also called Euclidean distance demodulator) is used. The nearest-neighbor demodulator computes estimates of the transmit sequence as

$$\hat{x}_\kappa = \arg\min_{s\in\mathcal{X}} \|y_\kappa - s\|^2 \,. \tag{7.95}$$

Note that the Euclidean distance rule is only optimal for Gaussian noise and should be modified accordingly if the channel behaves differently (e.g., when strong phase noise is present). Using the estimated symbol sequence $\hat{\boldsymbol{x}}$, the pre-FER SER is estimated as

$$\mathrm{SER}_{\mathrm{pre}} = \frac{1}{N_M} \sum_{\kappa=1}^{N_M} \mathbb{1}_{\{\hat{x}_\kappa \neq x_\kappa\}} \,. \tag{7.96}$$

The Pre-FEC SER should be used only as a threshold in particular cases, e.g., when symbol-wise HDD is used and the constellation size is matched to the symbol size of the code. In this case, for a fixed constellation size M, the achievable rate $R_{\text{HDD-SW}}$ given by (7.15) is directly linked to the symbol error probability (therein denoted δ), to which the SER converges as $N_M \to \infty$, and has no further dependence on the modulation format (except its size M).

Pre-FEC BER. The pre-FEC BER was once the most widespread threshold function and still is for FEC schemes with bit-wise HDD. Unfortunately, in many experimental publications, it is not explicitly stated how the pre-FEC BER is computed. Especially if higher-order constellations are used, different computation rules can lead to (slightly) different results.

The computation of the pre-FEC BER is only marginally more complicated than the computation of the pre-FEC SER. Possibly the easiest method is to start from the estimated symbols $\hat{x}_\kappa$, computed using (7.95), and to use these to recover the bit patterns by applying the binary labeling function $\mathsf{L}(\hat{x}_\kappa)$, assuming *implicitly* the use of BICM as CM technique. We denote by $\mathsf{L}(\hat{x}_\kappa) = (\hat{b}_\kappa^{(1)}, \ldots, \hat{b}_\kappa^{(m)})$ the bits that are assigned to the modulation symbol $\hat{x}_\kappa$, and similarly, we denote by $\mathsf{L}(x_\kappa) = (b_\kappa^{(1)}, \ldots, b_\kappa^{(m)})$ the transmit bits associated with x_κ. The pre-FEC BER $\mathrm{BER}_{\mathrm{pre}}$ is then given by

$$\mathrm{BER}_{\mathrm{pre}} = \frac{1}{mN_M} \sum_{\kappa=1}^{N_M} \sum_{i=1}^{m} \mathbb{1}_{\{b_\kappa^{(i)} \neq \hat{b}_\kappa^{(i)}\}} \,. \tag{7.97}$$

An alternative method for computing the pre-FEC BER is to compare the sign of the LLR $l_\kappa^{(i)}$ given by the bit-wise demodulator Φ^{-1} with the transmit bit sequence. The two approaches lead to different results. However, for most practical cases and constellations, the differences are negligible.

The pre-FEC BER should be used *only* if bit-wise HDD is utilized and the channel is sufficiently interleaved to remove all burst errors. In this case, the achievable rate $R_{\text{HDD-BW}}$ is directly linked to the average bit error probability (therein denoted $\bar{\epsilon}$), of which the pre-FEC BER is an estimate. The pre-FEC BER threshold can be used in some circumstances with SDD, but only if the FEC that is evaluated has been thoroughly simulated with a model that is sufficiently close to the experimental setup (e.g., using the same modulation format, quantization, fiber model, neighboring channel setup). Only in this case is the use of pre-FEC BER legitimate. Given the effort required to set up such an involved simulation, it may be easier to include the FEC code directly in the experiment.

Generalized Mutual Information. In the field of fiber-optic communications, the use of the GMI as decoding threshold has become dominant in the recent years. We would like to point out again that the notion of GMI is a much broader concept, introduced as

a bound on the achievable rate for mismatched decoding [7.87, 88]. In the case of BICM, the GMI equals the achievable rate $R_{\text{SDD-BW}}$. Hence, the term GMI is often used in a somewhat inaccurate way, interchangeably with the achievable rate of BICM.

For computing the GMI, we assume that all constellation symbols are equiprobable. There are multiple ways to compute the GMI [7.266]. We give here the one that we believe is easiest. In a first step, we use the bit-wise demodulator Φ^{-1} given by (7.93) to compute m LLRs per received symbol, i.e., $\Phi^{-1}(y_\kappa) = (l_\kappa^{(1)}, \ldots, l_\kappa^{(m)})$. Similarly, we compute the corresponding bit patterns of the transmit sequence using $\mathsf{L}(x_\kappa) = (b_\kappa^{(1)}, \ldots, b_\kappa^{(m)})$. The *bit-wise* GMI threshold can then be computed directly from the LLRs as

$$\text{GMI}_{\text{SDD-BW}} = \underbrace{\log_2 |\mathcal{X}|}_{=m} - \frac{1}{N_M} \inf_{s\geq 0} \sum_{i=1}^{m} \sum_{\kappa=1}^{N_m} \log_2\left(1 + \mathrm{e}^{s(-1)^{b_\kappa^{(i)}} l_\kappa^{(i)}}\right). \tag{7.98}$$

The computation includes a minimization (inf) over the variable $s \geq 0$. When the computation of the LLRs (7.93) is carried out using *exactly* the same channel PDF $p_{Y_i|B_i^{(j)}}(y_i|b_i^{(j)})$ that was used for transmission (e.g., in a simulation), the minimum is obtained for $s = 1$, and the minimization does not need to be carried out. In any other case, due to the unimodality of the objective function, the minimization can be easily carried out using, for example, the golden section search or built-in minimization functions of numerical software packages (e.g., `fminbnd` of MATLAB®).

The threshold $\text{GMI}_{\text{SDD-BW}}$ is an estimate of the achievable rate $R_{\text{SDD-BW}}$ and hence should be used *only* for FEC with bit-wise SDD (i.e., in the case of BICM) and if there is sufficient evidence that the code is approximately *universal* or, alternatively, if the GMI threshold of the code has been determined in a simulation mimicking sufficiently closely the transmission experiment, such that the code's lack of universality does not cause a difference.

PAS has recently become an attractive solution for realizing probabilistic constellation shaping and has found widespread use in optical communications. PAS is not a BICM scheme, but still uses a bit-wise demodulator Φ^{-1} that computes LLRs according to

$$l_i^{(j)} \triangleq \log\left(\frac{\sum_{x\in\mathcal{X}_0^{(j)}} p_{Y_i|\mathsf{X}_i}(y_i|x)P_X(x)}{\sum_{x\in\mathcal{X}_1^{(j)}} p_{Y_i|\mathsf{X}_i}(y_i|x)P_X(x)}\right), \tag{7.99}$$

with $P_X(x)$ being the prior distribution of using constellation symbol x. Applying the LLR $l_i^{(j)}$ given in (7.99) in (7.98) yields a value that we denote as $\bar{R}_{\text{SDD-BW}}$ and that was used in [7.267, Eq. (6)] to compute the so-called normalized generalized mutual information (NGMI), which has been shown to serve as a relatively accurate performance threshold. Note, however, that the NGMI should not be confused with the notion of GMI (despite the similarity in the name). For details regarding achievable rates of PAS, we refer the interested reader to [7.91].

Mutual Information. The estimation of the MI threshold $\text{MI}_{\text{SDD-SW}}$ can be subdivided into two steps. In a first step, we estimate the actual channel PDF $p_{Y|X}(y|x)$ as closely as possible neglecting potential memory effects. We denote the estimated channel PDF by $q_{Y|X}(y|x)$. The PDF can be modeled using kernel density estimators, histograms, and piecewise approximations, or using Gaussian mixture models (GMMs). In the latter case, using g mixtures, for every constellation symbol $\mathsf{X}_i \in \mathcal{X}$, the estimate is written as

$$q_{Y|X}(y|X = \mathsf{X}_i) = \sum_{j=1}^{g} w_{i,j} \mathcal{CN}(\mu_{i,j}, \sigma_{i,j}^2),$$
$$\text{with } \mathcal{CN}(\mu, \sigma^2) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{\|y-\mu\|^2}{\sigma^2}\right), \tag{7.100}$$

i.e., $\mathcal{CN}(\mu, \sigma^2)$ is the complex, circularly symmetric Gaussian PDF with (complex) mean μ and variance σ^2. The PDF $q_{Y|X}$ is parameterized by Mg 4-tuples $(w_{i,j}, \text{Re}\{\mu_{i,j}\}, \text{Im}\{\mu_{i,j}\}, \sigma_{i,j}^2)$ that need to be estimated from the measurement. The $4Mg$ parameters of the model can be estimated for instance using the expectation-maximization (EM) algorithm or built-in estimators of numerical computing environments (e.g., `fitgmdist` of MATLAB®). The MI threshold can then be estimated in a similar way as the GMI

$$\text{MI}_{\text{SDD-SW}} = \frac{1}{N_m} \sup_{s\geq 0} \sum_{\kappa=1}^{N_m} \log_2\left(\frac{M\left[q_{Y|X}(y_\kappa|x_\kappa)\right]^s}{\sum_{x'\in\mathcal{X}}\left[q_{Y|X}(y_\kappa|x')\right]^s}\right). \tag{7.101}$$

The computation includes a maximization (sup) over the variable $s \geq 0$, which can be easily carried out due to the unimodality of the objective function. If $q_{Y|X}(y|x)$ matches the true channel PDF $p_{Y|X}(y|x)$, the maximum is obtained for $s = 1$.

Often, especially in dispersion-uncompensated coherent fiber-optic communications, the GMM is not necessary, and a 2-D Gaussian PDF approximates the

true channel PDF sufficiently well as predicted by the GN model [7.99, 100] and its extensions. Such a 2-D Gaussian PDF yields good estimates of the MI, neglecting any cross-polarization and memory effects. In some circumstances, a 4-D Gaussian PDF can also be used [7.259, 265]. In optical receivers, it is usually not feasible to implement the GMM of the noise, and a complex circularly symmetric Gaussian PDF is usually used as a simple estimate of the true PDF. In this case, we can use a GMI-like argument to compute the threshold

$$\mathrm{MI}_{\text{SDD-SW}} = \frac{1}{N_m} \sup_{s\geq 0} \sum_{\kappa=1}^{N_m} \log_2 \left(\frac{M \exp(-s\|y_\kappa - x_\kappa\|^2)}{\sum_{x'\in\mathcal{X}} \exp(-s\|y_\kappa - x'\|^2)} \right). \tag{7.102}$$

Note that the maximization (sup) over s implicitly estimates the variance of the Gaussian PDF as $\sigma^2 = s_{\max}^{-1}$ [7.259, Example 2].

The MI threshold $\mathrm{MI}_{\text{SDD-SW}}$, which approximates $R_{\text{SDD-SW}}$, should be used only in cases where $R_{\text{SDD-SW}}$ is actually an achievable rate of the transmission system. This includes the use of nonbinary codes, e.g., nonbinary LDPC codes matched to the modulation format as CM scheme, with SDD. The MI threshold can also be used when multilevel coding with multistage decoding is employed as CM scheme. However, the latter is an example of a nonuniversal CM scheme.

Note that in many cases, the values of $\mathrm{MI}_{\text{SDD-SW}}$ and $\mathrm{GMI}_{\text{SDD-BW}}$ are relatively close (e.g., BICM with square QAM constellations and Gray coding, and PAS [7.255]). In these cases, the GMI threshold $\mathrm{GMI}_{\text{SDD-BW}}$ can be replaced by $\mathrm{MI}_{\text{SDD-SW}}$ without losing too much accuracy. This has advantages, as the MI is often easier to compute in the experiment. Computing the MI does not require the implementation of a bit-wise demapper, nor does it need the definition of a bit mapping. The MI can be directly computed from the complex samples at the output of the transmission experiments. In [7.268], MI and GMI thresholds were compared for a field system based on PAS, and were found to be very close.

7.7.2 Implementing Forward Error Correction in Experiments

In view of the pitfalls inherent in the use of thresholds, especially when there is uncertainty as to the universality of the FEC and CM scheme used, we suggest including FEC in the transmission experiments. In particular, the use of offline DSP in modern transmission experiments facilitates the inclusion of at least a basic offline FEC decoding in the evaluation phase of the experiments. This is additionally made easy by the availability, for example, of integrated LDPC encoding and decoding routines in common computational software such as MATLAB®. Essentially, there are two main possibilities for including FEC with actual decoding in transmission experiments, which we discuss in the following.

Implementing Complete Encoding and Decoding

The closest approximation to the true system performance is achieved by implementing the complete physical layer including FEC encoding and decoding. For this, we use an encoder that generates codewords $\boldsymbol{c}$ (for instance using random information words $\boldsymbol{u}$) which are then fed to the transmitter. After reception, DSP is carried out, and following frame alignment, decoding can be performed. The recovered information words can be compared with the original information words to estimate the post-FEC BER. While this approach should be used whenever possible, it has some drawbacks:

- In some transmission experiments, no programmable DAC is available, and the transmit sequence cannot be freely chosen, but is fixed and limited to some PRBS. In other cases, the memory depth of the DAC may be limited, such that a whole codeword cannot fit into the memory (especially if long SC-LDPC codes with large replication factor L are chosen).
- It is often possible that the code design will not yet be completed when transmission experiments are carried out, and the code designers would like to use the results of the transmission experiment to estimate channel characteristics in order to best adapt their code to the channel. This chicken-and-egg problem requires other approaches.

In both cases, we need to decouple the transmission experiment from the evaluation of the FEC performance. This is highlighted in what follows.

Implementing Decoding Using a Database of Measurements

In [7.269], the authors suggested reusing a database of measurements to evaluate the performance of multiple FEC schemes. Essentially the same method was later proposed in [7.270], which however requires a modification of the decoder and introduces some extra interleaving to enlarge the set of sequences, neglecting possible memory effects.

In the following, we describe the method reported in [7.269] that integrates the evaluation of FEC into the

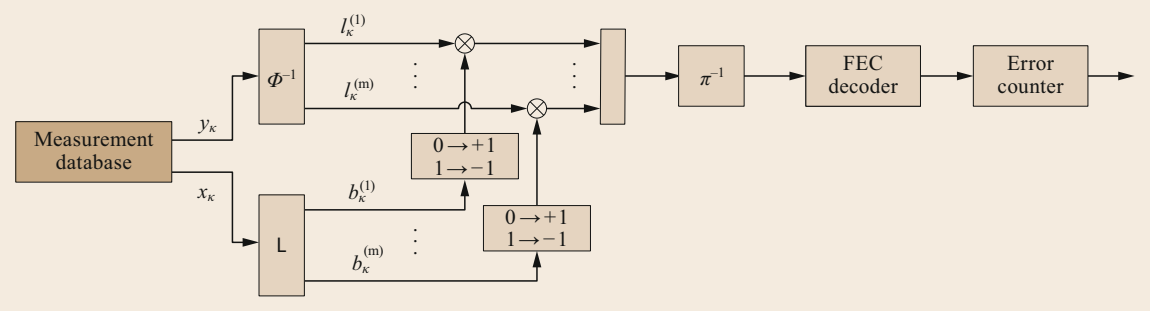


Fig. 7.55 Evaluating FEC performance from measurements for a BICM-based CM system

receiver processing chain in order to evaluate the performance of different codes. This method is based on the fact that the post-FEC BER performance of most practically employed FEC codes and, most importantly, their decoders, does *not* depend on the transmitted codeword, but only on the noise pattern. This property holds for LDPC and SC-LDPC codes together with the decoders that we introduced in Sect. 7.4 (see [7.111, Chap. 4] for a detailed discussion).

Our method is based on channel adapters introduced in [7.271] for analyzing CM schemes. In order to devise a strategy for performance assessment, we assume that the coding scheme to be tested generates one (or several) valid codewords. For simplicity, if linear codes are used, we can use the all-zero codeword $\boldsymbol{c} = (0, 0, \ldots, 0)$, which belongs to any linear code.

We explain the method for the case with BICM as CM scheme, but we stress that the method can be easily extended to other CM schemes (e.g., PAS) as well. The method is illustrated in Fig. 7.55. The first step of the method consists in generating an equivalent bitstream of length mN_M corresponding to the transmit sequence from the experimental database by

$$\begin{aligned} &\boldsymbol{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_{N_M}), \\ &\quad \text{with } \boldsymbol{b}_\kappa = (b_\kappa^{(1)}, \ldots, b_\kappa^{(m)}) = \mathsf{L}(x_\kappa)\,. \end{aligned} \tag{7.103}$$

Similarly, we compute a sequence of LLRs for the received samples y_k,

$$\begin{aligned} &\boldsymbol{l} = (\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_{N_M}), \\ &\quad \text{with } \boldsymbol{l}_\kappa = (l_\kappa^{(1)}, \ldots, l_\kappa^{(m)}) = \Phi^{-1}(y_\kappa)\,. \end{aligned} \tag{7.104}$$

Using both sequences, we generate a set of equivalent LLRs, corresponding to the transmission of the all-zero codeword as

$$\begin{aligned} \tilde{\boldsymbol{l}} &= (\tilde{\boldsymbol{l}}_1, \ldots, \tilde{\boldsymbol{l}}_{N_M}) \\ &\quad \text{with } \tilde{\boldsymbol{l}}_\kappa = \left(\tilde{l}_\kappa^{(1)}, \ldots, \tilde{l}_\kappa^{(m)}\right) \\ &\qquad = \left((-1)^{b_\kappa^{(1)}} l_\kappa^{(1)}, \ldots, (-1)^{b_\kappa^{(m)}} \tilde{l}_\kappa^{(m)}\right). \end{aligned} \tag{7.105}$$

This approach corresponds to transmitting the all-zero codeword $\boldsymbol{c}$, which is scrambled using a (time-varying, data-dependent) binary scrambler. This scrambler generates, by modulo-2 addition (i.e., XOR) of a scrambling sequence, the desired transmit bit sequence $\boldsymbol{b}$. This sequence corresponds to the sequence which, after modulation, yields the transmitted symbol sequence used in the experiment.

The sequence $\tilde{\boldsymbol{l}} = (\tilde{l}_1, \ldots, \tilde{l}_{mN_M})$ now consists of mN_M LLRs that can be fed to an FEC decoder. If $mN_M \gg n$, with n being the codeword length, we can partition the sequence $\tilde{\boldsymbol{l}}$ into subsequences of length n, by selecting $\tilde{n} \triangleq \lceil n/m \rceil$ consecutive vectors $\tilde{\boldsymbol{l}}_i$ and permuting the resulting sequence of LLRs with a de-interleaver π^{-1}. Note that in BICM with LDPC codes, interleaving is sometimes not explicitly carried out but assumed to be implicitly part of the code. However, when QC-LDPC codes are used, care must be taken, and we advise the reader to always explicitly carry out interleaving. The de-interleaved vectors are then fed into the FEC decoder, $f_{\text{FEC,dec}}$, which outputs a binary sequence corresponding to the estimated codeword $\hat{\boldsymbol{c}}$. The post-FEC BER BER_{post} is then obtained by counting the errors after decoding a sufficiently large number of such frames.

Often, the number of possible measurements is not large enough to obtain sufficiently accurate post-FEC BER estimates. In this case, we can extend the estimation using the interleaver method proposed in [7.270]. However, note that this method assumes that at the receiver, long interleaving across multiple codewords is carried out, breaking all correlation between neighboring symbols that arise due to the interplay of chromatic dispersion and fiber nonlinearities. Such a long interleaving will mask any nonstationary effects and may thus lead to overly optimistic performance estimates. Short burst errors, which can be caused for instance by DSP algorithms that cannot track phase or polarization changes during transmission, and cause uncorrectable blocks (UCBs), will not be captured by such an approach. Our advice is therefore to record sufficiently

long sequences and treat the sequences like a continuous bitstream.

7.7.3 Performance Example

To experimentally verify the proposed method, we consider a system using 8-QAM together with BICM. We reuse the measurement database that was used in [7.259] to evaluate nonbinary LDPC codes. We use two 8-QAM constellations $\mathcal{X}_1$ and $\mathcal{X}_2$, shown in Fig. 7.56b and c, together with the bit labeling that maximizes the GMI [7.272]. For illustration purposes, we use three quasi-cyclic regular LDPC codes with rates 0.8, 0.85, and 0.9 (corresponding to FEC overheads of $\approx 25\%$, 18%, and 11%, respectively) with VN degree $d_v = 3$, CN degree $d_c \in \{15, 20, 30\}$, and lifting factor $S = 128$. Each code has length $n = 38\,400$, i.e., 12 800 8-QAM symbols are always mapped to one LDPC codeword. Decoding takes place using $I = 10$ iterations with a layered scaled min-sum decoder ($\alpha = 0.75$). We first test the performance of the three LDPC codes in an AWGN channel. To this end, we first calculate $R_{\text{SDD-BW}}$ for the two constellations $\mathcal{X}_1$ and $\mathcal{X}_2$ as a function of E_s/N_0. This is achieved, for example, by applying the estimator given in (7.98) to the results of a simulation over an AWGN channel.

In Fig. 7.56a, we show the post-FEC BER as a function of $R_{\text{SDD-BW}}$. Changing the constellation for a given code can be interpreted as changing the equivalent channel. We see that the codes act approximately universally, and their performance only marginally depends on the chosen constellation. For a post-FEC BER of 10^{-4} (horizontal line), we find the equivalent achievable rate thresholds T_R for each rate.

To validate the AWGN results in Fig. 7.56a, we now consider a coherent fiber-optic communication system based on dual-polarization transmission at a symbol rate of 41.6 Gbaud. First, symbol sequences corresponding to constellation $\mathcal{X}_1$ are generated and tested using a high-speed DAC in a back-to-back configuration. A root-raised cosine pulse shaping signal (roll-off factor 0.1) is generated as described in [7.272], and two code rates ($R = 0.8$ and $R = 0.85$) are chosen, giving net data rates of approximately 200 and 212 Gbit/s, respectively.

The empirical achievable rate estimates $\text{GMI}_{\text{SDD-BW}}$, computed using (7.98), are shown as a function of the OSNR in Fig. 7.57a. We also show the GMI thresholds $T_{0.8} = 2.57$ and $T_{0.85} = 2.69$. These thresholds are then used to determine equivalent OSNR thresholds (vertical lines). The measured data are then used to perform LDPC code decoding using the method described in Sect. 7.7.2. The obtained results are shown in Fig. 7.57b by solid markers. Additionally, from the estimated achievable rate values, we interpolate the estimated post-FEC BER values using the AWGN simulation results from Fig. 7.56a, which are given by the thin dotted lines. We observe an agreement between the predicted post-FEC BER and actual post-FEC BER. However, we also note that the performance of the actual decoding deviates from the prediction, especially at low error rates. This is mostly due to some non-stationarity of the measurements, with some frames that are more affected by the noise than others and cannot be reliably decoded.

In order to show that the proposed method also works for a transmission over a link, we apply the method to a transmission experiment using both constellations

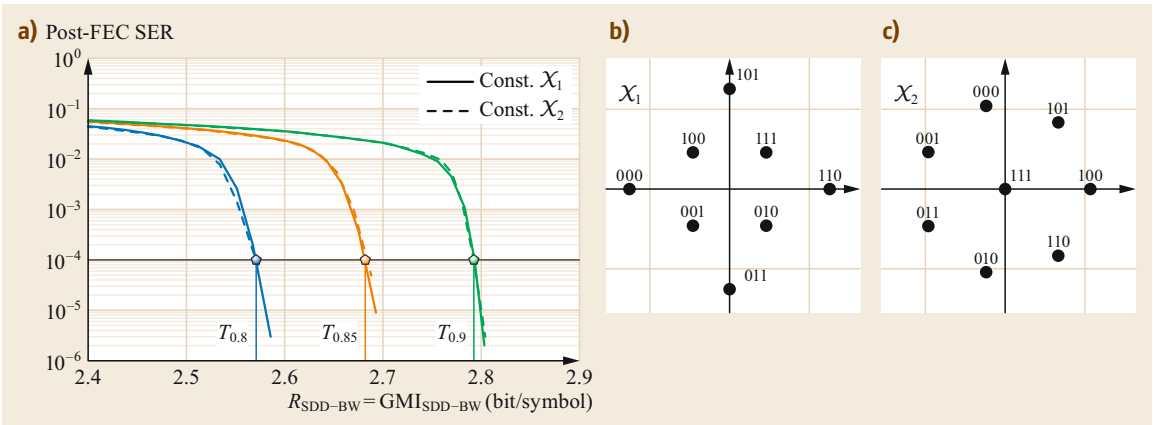


Fig. 7.56a–c AWGN simulation results of regular LDPC codes of rates $R \in \{0.8, 0.85, 0.9\}$ with $d_v = 3$ and $d_c \in \{15, 20, 30\}$ with layered normalized min-sum decoding ($\alpha = 0.75$) and $I = 10$ decoding iterations **(a)**. Constellations utilized with GMI-maximizing bit labeling (b_1, b_2, b_3) **(b,c)**. **(a)** Simulation results; **(b)** constellation $\mathcal{X}_1$; **(c)** constellation $\mathcal{X}_2$

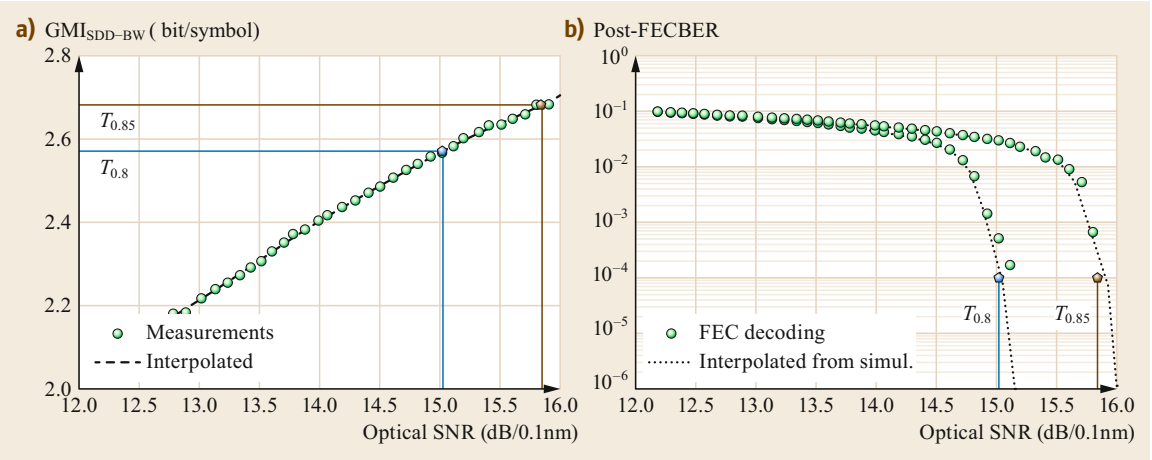


Fig. 7.57a,b Back-to-back results. Empirically estimated values of $\mathrm{GMI}_{\mathrm{SDD\text{-}BW}}$ and interpolated values together with the thresholds to achieve a post-FEC BER of 10^{-4} with ($d_v = 3$, $d_c \in \{15, 20\}$) regular LDPC codes and layered scaled-min-sum decoding ($\alpha = 0.75$) **(a)** and interpolated post-FEC BER (from AWGN simulations with the same setup as that of Fig. 7.54) together with actual decoding **(b)**

X_1 and X_2 over a recirculating loop, described in detail in [7.273]. We recapitulate the experimental setup in the following. The transmission testbed consists of one narrow-linewidth laser under test at 1545.72 nm, and additionally 63 loading channels spaced by 50 GHz. The output of the laser under test is sent into a dual-polarization I/Q modulator driven by a pair of DACs operating at 65 Gsamples/s. Multiple delayed-decorrelated sequences of 2^{15} bits are used to generate the multilevel drive signals. Pilot symbols and a sequence for frame synchronization are additionally inserted.

The symbol sequences are oversampled by a factor of ≈ 1.56 and pulse-shaped by a root-raised cosine function with roll-off of 0.1. The load channels are separated into odd and even sets of channels and modulated independently with the same constellation as the channel under test using separate I/Q modulators. Odd and even sets are then polarization-multiplexed by dividing, decorrelating, and recombining through a polarization beam combiner with an approximate delay of 10 ns. The test channel and the loading channels are passed into separate low-speed (< 10 Hz) polarization scramblers and spectrally combined through a wavelength-selective switch (WSS). The resulting multiplex is boosted through a single-stage EDFA and sent into the recirculating loop. The loop consists of four 100 km-long dispersion-uncompensated spans of standard single-mode fiber with hybrid Raman–EDFA optical repeaters to compensate for the fiber loss. The Raman pre-amplifier is designed to provide ≈ 10 dB on–off gain. Loop-synchronous polarization scrambling is used, and power equalization is performed by a 50 GHz-grid WSS inserted at the end of the loop.

At the receiver side, the channel under test is selected by a tunable filter and sent into a dual-polarization coherent receiver feeding four balanced photodiodes. Their electrical signals are sampled at 80 Gsamples/s by a real-time digital oscilloscope with 33 GHz electrical bandwidth. For each measurement, five different sets of 20 µs are stored. The received samples are processed offline by a DSP, including chromatic dispersion compensation, polarization demultiplexing by a 25 tap $T/2$ spaced butterfly equalizer with blind adaptation based on a multi-modulus algorithm, frequency recovery using the fourth and seventh power (depending on the constellation) periodogram, and phase recovery using the blind phase search algorithm. In the latter, equally spaced test phases in the interval $[-\pi/4; \pi/4]$ (constellation X_1) or in the interval $[-\pi/7; \pi/7]$ (constellation X_2) are used with correspondingly modified phase unwrapping.

We consider transmission over a distance of 3200 km, corresponding to eight round trips in the recirculating loop. Figure 7.58a shows the estimated $\mathrm{GMI}_{\mathrm{SDD\text{-}BW}}$ as a function of the input power P_{in} per WDM channel [7.273, Fig. 3a]. Additionally, we show the achievable rate thresholds T_R for $R \in \{0.8, 0.85, 0.9\}$. The thresholds give us the region of launch powers at which reliable transmission is possible. To be precise, whenever the estimated achievable rate lies above the threshold T_R, it means that transmission with a post-FEC BER below 10^{-4} is possible. For example, consider the horizontal line in Fig. 7.58a corresponding to $T_{0.9}$. We can see that with constellation X_2, we are just barely above the line for $P_{\mathrm{in}} \in \{-2\,\mathrm{dBm}, -1\,\mathrm{dBm}\}$, which means that decoding is also

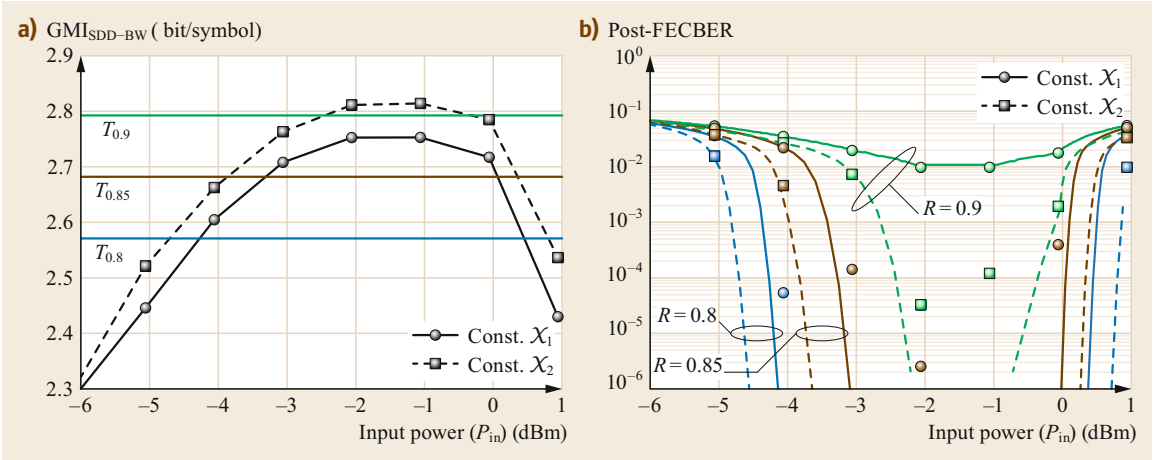


Fig. 7.58a,b Transmission results. Estimated GMI$_{SDD-BW}$ for two constellations $\mathcal{X}_1$ and $\mathcal{X}_2$ together with thresholds to achieve a post-FEC BER of 10^{-4} with LDPC codes ($d_v = 3$, $d_c \in \{15, 20, 30\}$, rates $R \in \{0.8, 0.85, 0.9\}$) with layered scaled-min-sum decoding ($\alpha = 0.75$) **(a)** and interpolated post-FEC BER (from AWGN simulations with the same setup as in Fig. 7.54) together with actual decoding given by markers **(b)**

only barely possible. Contrarily, with constellation $\mathcal{X}_1$, we do not cross the line, and reliable decoding is not possible.

In Fig. 7.58b, we use the simulation results of Fig. 7.56a to estimate the post-FEC performance of the transmission system by interpolation. The interpolated curves are given by the solid (constellation $\mathcal{X}_1$) and dash-dotted (constellation $\mathcal{X}_2$) lines. Additionally, we carried out actual decoding using the LDPC codes introduced before and the method of Sect. 7.7.2. The post-FEC BER results after decoding are given by the solid markers in the figure. We can see that the estimates from interpolation match the actual decoding performance, especially for high BERs. However, we see a deviation at low BERs around 10^{-4}, which is caused by non-stationarity in the measurements. For instance, at P_{in} of -1 dB, a burst caused by the DSP not following the polarization rotation quickly enough occured. The average GMI$_{SDD-BW}$ was still below the threshold. However, the burst causes a frame that cannot be decoded by the LDPC decoder and hence yields a post-FEC BER above 10^{-4}.

7.8 Conclusion and Outlook

A vast body of research on FEC exists today, spanning over seven decades. For most of the classical communication channels, coding schemes that asymptotically achieve the theoretical capacity limits are already available. There also exist very powerful coding schemes that yield excellent performance for short block lengths. Surprisingly, however, despite being a relatively old field, and notwithstanding the enormous successes achieved over the past decades, coding is still a very vibrant research topic. One of the main reasons is that it transcends the classical transmission problem. Today, coding plays a crucial role in myriad applications, from distributed storage, caching and computing, to privacy and post-quantum cryptography. We are convinced that coding will continue to be an indispensable tool for future applications, many of which we cannot even envisage, and expertise and frontline research in this area will certainly be necessary.

For the classical data transmission problem, while most of the major theoretical breakthroughs and code constructions are probably behind us, there are still important open challenges to be addressed. This is particularly true in the area of optical communications, which poses severe constraints in terms of complexity, throughput, and power consumption. For example, finding the best scheme yielding the highest coding gains under heavy complexity constraints, and having a highly parallelizable, efficient decoder hardware architecture enabling high data throughput, remains a very challenging engineering task.

Future research on FEC for optical communications will likely focus on low-complexity coding schemes and

Table 7.5 Comparison of state-of-the-art FEC codes

Name	Ref.	Dec.	Rate R (OH)	NCG (dB)	$\mathrm{BER_{pre}}$	$\mathrm{GMI_{SDD\text{-}BW}}$
LDPC super FEC	[7.46, App. I.6]	HDD	0.937 (6.7%)	8.02[d]	0.00112	—
RS+Hamm. prod.	[7.46, App. I.5]	HDD	0.937 (6.7%)	8.5	0.0019	—
Concat. BCH	[7.46, App. I.3]	HDD	0.937 (6.7%)	8.99	0.00310	—
SP-BCH	[7.274]	HDD	0.937 (6.7%)	9.4[a]	0.0045	—
Staircase	[7.49]	HDD	0.937 (6.7%)	9.41	0.00465	—
Swizzle	[7.275]	HDD	0.937 (6.7%)	9.45	0.00484	—
Braided BCH	[7.50]	HDD	0.937 (6.7%)	9.48[e]	0.00498	—
Concat. BCH	[7.276]	HDD	0.936 (6.8%)	8.91	0.0029	—
Orth. concat. BCH	[7.46, App. I.7]	HDD	0.912 (9.6%)	9.19[f]	0.00444	—
CI-BCH	[7.275, 277]	HDD	0.893 (12%)	10.0	0.00876	—
CI-BCH	[7.275, 277]	HDD	5/6 (20%)	10.5	0.01524	—
Orth. concat. BCH	[7.46, App. I.7]	HDD	0.805 (24.3%)	10.06[f]	0.01302	—
Algebraic LDPC	[7.278]	SDD	0.952 (5.05%)	9.35[c]	0.0041	0.9834
RS+Hamm. prod.	[7.46, App. I.5]	SDD[g]	0.937 (6.7%)	9.4	0.0045	0.9817
Viasat TPC	[7.279]	SDD	0.935 (7%)	10.2	0.0087	0.9654
400G ZR[h]	[7.280]	SDD	0.871 (14.8%)	10.4	0.0125	0.9510
Viasat TPC	[7.281]	SDD	0.87 (15%)	11.0	0.0183	0.9290
Low-power LDPC	[7.282]	SDD	5/6 (20%)	9.5	0.0075	0.9701
Impr. low-power LDPC	[7.283]	SDD	5/6 (20%)	11.0[a]	0.0204	0.9216
LDPC	[7.86]	SDD	5/6 (20%)	11.24[b]	0.0233	0.9110
LDPC	[7.64]	SDD	5/6 (20%)	11.3	0.0241	0.9080
Viasat TPC	[7.279]	SDD	5/6 (20%)	11.3	0.0240	0.9080
Convolutional LDPC	[7.84]	SDD	5/6 (20%)	11.51	0.0267	0.8984
Concatenated LDPC+TPC	[7.284]	SDD	0.830 (20.5%)	10.8[i]	0.0183	0.9288
SC-LDPC	[7.83]	SDD	4/5 (25%)	12.02[c]	0.0373	0.8605
SC-LDPC	[7.165]	SDD	4/5 (25%)	12.14[a]	0.0395	0.8525

[a] Extrapolated from software verification down to 10^{-12}
[b] Extrapolated from FPGA verification down to 10^{-13}
[c] Extrapolated from FPGA verification down to 10^{-14}
[d] The LDPC super FEC allows lower latency due to lower block length than other codes for HDD defined in [7.46] at a cost of a lower NCG
[e] The extra NCG compared with staircase codes is mostly due to the use of an improved decoding algorithm, which however doubles the memory usage. Without this algorithm, the NCG is similar to that of staircase codes
[f] Extrapolated from software verification. System overhead larger due to extra required stuffing bits
[g] Performance evaluated only for a simplified SDD with 4-level channel output quantization
[h] Concatenated Hamming code and staircase code, optimized for SDD with low power consumption
[i] Outer code not implemented and assuming ideal interleaving between inner and outer code

decoding algorithms that allow us to approach capacity limits, and we believe that the most interesting contributions will lie in this area. Two promising research lines in this direction are soft-aided decoding algorithms for product-like codes [7.223, 224] (see also Sect. 7.5.8) and hybrid HDD/SDD schemes, where an outer hard-decision FEC (a staircase code) is concatenated with an inner soft-decision FEC (typically a low-density parity-check code) [7.221, 222]. The latter schemes are now emerging in standards for low-complexity short-reach optical communications [7.280].

To conclude the chapter, we present to the reader an overview of some state-of-the-art FEC schemes that have been proposed for use in optical communications and whose error rate performance has been evaluated down to a BER of 10^{-15} or extrapolated from a reasonably close value. In Table 7.5, we provide a selection of popular and recently proposed FEC schemes for both HDD and SDD. Besides the achievable NCG, we also provide the pre-FEC BER threshold, $\mathrm{BER_{pre}}$, and the GMI threshold, $\mathrm{GMI_{SDD\text{-}BW}}$, if SDD is employed. In the case of SDD, $\mathrm{BER_{pre}}$ is computed assuming BPSK modulation over the AWGN channel. However, we would like to point out (see also Sect. 7.7) that it is not necessarily a good metric for performance estimation, which is why it is given in gray font. The achievable NCGs of the schemes from Table 7.5 are also visualized in Fig. 7.59.

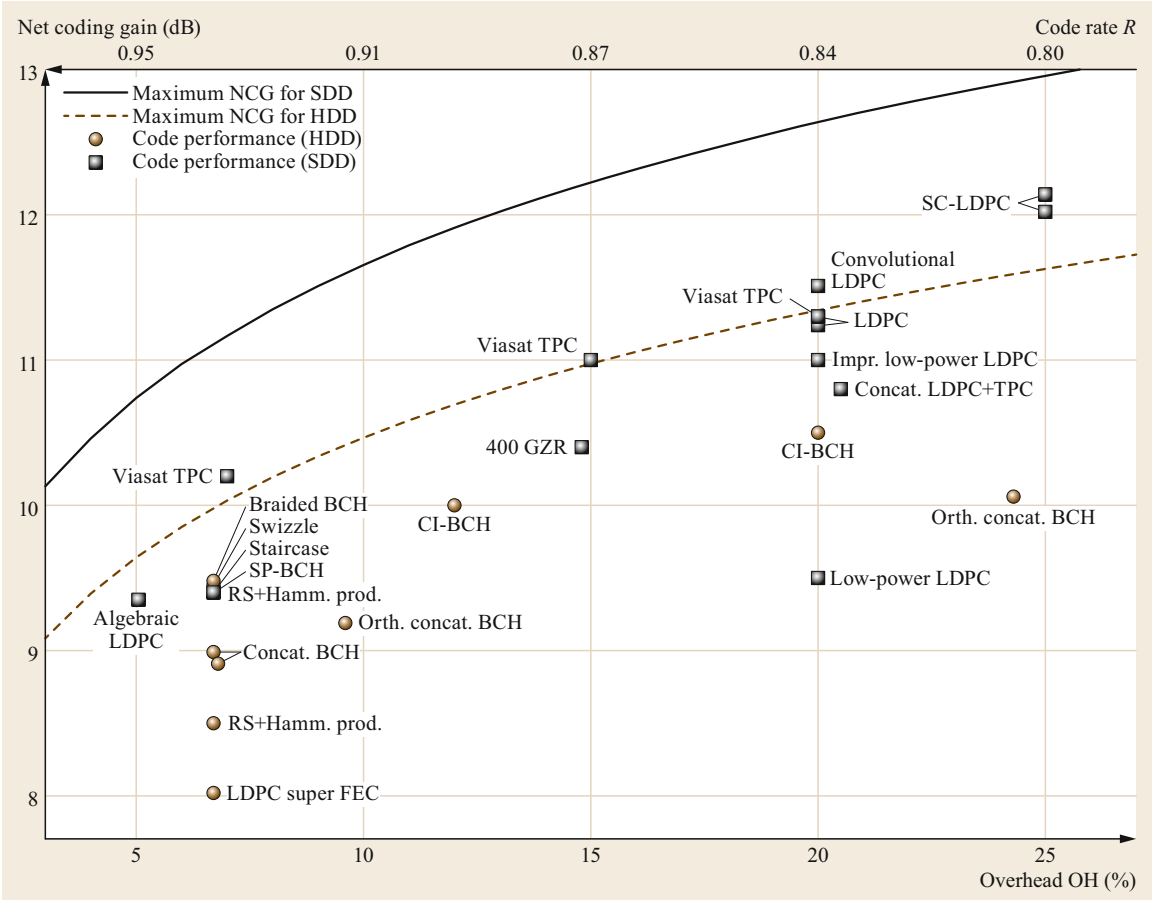


Fig. 7.59 Illustration of NCGs (at an output BER of 10^{-15}) obtained by different state-of-the-art FEC schemes with HDD and SDD taken from Tab. 7.5

Acknowledgments. Laurent Schmalen would like to acknowledge the support of Sebastian Cammerer (University of Stuttgart) for providing the polar coding simulation results shown in Fig. 7.24, Rafael Rios–Müller (Nokia Bell Labs) for carrying out the experiments used in Sect. 7.7.3, and Detlef Suikat and Detlef Rösener for programming the FPGA platform used in Sect. 7.4.5.

References

7.1 C.E. Shannon: A mathematical theory of communication, Bell Syst. Tech. J. **27**(379–423), 623–656 (1948)

7.2 A.G. Dimakis, P.B. Godfrey, Y. Wu, M.J. Wainwright, K. Ramchandran: Network coding for distributed storage systems, IEEE Trans. Inf. Theory **56**(9), 4539–4551 (2010)

7.3 M.A. Maddah-Ali, U. Niesen: Fundamental limits of caching, IEEE Trans. Inf. Theory **60**(5), 2856–2867 (2014)

7.4 G. Liva: Graph-based analysis and optimization of contention resolution diversity slotted ALOHA, IEEE Trans. Commun. **59**(2), 477–487 (2011)

7.5 S. Li, M.A. Maddah-Ali, A.S. Avestimehr: A unified coding framework for distributed computing with straggling servers. In: *Proc. IEEE Globecom Work. (GC Wkshps), Washington, DC, Dec* (2016)

7.6 S. Li, M.A. Maddah-Ali, A.S. Avestimehr: Coding for distributed fog computing, IEEE Commun. Mag. **55**(4), 34–40 (2017)

7.7 D. Pearson: High-speed QKD reconciliation using forward error correction. In: *Proc. AIP Int. Conf. Quantum Commun., Measurement Comp*, Vol. 734 (2004) pp. 299–302, Glasgow

7.8 R.J. McEliece: A public-key cryptosystem based on algebraic coding theory, Deep Space Netw. Prog. Rep. **44**, 114–116 (1978)

7.9 W. Grover: Error correction in dispersion-limited lightwave systems, J. Lightwave Technol. **6**, 643–654 (1988)

7.10 R. Hamming: Error detecting and error correcting codes, Bell Syst. Tech. J. **26**, 147–160 (1950)

7.11 D.J. Costello Jr., J. Hagenauer, H. Imai, S.B. Wicker: Applications of error-control coding, IEEE Trans. Inf. Theory **44**(6), 2531–2560 (1998)

7.12 D.J. Costello Jr., G.D. Forney Jr.: Channel coding: the road to capacity, Proceedings IEEE **95**(6), 1150–1177 (2007)

7.13 M.J.E. Golay: Notes on digital coding, Proc. Inst. Radio Eng. **37**(6), 657–657 (1949)

7.14 A. Barg: At the dawn of the theory of codes, Math. Intell. **15**(1), 20–26 (1993)

7.15 F. MacWilliams, N. Sloane: *The Theory of Error-Correcting Codes*, 2nd edn. (North-Holland, Amsterdam 1978)

7.16 D.E. Muller: Application of Boolean algebra to switching circuit design and to error detection, IRE Trans. Electron. Comput. **3**, 6–12 (1954)

7.17 I. Reed: A class of multiple-error-correcting codes and the decoding scheme, IRE Trans. Inf. Theory **4**(4), 38–49 (1954)

7.18 P. Elias: Error-free coding, IRE Trans. Inf. Theory **PGIT-4**, 29–37 (1954)

7.19 R.D. Cideciyan, S. Furrer, M.A. Lantz: Product codes for data storage on magnetic tape, IEEE Trans. Magn. **53**(2), 1–10 (2017)

7.20 S. Emmadi, K.R. Narayanan, H.D. Pfister: Half-product codes for flash memory. In: *Proc. Non-Volatile Memories Workshop* (2015), San Diego

7.21 M. Wang: WiMAX physical layer: Specifications overview and performance evaluation. In: *Proc. Consumer Commun. Networking Conf. (CCNC)* (2011) pp. 10–12

7.22 J. Justesen, K.J. Larsen, L.A. Pedersen: Error correcting coding for OTN, IEEE Commun. Mag. **48**(9), 70–75 (2010)

7.23 P. Elias: Coding for noisy channels. In: *IRE Convention Record, Part IV* (1955) pp. 37–46

7.24 ETSI: *Recommendation GSM 05.03* (ETSI, Sophia Antipolis 1994)

7.25 IEEE: Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 5: Enhancements for higher throughput. In: *IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements* (2005), Std 802.11-2005, Tech. Rep.

7.26 IEEE: Part 3: Carrier sense multiple access with collision detection (CSMA/CD): Access method and physical layer specifications. In: *802.3, Local and Metropolitan Area Networks-specific Requirements* (2006)

7.27 A. Viterbi: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, IEEE Trans. Inf. Theory **13**(2), 260–269 (1967)

7.28 G.D. Forney Jr.: The Viterbi algorithm: A personal history. In: *Proc. Viterbi Conf.* (2005), University of Southern California, Los Angeles

7.29 T.T. Ha, R.L. Borchardt, E.-S. Hwang: Direct detection optical systems employing OOK modulation and error correction codes, J. Opt. Commun. **17**(6), 212–215 (1996)

7.30 T. Wuth, E. Agrell, M. Karlsson, M. Sköld: Fiber communications using convolutional coding and bandwidth-efficient modulation, Opt. Express **14**(2), 542–555 (2006)

7.31 A. Bisplinghoff, N. Beck, M. Ene, M. Danninger, T. Kupfer: Phase slip tolerant, low power multi-level coding for 64QAM with 12.9 dB NCG. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2016) pp. 1–3, Anaheim

7.32 A. Hocquenghem: Codes correcteurs d'erreurs, Chiffres **2**, 147–156 (1959)

7.33 R.C. Bose, D.K. Ray-Chaudhuri: On a class of error correcting binary group codes, Inform. Control **3**(1), 68–79 (1960)

7.34 R. Blahut: *Algebraic Codes for Data Transmission* (Cambridge Univ. Press, New York 2003)

7.35 S. Lin, D.J. Costello Jr.: *Error Control Coding*, 2nd edn. (Prentice Hall, Upper Saddle River 2004)

7.36 S. Quiroga, D. Torres, A. Veloz: In band FEC decoder for SONET/SDH at 2.5 Gbit/s and 10 Gbit/s. In: *Proc. Int. Conf. Elec. Electron. Eng. (ICEEE)* (2004) pp. 70–73

7.37 I.S. Reed, G. Solomon: Polynomial codes over certain finite fields, J. Soc. Ind. Appl. Math. **8**(2), 300–304 (1960)

7.38 E. Berlekamp: *Algebraic Coding Theory* (McGraw-Hill, New York 1968)

7.39 J. Massey: Shift-register synthesis and BCH decoding, IEEE Trans. Inf. Theory **15**(1), 122–127 (1969)

7.40 S. Yamamoto, H. Takahira, E. Shibano, M. Tanaka, Y. Chen: BER performance improvement by forward error correcting code in 5 Gbit/s 9000 km EDFA transmission system, Electron. Lett. **30**, 718–719 (1994)

7.41 L. Schmalen, A. de Lind van Wijngaarden, S. ten Brink: Forward error correction in optical core and access networks, Bell Labs Tech. J. **18**(3), 39–66 (2013)

7.42 ITU-T: *Recommendation G.975: Forward error correction for submarine systems* 1996)

7.43 F. Chang, K. Onohara, T. Mizuochi: Forward error correction for 100 G transport networks, IEEE Commun. Mag. **48**, S48–S55 (2010)

7.44 G.D. Forney: *Concatenated codes, Ph.D. dissertation* (Massachusetts Institute of Technology (MIT), Cambridge 1965)

7.45 ETSI: *DVB-S, ETSI standard EN 300 421 v.1.1.2* (ETSI, Sophia Antipolis 1997)

7.46 ITU-T: *Recommendation G.975.1: Forward error correction for high bit-rate DWDM submarine systems* (ITU, Geneva 2004)

7.47 A. Afshar: Deployment of 100G in the metro network, applications and drivers, Cortina White Paper, Tech. Rep. 20.

7.48 M. Scholten, T. Coe, J. Dillard: Continuously-interleaved BCH (CI-BCH) FEC delivers best in class NECG for 40G and 100G metro applications. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2010), San Diego

7.49 B.P. Smith, A. Farhood, A. Hunt, F.R. Kschischang, J. Lodge: Staircase codes: FEC for 100 Gb/s OTN, J. Lightwave Technol. **30**(1), 110–117 (2012)

7.50 Y.-Y. Jian, H.D. Pfister, K.R. Narayanan, R. Rao, R. Mazahreh: Iterative hard-decision decoding of braided BCH codes for high-speed optical communication. In: *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)* (2013), Atlanta

7.51 D.C. MacKay, R. Neal: Near Shannon limit performance of low density parity check codes, Electron. Lett. **32**, 1645–1646 (1996)

7.52 M. Luby, M. Mitzenmacher, A. Shokrollahi, D.A. Spielman, V. Stemann: Practical loss-resilient codes. In: *Proc. ACM Symp. Theory Comput.* (1997) pp. 150–159

7.53 R.G. Gallager: Low-density parity-check codes, IRE Trans. Inf. Theory **8**(1), 21–28 (1962)

7.54 R.G. Gallager: *Low-Density Parity-Check Codes* (M.I.T. Press, Cambridge 1963)

7.55 IEEE: *802.11, wireless LAN medium access control (MAC) and physical layer (PHY) specifications* (IEEE, Piscataway Township 2012)

7.56 ETSI: *DVB-S2, ETSI standard EN 302 307 v.1.2.1* (ETSI, Sophia Antipolis 2009)

7.57 IEEE: *802.3an, local and metropolitan area networks-specific requirements Part 3: Carrier sense multiple access with collision detection (CSMA/CD): Access method and physical layer specifications* (IEEE, Piscataway Township 2006)

7.58 R. Pyndiah: Near optimum decoding of product codes: Block turbo codes, IEEE Trans. Commun. **46**(8), 1003–1010 (1998)

7.59 T. Mizuochi, Y. Miyata, T. Kobayashi, K. Ouchi, K. Kuno, K. Kubo, K. Shimizu, H. Tagami, H. Yoshida, H. Fujita, M. Akita, K. Motoshima: Forward error correction based on block turbo code with 3-bit soft decision for 10-Gb/s optical communication systems, IEEE J. Sel. Top. Quantum Electron. **10**(2), 376–386 (2004)

7.60 B. Vasic, I.B. Djordjevic: Low-density parity check codes for long-haul optical communication systems, IEEE Photon. Technol. Lett. **14**(8), 1208–1210 (2002)

7.61 Y. Miyata, K. Kubo, H. Yoshida, T. Mizuochi: Proposal for frame structure of optical channel transport unit employing LDPC codes for 100 Gb/s FEC. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2009)

7.62 I.B. Djordjevic, B. Vasic: Nonbinary LDPC codes for optical communication systems, IEEE Photon. Technol. Lett. **17**(10), 2224–2226 (2005)

7.63 T. Richardson: Error floors of LDPC codes. In: *Proc. Allerton Annu. Conf. Commun. Control Comp.* (2003)

7.64 D.A. Morero, M.A. Castrillon, F.A. Ramos, T.A. Goette, O.E. Agazzi, M.R. Hueda: Non-concatenated FEC codes for ultra-high speed optical transport networks. In: *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)* (2011) pp. 1–5

7.65 E. Arıkan: Channel polarization: A method for constructing capacity-achieving codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2008) pp. 1173–1177

7.66 E. Arıkan: Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, IEEE Trans. Inf. Theory **55**(7), 3051–3073 (2009)

7.67 I. Tal, A. Vardy: List decoding of polar codes, IEEE Trans. Inf. Theory **61**(5), 2213–2226 (2015)

7.68 ETSI: *3GPP TR 38.802 V14.2.0, Technical Specification Group Radio Access Network: Study on new radio access technology physical layer aspects (release 14), 3GPP* (ETSI, Sophia Antipolis 2017), Tech. Rep.

7.69 G. Sarkis, P. Giard, A. Vardy, C. Thibeault, W.J. Gross: Fast list decoders for polar codes, IEEE J. Sel. Areas Commun. **34**(2), 318–328 (2016)

7.70 P. Giard, C. Thibeault, W.J. Gross: *High-Speed Decoders for Polar Codes* (Springer, Berlin, Heidelberg 2016)

7.71 A.J. Felström, K.S. Zigangirov: Time-varying periodic convolutional codes with low-density parity-check matrix, IEEE Trans. Inf. Theory **45**(6), 2181–2191 (1999)

7.72 M. Lentmaier, G. Fettweis, K. Zigangirov, D. Costello: Approaching capacity with asymptotically regular LDPC codes. In: *Proc. Inform. Theory Appl. Workshop (ITA)* (2009)

7.73 M. Lentmaier, A. Sridharan, D.J. Costello, K. Zigangirov: Iterative decoding threshold analysis for LDPC convolutional codes, IEEE Trans. Inf. Theory **56**(10), 5274–5289 (2010)

7.74 M. Lentmaier, D. Mitchell, G. Fettweis, D. Costello: Asymptotically good LDPC convolutional codes with AWGN channel thresholds close to the shannon limit. In: *Proc. Int. Symp. Turbo Codes Iterative Inform. Process. (ISTC)* (2010), Brest

7.75 S. Kudekar, T. Richardson, R. Urbanke: Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC, IEEE Trans. Inf. Theory **57**(2), 803–834 (2011)

7.76 S. Kudekar, T. Richardson, R. Urbanke: Spatially coupled ensembles universally achieve capacity under belief propagation, IEEE Trans. Inf. Theory **59**(12), 7761–7813 (2013)

7.77 A. Iyengar, M. Papaleo, P. Siegel, J. Wolf, A. Vanelli-Coralli, G. Corazza: Windowed decoding of protograph-based LDPC convolutional codes over erasure channels, IEEE Trans. Inf. Theory **58**(4), 2303–2320 (2012)

7.78 S. Moloudi, M. Lentmaier, A. Graell i Amat: Spatially coupled turbo-like codes, IEEE Trans. Inf. Theory **63**(10), 6199–6215 (2017)

7.79 V. Aref, N. Macris, R. Urbanke, M. Vuffray: Lossy source coding via spatially coupled LDGM ensembles. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2012) pp. 373–377, Cambridge

7.80 K. Takeuchi, T. Tanaka, T. Kawabata: Improvement of BP-based CDMA multiuser detection by spatial coupling. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2011) pp. 1489–1493, St. Petersburg

7.81 D.L. Donoho, A. Javanmard, A. Montanari: Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing, IEEE Trans. Inf. Theory **59**(11), 7434–7464 (2013)

7.82 Z. Wu, J.K. Fischer, B. Lankl: Experimental investigation of polar code performance for coherent UDWDM PONs. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2015)

7.83 L. Schmalen, D. Suikat, D. Rösener, V. Aref, A. Leven, S. ten Brink: Spatially coupled codes and optical fiber communications: An ideal match? In: *Proc. IEEE Workshop Signal Process. Adv. Wirel. Commun. (SPAWC)* (2015) pp. 460–464

7.84 D. Chang, F. Yu, Z. Xiao, N. Stojanovic, F.N. Hauske, Y. Cai, C. Xie, L. Li, X. Xu, Q. Xiong: LDPC convolutional codes using layered decoding algorithm for high speed coherent optical transmission. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2012) p. OW1H.4

7.85 I.B. Djordjevic, W. Ryan, B. Vasic: *Coding for Optical Channels* (Springer, Cham 2010)

7.86 A. Leven, L. Schmalen: Status and recent advances on forward error correction technologies for lightwave systems, J. Lightw. Technol. **32**(16), 2735–2750 (2014)

7.87 N. Merhav, G. Kaplan, A. Lapidoth, S. Shamai Shitz: On information rates for mismatched decoders, IEEE Trans. Inf. Theory **40**(6), 1953–1967 (1994)

7.88 A. Ganti, A. Lapidoth, I.E. Telatar: Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit, IEEE Trans. Inf. Theory **46**(7), 2315–2328 (2000)

7.89 T. Fehenberger, A. Alvarado, P. Bayvel, N. Hanik: On achievable rates for long-haul fiber-optic communications, Opt. Express **23**(7), 9183–9191 (2015)

7.90 A. Sheikh, A. Graell i Amat, G. Liva: Achievable information rates for coded modulation with hard decision decoding for coherent fiber-optic systems, J. Lightwave Technol. **35**(23), 5069–5078 (2017)

7.91 G. Böcherer: Achievable rates for probabilistic shaping, http://arxiv.org/abs/arXiv:1707.01134 (2017)

7.92 I.B. Djordjevic, B. Vasic, M. Ivkovic, I. Gabitov: Achievable information rates for high-speed long-haul optical transmission, J. Lightwave Technol. **23**(11), 3755–3763 (2005)

7.93 E. Agrell, A. Alvarado, G. Durisi, M. Karlsson: Capacity of a nonlinear optical channel with finite memory, J. Lightwave Technol. **32**(16), 2862–2876 (2014)

7.94 R.J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel: Capacity limits of optical fiber networks, J. Lightwave Technol. **28**(4), 662–701 (2010)

7.95 M. Secondini, E. Forestieri, G. Prati: Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps, J. Lightwave Technol. **31**(23), 3839–3852 (2013)

7.96 G. Caire, G. Taricco, E. Biglieri: Bit-interleaved coded modulation, IEEE Trans. Inf. Theory **44**(3), 927–946 (1998)

7.97 A. Martinez, A. Guillén i Fàbregas, G. Caire, F.M.J. Willems: Bit-interleaved coded modulation revisited: A mismatched decoding perspective, IEEE Trans. Inf. Theory **55**(6), 2756–2765 (2009)

7.98 G. Kaplan, S. Shamai Shitz: Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment, AEÜ **47**(4), 228–230 (1993)

7.99 P. Poggiolini: The GN model of non-linear propagation in uncompensated coherent optical systems, J. Lightwave Technol. **30**(24), 3857–3879 (2012)

7.100 P. Johannisson, E. Agrell: Modeling of nonlinear signal distortion in fiber-optic networks, J. Lightwave Technol. **32**(23), 3942–3950 (2014)

7.101 W.E. Ryan, S. Lin: *Channel Codes. Classical and Modern* (Cambridge Univ. Press, Cambridge 2009)

7.102 R.L. Dobrushin: Asymptotic optimality of group and systematic codes for some channels, Theory Probab. Appl. **8**(1), 47–60 (1963)

7.103 A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri: EGN model of non-linear fiber propagation, Opt. Express **22**(13), 16335–16362 (2014)

7.104 N.A. Shevchenko, J.E. Prilepsky, S.A. Derevyanko, A. Alvarado, P. Bayvel, S.K. Turitsyn: A lower bound on the per soliton capacity of the nonlinear optical fibre channel. In: *Proc. IEEE Inform. Theory Workshop (ITW)* (2015) pp. 104–108

7.105 A. Alvarado, L. Szczecinski, R. Feick: On the distribution of extrinsic L-values in gray-mapped 16-QAM. In: *Proc. Int. Conf. Wirel. Commun. Mobile Comp. (IWCMC)* (2007) pp. 329–336

7.106 Top500: The List, https://www.top500.org/ (2019)

7.107 S. Dolinar, D. Divsalar, F. Pollara: *Code performance as a function of block size* (Jet Propulsion Laboratory, Pasadena 1998) pp. 42–133, TMO Prog. Rep.

7.108 Y. Polyanskiy, H. Poor, S. Verdu: Channel coding rate in the finite blocklength regime, IEEE Trans. Inf. Theory **56**(5), 2307–2359 (2010)

7.109 C.E. Shannon: Probability of error for optimal codes in a gaussian channel, Bell Syst. Tech. J. **38**, 611–656 (1959)

7.110 T. Erseghe: Coding in the finite-blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations, IEEE Trans. Inf. Theory **62**(12), 6854–6883 (2016)

7.111 T. Richardson, R. Urbanke: *Modern Coding Theory* (Cambridge Univ. Press, Cambridge 2008)

7.112 T.J. Richardson, M.A. Shokrollahi, R.L. Urbanke: Design of capacity-approaching irregular low-density parity-check codes, IEEE Trans. Inf. Theory **47**(2), 619–637 (2001)

7.113 D. Divsalar, C. Jones, S. Dolinar, J. Thorpe: Protograph based LDPC codes with minimum distance linearly growing with block size. In: *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Vol. 3 (2005) p. 5

7.114 Consultative Committee for Space Data Systems (CCSDS): *TM synchronization and channel coding, recommended standard, CCSDS 131.0-B-2, blue book* (CCSDS, Washington 2011), Tech. Rep.

7.115 R.M. Tanner: A recursive approach to low complexity codes, IEEE Trans. Inf. Theory **27**(5), 533–547 (1981)

7.116 L. Dolecek, P. Lee, Z. Zhang, V. Anantharam, B. Nikolic, M. Wainwright: Predicting error floors of structured LDPC codes: Deterministic bounds and estimates, IEEE J. Sel. Areas Commun. **27**(6), 908–917 (2009)

7.117 T. Richardson, R. Urbanke: The capacity of low-density parity-check codes under belief propagation decoding, IEEE Trans. Inf. Theory **47**(2), 599–618 (2001)

7.118 S. Jayasooriya, M. Shirvanimoghaddam, L. Ong, G. Lechner, S.J. Johnson: A new density evolution approximation for LDPC and multi-edge type LDPC codes, IEEE Trans. Commun. **64**(10), 4044–4056 (2016)

7.119 S.-Y. Chung, G.D. Forney Jr., T. Richardson, R. Urbanke: On the design of low-density parity-check codes within 0.0045 db of the Shannon limit, IEEE Commun. Lett. **5**(2), 58–60 (2001)

7.120 S. ten Brink, G. Kramer, A. Ashikhmin: Design of low-density parity-check codes for modulation and detection, IEEE Trans. Commun. **52**(4), 670–678 (2004)

7.121 L. Schmalen, S. ten Brink, A. Leven: Advances in detection and error correction for coherent optical communications: Regular, irregular, and spatially coupled LDPC code designs. In: *Enabling Technologies for High Spectral-Efficiency Coherent Optical Communication Networks*, ed. by X. Zhou, C. Xie (Wiley, Hoboken 2016), Chap. 3

7.122 B. Vasic, I.B. Djordjevic, R.K. Kostuk: Low-density parity check codes and iterative decoding for long-haul optical communication systems, J. Lightwave Technol. **21**(2), 438 (2003)

7.123 G. Liva, S. Song, L. Lan, Y. Zhang, S. Lin, W. Ryan: Design of LDPC codes: A survey and new results, J. Commun. Software Syst. **2**(3), 191–211 (2006)

7.124 N. Bonello, S. Chen, L. Hanzo: Design of low-density parity-check codes, IEEE Veh. Technol. Mag. **6**(4), 1574–1606 (2011)

7.125 X.-Y. Hu, E. Eleftheriou, D.M. Arnold: Regular and irregular progressive edge-growth Tanner graphs, IEEE Trans. Inf. Theory **51**(1), 386–398 (2005)

7.126 T. Tian, C. Jones, J.D. Villasenor, R.D. Wesel: Construction of irregular LDPC codes with low error floors. In: *Proc. IEEE Int. Conf. Commun. (ICC)* (2003) pp. 3125–3129

7.127 H. Xiao, A.H. Banihashemi: Improved progressive-edge-growth (PEG) construction of irregular LDPC codes, IEEE Commun. Lett. **8**(12), 715–717 (2004)

7.128 T. Tian, C.R. Jones, J.D. Villasenor, R.D. Wesel: Selective avoidance of cycles in irregular LDPC code construction, IEEE Trans. Commun. **52**(8), 1242–1247 (2004)

7.129 T. Richardson, R. Urbanke: Multi-edge type LDPC codes. In: *Workshop honoring Prof. Bob McEliece on his 60th birthday* (California Institute of Technology, Pasadena 2002) pp. 24–25

7.130 J. Thorpe: *Low-density parity-check (LDPC) codes constructed from protographs* (NASA JPL, Pasadena 2003), IPN Progress Report 42-154

7.131 J. Li, S. Lin, K. Abdel-Ghaffar, W.E. Ryan, D.J. Costello Jr.: *LDPC Code Designs, Constructions, and Unification* (Cambridge Univ. Press, Cambridge 2016)

7.132 M.P.C. Fossorier: Quasi-cyclic low-density parity-check codes from circulant permutation matrices, IEEE Trans. Inf. Theory **50**(8), 1788–1793 (2004)

7.133 R. Smarandache, P.O. Vontobel: Quasi-cyclic LDPC codes: Influence of proto- and Tanner-graph structure on minimum Hamming distance upper bounds, IEEE Trans. Inf. Theory **58**(2), 585–607 (2012)

7.134 X. Jiang, M.H. Lee: Large girth quasi-cyclic LDPC codes based on the Chinese remainder theorem, IEEE Commun. Lett. **13**(5), 342–344 (2009)

7.135 T. Richardson, R. Urbanke: Efficient encoding of low-density parity-check codes, IEEE Trans. Inf. Theory **47**(2), 638–656 (2001)

7.136 D.J.C. MacKay: *Information Theory, Inference and Learning Algorithms* (Cambridge Univ. Press, Cambridge 2003)

7.137 J. Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, New York 1988)

7.138 G. Lechner, T. Pedersen, G. Kramer: Analysis and design of binary message passing decoders, IEEE Trans. Commun. **60**(3), 601–607 (2012)

7.139 J. Hagenauer, E. Offer, L. Papke: Iterative decoding of binary block and convolutional codes, IEEE Trans. Inf. Theory **42**(2), 429–445 (1996)

7.140 D.R. Pauluzzi, N.C. Beaulieu: A comparison of SNR estimation techniques for the AWGN channel, IEEE Trans. Commun. **48**(10), 1681–1691 (2000)

7.141 C.F. Mecklenbräuker, S. Paul: On estimating the signal to noise ratio from BPSK signals. In: *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Vol. 4 (2005)

7.142 T.K. Moon: *Error Correction Coding – Mathematical Methods and Algorithms* (Wiley, Hoboken 2005)

7.143 J. Chen, M.P.C. Fossorier: Near optimum universal belief propagation based decoding of low-density parity check codes, IEEE Trans. Commun. **50**(3), 406–414 (2002)

7.144 M. Ardakani, F.R. Kschischang: Gear-shift decoding, IEEE Trans. Commun. **54**(7), 1235–1242 (2006)

7.145 J. Zhao, F. Zarkeshvari, A.H. Banihashemi: On implementation of min-sum algorithm and its modifications for decoding low-density parity-check (LDPC) codes, IEEE Trans. Commun. **53**(4), 549–554 (2005)

7.146 D. Hocevar: A reduced complexity decoder architecture via layered decoding of LDPC codes. In: *Proc. IEEE Workshop Signal Process. Syst. (SiPS)* (2004)

7.147 A. Leven, L. Schmalen: Implementation aspects of coherent transmit and receive functions in application-specific integrated circuits. In: *Optical Fiber Telecommunications VI-A*, ed. by I.P. Kaminow, T. Li, A.E. Willner (Academic, New York 2013), Chap. 15

7.148 A. Wyner, R. Ash: Analysis of recurrent codes, IEEE Trans. Inf. Theory **9**(3), 143–156 (1963)

7.149 A. Yedla, Y.-Y. Jian, P.S. Nguyen, H.D. Pfister: A simple proof of Maxwell saturation for coupled scalar

recursions, IEEE Trans. Inf. Theory **60**(11), 6943–6965 (2014)
7.150 A. Yedla, Y.-Y. Jian, P. Nguyen, H. Pfister: A simple proof of threshold saturation for coupled vector recursions. In: *Proc. IEEE Inform. Theory Workshop (ITW)* (2012) pp. 25–29, Lausanne
7.151 A. Piemontese, A. Graell i Amat, G. Colavolpe: Nonbinary spatially-coupled LDPC codes on the binary erasure channel. In: *Proc. IEEE Int. Conf. Commun. (ICC)* (2013) pp. 3270–3274, Budapest
7.152 I. Andriyanova, A. Graell i Amat: Threshold saturation for nonbinary SC-LDPC codes on the binary erasure channel, IEEE Trans. Inf. Theory **62**(5), 2622–2638 (2016)
7.153 ITU-T: *Recommendation G.709: Interfaces for the optical transport network* (ITU-T, Genf 2016)
7.154 W. Lautenschlaeger, N. Benzaoui, F. Buchali, L. Dembeck, R. Dischler, B. Franz, U. Gebhard, J. Milbrandt, Y. Pointurier, D. Roesener, L. Schmalen, A. Leven: Optical ethernet – flexible optical metro networks, J. Lightwave Technol. **35**(12), 2346–2357 (2017)
7.155 S. Kudekar, C. Méasson, T. Richardsony, R. Urbanke: Threshold saturation on BMS channels via spatial coupling. In: *Proc. Int. Symp. Turbo Codes Iterative Inform. Process. (ISTC)* (2010) pp. 309–313
7.156 C. Häger, A. Graell i Amat, A. Alvarado, F. Brännström, E. Agrell: Optimized bit mappings for spatially coupled LDPC codes over parallel binary erasure channels. In: *Proc. IEEE Int. Conf. Commun. (ICC)* (2014) pp. 2064–2069, Sydney
7.157 S. Cammerer, V. Aref, L. Schmalen, S. ten Brink: Triggering wave-like convergence of tail-biting spatially coupled LDPC codes. In: *Conf. Inform. Sci. Syst. (CISS)* (2016) pp. 93–98
7.158 L. Schmalen, V. Aref, F. Jardel: Non-uniformly coupled LDPC codes: Better thresholds, smaller rate-loss, and less complexity. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2017), Aachen
7.159 S. Kudekar, T. Richardson, R. Urbanke: Wave-like solutions of general one-dimensional spatially coupled systems, IEEE Trans. Inf. Theory **61**(8), 4117–4157 (2015)
7.160 M. Tavares: *On low-density parity-check convolutional codes: Constructions, analysis and VLSI implementations, Ph.D. dissertation* (TU Dresden, Dresden 2010)
7.161 A.E. Pusane, A.J. Feltström, A. Sridharan, M. Lentmaier, K.S. Zigangirov, D.J. Costello Jr.: Implementation aspects of LDPC convolutional codes, IEEE Trans. Commun. **56**(7), 1060–1069 (2008)
7.162 V. Aref, L. Schmalen, S. ten Brink: On the convergence speed of spatially coupled LDPC ensembles. In: *Proc. Allerton Annu. Conf. Commun. Control Comp.* (2013) p. 2013, arXiv: 1307.3780
7.163 R. El-Khatib, N. Macris: The velocity of the decoding wave for spatially coupled codes on BMS channels. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2016) pp. 2119–2123
7.164 R. El-Khatib, N. Macris: The velocity of the propagating wave for spatially coupled systems with applications to LDPC codes, IEEE Trans. Inf. Theory **64**(11), 7113–7131 (2018)
7.165 L. Schmalen, V. Aref, J. Cho, D. Suikat, D. Rösener, A. Leven: Spatially coupled soft-decision error correction for future lightwave systems, J. Lightwave Technol. **33**(5), 1109–1116 (2015)
7.166 P.M. Olmos, R. Urbanke: A scaling law to predict the finite-length performance of spatially-coupled LDPC codes, IEEE Trans. Inf. Theory **61**(6), 3164–3184 (2015)
7.167 M. Stinner, P.M. Olmos: Analyzing finite-length protograph-based spatially coupled LDPC codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2014) pp. 891–895
7.168 J. Cho, L. Schmalen: Construction of protographs for large-girth structured LDPC convolutional codes. In: *Proc. IEEE Int. Conf. Commun. (ICC)* (2015) pp. 4412–4417
7.169 M. Battaglioni, A. Tasdighi, G. Cancellieri, F. Chiaraluce, M. Baldi: Design and analysis of time-invariant SC-LDPC codes with small constraint length, IEEE Trans. Commun. **66**(3), 918–931 (2018)
7.170 D. Achlioptas, H. Hassani, W. Liu, R. Urbanke: Time-invariant LDPC convolutional codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2017), Aachen
7.171 L. Schmalen, D. Suikat, D. Rösener, A. Leven: Evaluation of left-terminated spatially coupled LDPC codes for optical communications. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2014)
7.172 K. Tazoe, K. Kasai, K. Sakaniwa: Efficient termination of spatially-coupled codes. In: *Proc. IEEE Inform. Theory Workshop (ITW)* (2012)
7.173 M.R. Sanatkar, H.D. Pfister: Increasing the rate of spatially-coupled codes via optimized irregular termination. In: *Proc. Int. Symp. Turbo Codes Iterative Inform. Process. (ISTC)* (2016)
7.174 H. Kwak, J. Kim, Jong-Seon: Rate-loss reduction of SC-LDPC codes by optimizing reliable variable nodes via expected graph evolution. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2017) pp. 2930–2934
7.175 T. Jerkovits, G. Liva, A. Graell i Amat: Improving the decoding threshold of tailbiting spatially coupled LDPC codes by energy shaping, IEEE Commun. Lett. **22**(4), 660–663 (2018)
7.176 C. Häger, A. Graell i Amat, F. Brännström, A. Alvarado, E. Agrell: Comparison of terminated and tailbiting spatially coupled LDPC codes with optimized bit mapping for PM-64-QAM. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2014)
7.177 C. Häger, A. Graell i Amat, F. Brännström, A. Alvarado, E. Agrell: Improving soft FEC performance for higher-order modulations via optimized bit channel mappings, Opt. Express **22**(12), 14544–14558 (2014)
7.178 C. Häger, A. Graell i Amat, F. Brännström, A. Alvarado, E. Agrell: Terminated and tailbiting spatially coupled codes with optimized bit mappings for spectrally efficient fiber-optical systems, J. Lightwave Technol. **33**(7), 1275–1285 (2015)
7.179 S. Cammerer, L. Schmalen, V. Aref, S. ten Brink: Wave-like decoding of tail-biting spatially coupled LDPC codes through iterative demapping. In: *Proc.*

Int. Symp. Turbo Codes Iterative Inform. Process. (ISTC) (2016) pp. 121–125
7.180 L. Schmalen, D. Suikat, V. Aref, D. Rösener: On the design of capacity-approaching unit-memory spatially coupled LDPC codes for optical communications. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2016) pp. 1–3
7.181 T. Koike-Akino, S.C. Draper, Y. Wang, K. Sugihara, W. Matsumoto, D.S. Millar, K. Parsons, V. Arlunno, K. Kojima: Optimal layered scheduling for hardware-efficient windowed decoding of LDPC convolutional codes. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2016) pp. 1–3
7.182 K. Klaiber, S. Cammerer, L. Schmalen, S. ten Brink: Avoiding burst-like error patterns in windowed decoding of spatially coupled LDPC codes. In: *Proc. Int. Symp. Turbo Codes Iterative Inform. Process. (ISTC)* (2018), Hong Kong
7.183 D.J. Costello Jr., L. Dolecek, T. Fuja, J. Kliewer, D. Mitchell, R. Smarandache: Spatially coupled sparse codes on graphs: Theory and practice, IEEE Commun. Mag. **52**(7), 168–176 (2014)
7.184 E. Şaşoğlu, I. Telatar, E. Arıkan: Polarization for arbitrary discrete memoryless channels. In: *Proc. IEEE Inform. Theory Workshop (ITW)* (2009) pp. 144–148
7.185 D. Sutter, J.M. Renes, F. Dupuis, R. Renner: Achieving the capacity of any DMC using only polar codes. In: *Proc. IEEE Inform. Theory Workshop (ITW)* (2012) pp. 114–118
7.186 N. Stolte: *Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung, Ph.D. dissertation* (Technische Universität Darmstadt, Darmstadt 2002)
7.187 I. Tal, A. Vardy: How to construct polar codes, IEEE Trans. Inf. Theory **59**(10), 6562–6582 (2013)
7.188 E. Arıkan: Systematic polar coding, IEEE Commun. Lett. **15**(8), 860–862 (2011)
7.189 T. Ahmad: *Polar codes for optical communications, Ph.D. dissertation* (Bilkent University, Ankara 2016)
7.190 B. Li, H. Shen, D. Tse: Parallel decoders of polar codes, https://arxiv.org/abs/1309.1026 (2013)
7.191 G. Sarkis, P. Giard, A. Vardy, C. Thibeault, W.J. Gross: Fast polar decoders: Algorithm and implementation, IEEE J. Sel. Areas Commun. **32**(5), 946–957 (2014)
7.192 S. Cammerer, M. Ebada, A. Elkelesh, S. ten Brink: Sparse graphs for belief propagation decoding of polar codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2018) pp. 1465–1469, Vail
7.193 C. Berrou, A. Glavieux: Near optimum error correcting coding and decoding: Turbo-codes, IEEE Trans. Commun. **44**(10), 1261–1271 (1996)
7.194 J. Zhang, Y. Wang, M.P.C. Fossorier, J.S. Yedidia: Iterative decoding with replicas, IEEE Trans. Inf. Theory **53**(5), 1644–1663 (2007)
7.195 M. Mondelli, S.H. Hassani, R.L. Urbanke: Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors, IEEE Trans. Inf. Theory **62**(12), 6698–6712 (2016)
7.196 Z. Zhang, L. Dolecek, B. Nikolic, V. Anantharam, M. Wainwright: Investigation of error floors of structured low-density parity-check codes by hardware emulation. In: *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)* (2006)
7.197 M. Arabaci, I.B. Djordjevic: An alternative FPGA implementation of decoders for quasi-cyclic LDPC codes. In: *Proc. Telecommun. Forum (TELFOR)* (2008)
7.198 D.-U. Lee, J.D. Villasenor, W. Luk, P.H.W. Leong: A hardware Gaussian noise generator using the Box-Muller method and its error analysis, IEEE Trans. Comput. **55**(6), 659–671 (2006)
7.199 F. Yu, M. Li, N. Stojanovic, C. Xie, Z. Xiao, L. Li: FPGA demonstration of stretched continuously interleaved BCH code with low error floor for short-range optical transmission. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2017), Los Angeles
7.200 A.J. Feltström, D. Truhachev, M. Lentmaier, K.S. Zigangirov: Braided block codes, IEEE Trans. Inf. Theory **55**(6), 2640–2658 (2009)
7.201 S. Benedetto, E. Biglieri: *Principles of Digital Transmission: With Wireless Applications* (Kluwer Academic, Norwell 1999)
7.202 European Telecommunications Standards Institute: *Digital video broadcasting (DVB); upper layer FEC for DVB systems, ETSI TR 102 993, 2007* (ETSI, Sophia Antipolis 2011), Tech. Rep.
7.203 Consultative Committee for Space Data Systems: *CCSDS 131.0-B-2, CCSDS recommended standard for TM synchronization and channel coding* (CCSDS, Washington 2011), Tech. Rep.
7.204 J.S. Plank: A tutorial on Reed-Solomon coding for fault-tolerance in RAID-like systems, Softw. Pract. Exp. **27**(9), 995–1012 (1997)
7.205 L.M. Tolhuizen: More results on the weight enumerator of product codes, IEEE Trans. Inf. Theory **48**(9), 2573–2577 (2002)
7.206 S. Hirasawa, M. Kasahara, Y. Sugiyama, T. Namekawa: Modified product codes, IEEE Trans. Inf. Theory **30**(2), 299–306 (1984)
7.207 M. Alipour, O. Etesami, G. Maatouk, A. Shokrollahi: Irregular product codes. In: *Proc. IEEE Inform. Theory Workshop (ITW)* (2012) pp. 197–201, Lausanne
7.208 L.M. Zhang, F.R. Kschischang: Staircase codes with 6% to 33% overhead, J. Lightwave Technol. **32**(10), 1999–2002 (2014)
7.209 C. Häger, A. Graell i Amat, H.D. Pfister, A. Alvarado, F. Brännström, E. Agrell: On parameter optimization for staircase codes. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2015)
7.210 A. Sheikh, A. Graell i Amat, M. Karlsson: Nonbinary staircase codes for spectrally and energy efficient fiber-optic systems. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2017) p. W1J.1, Los Angeles
7.211 M. Schwartz, P.H. Siegel, A. Vardy: On the asymptotic performance of iterative decoders for product codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2005) pp. 1758–1762, Adelaide
7.212 H.D. Pfister, S.K. Emmadi, K. Narayanan: Symmetric product codes. In: *Proc. Inform. Theory Appl. Workshop (ITA)* (2015) pp. 282–290, San Diego
7.213 C. Häger, H.D. Pfister, A. Graell i Amat, F. Brännström: Density evolution and error floor analysis for staircase and braided codes. In:

Proc. Opt. Fiber Commun. Conf. (OFC) (2016) pp. 1–3, Los Angeles

7.214 C. Häger, H.D. Pfister, A. Graell i Amat, F. Brännström: Deterministic and ensemble-based spatially-coupled product codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2016) pp. 2114–2118, Barcelona

7.215 C. Häger, H.D. Pfister, A. Graell i Amat, F. Brännström: Density evolution for deterministic generalized product codes on the binary erasure channel at high rates, IEEE Trans. Inf. Theory **63**(7), 4357–4378 (2017)

7.216 D. Chase: Class of algorithms for decoding block codes with channel measurement information, IEEE Trans. Inf. Theory **18**(1), 170–182 (1972)

7.217 C. Häger, H.D. Pfister: Approaching miscorrection-free performance of product codes with anchor decoding, IEEE Trans. Commun. **66**(7), 2797–2808 (2018)

7.218 J. Hagenauer: The turbo principle: Tutorial introduction and state of the art. In: *Proc. Int. Symp. Turbo Codes Iterative Inform. Process. (ISTC)* (1997), Brest

7.219 Y.Y. Jian, H.D. Pfister, K.R. Narayanan: Approaching capacity at high rates with iterative hard-decision decoding. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2012) pp. 2696–2700, Cambridge

7.220 D. Truhachev, A. Karami, L. Zhang, F. Kschischang: Decoding analysis accounting for mis-corrections for spatially-coupled split-component codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2016) pp. 2124–2128, Barcelona

7.221 L.M. Zhang, F.R. Kschischang: Low-complexity soft-decision concatenated LDGM-staircase FEC for high-bit-rate fiber-optic communication, J. Lightwave Technol. **35**(18), 3991–3999 (2017)

7.222 M. Barakatain, F.R. Kschischang: Low-complexity concatenated LDPC-staircase codes, J. Lightwave Technol. **36**(12), 2443–2449 (2018)

7.223 A. Sheikh, A. Graell i Amat, G. Liva: Iterative bounded distance decoding of product codes with scaled reliability. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2018), Rome

7.224 A. Sheikh, A. Graell i Amat, G. Liva: On low-complexity decoding of product codes for high-throughput fiber-optic systems. In: *Proc. Int. Symp. Turbo Codes Iterative Inform. Process. (ISTC)* (2018), Hong Kong

7.225 C. Fougstedt, A. Sheikh, A. Graell i Amat, G. Liva, P. Larsson-Edefors: Energy-efficient soft-assisted product decoders. In: Proc. Opt. Fiber Commun. Conf. (OFC), p. W3H.6. https://arxiv.org/abs/1810.12054 (2019)

7.226 G. Ungerboeck, I. Csajka: On improving data-link performance by increasing channel alphabet and introducing sequence decoding. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (1976), Ronneby

7.227 G. Ungerboeck: Channel coding with multilevel/phase signals, IEEE Trans. Inf. Theory **28**(1), 55–67 (1982)

7.228 E. Biglieri, D. Divsalar, P.J. McLane, M.K. Simon: *Introduction to Trellis-Coded Modulation with Applications*, 2nd edn. (Prentice Hall, Englewood Cliffs 1992)

7.229 P. Robertson, T. Worz: Bandwidth-efficient turbo trellis-coded modulation using punctured component codes, IEEE J. Sel. Areas Commun. **16**(2), 206–218 (1998)

7.230 L.-F. Wei: Trellis-coded modulation with multidimensional constellations, IEEE Trans. Inf. Theory **33**(4), 483–501 (1987)

7.231 S. Benedetto, G. Olmo, P. Poggiolini: Trellis coded polarization shift keying modulation for digital optical communications, IEEE Trans. Commun. **43**(234), 1591–1602 (1995)

7.232 M. Magarini, R.J. Essiambre, B.E. Basch, A. Ashikhmin, G. Kramer, A.J. de Lind van Wijngaarden: Concatenated coded modulation for optical communications systems, IEEE Photon. Technol. Lett. **22**(16), 1244–1246 (2010)

7.233 H. Imai, S. Hirakawa: A new multilevel coding method using error-correcting codes, IEEE Trans. Inf. Theory **23**(3), 371–377 (1977)

7.234 U. Wachsmann, R.F.H. Fischer, J.B. Huber: Multilevel codes: theoretical concepts and practical design rules, IEEE Trans. Inf. Theory **45**(5), 1361–1391 (1999)

7.235 E. Zehavi: 8-PSK trellis codes for a Rayleigh channel, IEEE Trans. Commun. **40**(5), 873–884 (1992)

7.236 X. Li, J.A. Ritcey: Bit-interleaved coded modulation with iterative decoding, IEEE Commun. Lett. **1**(6), 169–171 (1997)

7.237 X. Li, J.A. Ritcey: Trellis-coded modulation with bit interleaving and iterative decoding, IEEE J. Sel. Areas Commun. **17**(4), 715–724 (1999)

7.238 S. ten Brink, J. Speidel, R.H. Han: Iterative demapping for QPSK modulation, Electron. Lett. **34**(15), 1459–1460 (1998)

7.239 A. Bennatan, D. Burshtein: Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels, IEEE Trans. Inf. Theory **52**(2), 549–583 (2006)

7.240 G.D. Forney, F. Ungerboeck: Modulation and coding for linear Gaussian channels, IEEE Trans. Inf. Theory **44**(6), 2384–2415 (1998)

7.241 A.R. Calderbank, L.H. Ozarow: Nonequiprobable signaling on the Gaussian channel, IEEE Trans. Inf. Theory **36**(4), 726–740 (1990)

7.242 G.D. Forney: Trellis shaping, IEEE Trans. Inf. Theory **38**(2), 281–300 (1992)

7.243 F. Steiner, G. Böcherer: Comparison of geometric and probabilistic shaping with application to ATSC 3.0. In: *Proc. Int. ITG Conf. Source Channel Coding (SCC)* (2017), Hamburg

7.244 B.P. Smith, F.R. Kschischang: A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications, J. Lightwave Technol. **30**(13), 2047–2053 (2012)

7.245 L. Beygi, E. Agrell, J.M. Kahn, M. Karlsson: Rate-adaptive coded modulation for fiber-optic communications, J. Lightwave Technol. **32**(2), 333–343 (2014)

7.246 F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, W. Idler: Rate adaptation and reach

increase by probabilistically shaped 64-QAM: An experimental demonstration, J. Lightw. Technol. **34**(7), 1599–1609 (2016)
7.247 M.P. Yankov, F.D. Ros, E.P. da Silva, S. Forchhammer, K.J. Larsen, L.K. Oxenløwe, M. Galili, D. Zibar: Constellation shaping for WDM systems using 256QAM/1024QAM with probabilistic optimization, J. Lightwave Technol. **34**(22), 5146–5156 (2016)
7.248 A. Ghazisaeidi, I. Fernandez de Jauregui Ruiz, R. Rios-Müller, L. Schmalen, P. Tran, P. Brindel, A.C. Meseguer, Q. Hu, F. Buchali, G. Charlet, J. Renaudier: Advanced C+L-band transoceanic transmission systems based on probabilistically shaped PDM-64QAM, J. Lightwave Technol. **35**(7), 1291–1299 (2017)
7.249 A. Sheikh, A. Graell i Amat, G. Liva: Probabilistically-shaped coded modulation with hard decision decoding for coherent optical systems. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2017), Gothenburg
7.250 A. Sheikh, A. Graell i Amat, G. Liva, F. Steiner: Probabilistic amplitude shaping with hard decision decoding and staircase codes, J. Lightwave Technol. **36**(9), 1689–1697 (2018)
7.251 G. Böcherer, F. Steiner, P. Schulte: Bandwidth efficient and rate-matched low-density parity-check coded modulation, IEEE Trans. Commun. **63**(12), 4651–4665 (2015)
7.252 Alcatel-Lucent: *The 400G photonic service engine – leaping toward a future of faster speeds and higher capacity*, Alcatel-Lucent Strategic White Paper, Tech. Rep. (2012)
7.253 A. Leven, F. Vacondio, L. Schmalen, S. ten Brink, W. Idler: Estimation of soft FEC performance in optical transmission experiments, IEEE Photon. Technol. Lett. **20**(23), 1547–1549 (2011)
7.254 W. Idler, F. Buchali: Higher-order modulation formats – concepts and enabling devices. In: *Fibre Optic Communication*, Springer Ser. Opt. Sci., Vol. 161, ed. by H. Venghaus, N. Grote (Springer, Heidelberg, Berlin, New York 2017) pp. 291–357
7.255 F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, W. Idler: Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration, J. Lightwave Technol. **34**(7), 1599–1609 (2016)
7.256 A. Alvarado, E. Agrell, D. Lavery, R. Maher, P. Bayvel: Replacing the soft-decision FEC limit paradigm in the design of optical communication systems, J. Lightwave Technol. **33**(20), 4338–4352 (2015)
7.257 A. Alvarado, E. Agrell, D. Lavery, R. Maher, P. Bayvel: Corrections to 'replacing the soft-decision FEC limit paradigm in the design of optical communication systems', J. Lightwave Technol. **34**(2), 722 (2016)
7.258 L. Schmalen, A. Alvarado, R. Rios-Müller: Predicting the performance of nonbinary forward error correction in optical transmission experiments. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2016) p. M2A.2
7.259 L. Schmalen, A. Alvarado, R. Rios-Müller: Performance prediction of nonbinary forward error correction in optical transmission experiments, J. Lightwave Technol. **35**(4), 1015–1027 (2017)
7.260 L. Schmalen: Performance metrics for communication systems with forward error correction. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2018), Rome
7.261 M. Franceschini, G. Ferrari, R. Raheli: Does the performance of LDPC codes depend on the channel?, IEEE Trans. Commun. **54**(12), 2129–2132 (2006)
7.262 I. Sason, B. Shuval: On universal LDPC code ensembles over memoryless symmetric channels, IEEE Trans. Inf. Theory **57**(8), 5182–5202 (2011)
7.263 A. Sanaei, M. Ramezani, M. Ardakani: On the design of universal LDPC codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2008) pp. 802–806, Toronto
7.264 S.H. Hassani, R. Urbanke: Universal polar codes. In: *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (2014) pp. 1451–1455, Honolulu
7.265 T.A. Eriksson, T. Fehenberger, P.A. Andrekson, M. Karlsson, N. Hanik, E. Agrell: Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments, J. Lightwave Technol. **34**(9), 2256–2266 (2016)
7.266 A. Alvarado, T. Fehenberger, B. Chen, F.M.J. Willems: Achievable information rates for fiber optics: applications and computations, J. Lightwave Technol. **36**(2), 424–439 (2018)
7.267 J. Cho, L. Schmalen, P. Winzer: Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2017)
7.268 W. Idler, F. Buchali, L. Schmalen, E. Lach, R.-P. Braun, G. Böcherer, P. Schulte, F. Steiner: Field trial of a 1 Tb/s super-channel network using probabilistically shaped constellations, J. Lightwave Technol. **35**(8), 1399–1406 (2017)
7.269 L. Schmalen, F. Buchali, A. Leven, S. ten Brink: A generic tool for assessing the soft-FEC performance in optical transmission experiments, IEEE Photon. Technol. Lett. **24**(1), 40–42 (2012)
7.270 N. Stojanovic, Y. Zhao, D. Chang, Z. Xiao, F. Yu: Reusing common uncoded experimental data in performance estimation of different FEC codes, IEEE Photon. Technol. Lett. **25**(24), 2494–2497 (2013)
7.271 J. Hou, P.H. Siegel, L.B. Milstein, H.D. Pfister: Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes, IEEE Trans. Inf. Theory **49**(9), 2141–2155 (2003)
7.272 R. Rios-Müller, J. Renaudier, L. Schmalen, G. Charlet: Joint coding rate and modulation format optimization for 8QAM constellations using BICM mutual information. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2015) p. W3K.4
7.273 R. Rios-Müller, J. Renaudier, P. Tran, G. Charlet: Experimental comparison of two 8-QAM constellations at 200 Gb/s over ultra long-haul transmission link. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2014) p. P.5.1, Cannes
7.274 Z. Wang: Super-FEC codes for 40/100 Gbps networking, IEEE Commun. Lett. **16**(12), 2056–2059 (2012)
7.275 G. Tzimpragos, C. Kachris, I.B. Djordjevic, M. Cvijetic, D. Soudris, I. Tomkos: A survey on FEC codes for 100 G and beyond optical networks, IEEE Commun. Surv. Tutor. **18**(1), 209–221 (2016)

7.276 K. Lee, H.-G. Kang, J.-I. Park, H. Lee: 100GB/S two-iteration concatenated BCH decoder architecture for optical communications. In: *Proc. IEEE Workshop Signal Process. Syst. (SiPS)* (2010) pp. 404–409
7.277 M. Scholten, T. Coe, J. Dillard, F. Chang: Enhanced FEC for 40G / 100G. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2009), Vienna, presentation in Workshop WS-1
7.278 J. Li, K. Liu, S. Lin, K. Abdel-Ghaffar: Algebraic quasi-cyclic LDPC codes: Construction, low error-floor, large girth and a reduced-complexity decoding scheme, IEEE Trans. Commun. **62**(8), 2626–2637 (2014)
7.279 Viasat Inc.: Viasat SDFEC 66200 – 200 Gbps SDFEC, https://www.viasat.com/sites/default/files/media/documents/sdfec_66200_datasheet_003_web_0.pdf (2017)
7.280 B. Smith, I. Lyobomirsky, S. Bhoja: Leveraging 400G ZR FEC technology. In: *IEEE 802.3 Beyond 10km Optical PHYs Study Group* (2017), Orlando
7.281 Viasat Inc.: ECC66100 Series SD-FEC Encoder/Decoder Cores, https://www.viasat.com/sites/default/files/media/documents/ecc_66100_datasheet_2_pgr_007_web_0.pdf (2017)
7.282 K. Cushon, P. Larsson-Edefors, P. Andrekson: Low-power 400-Gbps soft-decision LDPC FEC for optical transport networks, J. Lightwave Technol. **34**(18), 4304–4311 (2016)
7.283 K. Cushon, P. Larsson-Edefors, P. Andrekson: Improved low-power LDPC FEC for coherent optical systems. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2017) pp. 1–3, Gothenburg
7.284 K. Onohara, Y. Miyata, K. Sugihara, T. Sugihara, K. Kubo, H. Yoshida, K. Koguchi, T. Mizuochi: Implementation of soft-decision forward error correction for 100G digital coherent system. In: *Proc. Opto-Electron. Commun. Conf. (OECC)* (2011) pp. 423–424

Alexandre Graell i Amat

Dept. of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden
alexandre.graell@chalmers.se

Alexandre Graell i Amat received his PhD from Politecnico di Torino, Italy, in 2004. He is currently a professor at the Department of Electrical Engineering of Chalmers University of Technology, Sweden. His research interests are in the field of coding theory and its application to distributed storage and computing, privacy, and fiber-optic communications.

Laurent Schmalen

Communications Engineering Lab (CEL)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
laurent.schmalen@kit.edu

Laurent Schmalen received his PhD from RWTH Aachen University in Germany in 2011. Until 2019, he worked as a researcher and Department Head at Nokia Bell Labs in Stuttgart, Germany. He is now a professor at Karlsruhe Institute of Technology (KIT). His work focuses on forward error correcting codes, digital modulation schemes, and machine learning algorithms for high-speed data transmission.

8. Optical Node Architectures

Brandon Collings, Mark Filer

This chapter provides an overview of prominent optical network node architectures beginning with basic nonwavelength channel reconfigurable nodes through to the most advanced reconfigurable node configurations. The generic functionality of an optical network node is summarized as it pertains to network performance (Sect. 8.1), functionality and network topology, including characteristics such as colorless, directionless and contentionless switching and flexible spectrum channel definition. A variety of node architectures are then described in detail (Sect. 8.2), highlighting the capabilities and flexibilities these architectures provide and what the key devices necessary to construct each node architecture are and their attributes. A general overview of how the performance of the overall network depends upon the various node architectures and the implementing devices (Sect. 8.3), including the addition of optical noise, optical filtering and optical channel crosstalk is provided, followed by a description of general approaches to optically monitoring and validating the optical signals traversing the node. Finally, a summary of general forward-looking trends for reconfigurable optical network nodes is provided.

Optical communication in essence involves the transportation of optical, data-carrying signals between two desired locations separated by a physical distance, generally over optical fiber, in order to provide communication of that data between those two locations. Optical communication networking, however, refers to an optical transportation infrastructure that provides optical communication between multitudes of separate locations using a common infrastructure over which those optical data-carrying signals propagate. In practice, this common infrastructure generally consists of a network of optical nodes located at traffic communication end-points and interconnected by optical fibers. Leveraging such a common infrastructure provides economic advantages as a large number of communication links can be supported between diverse locations by that infrastructure. In addition, an operational advantage is generally further achieved as additional communication links can be introduced onto that network over time as the demand arises, providing the network operator the ability to grow the overall traffic-carrying capability of their network in alignment with their needs. As traffic capacity growth patterns are generally difficult to accurately predict, network operators are generally interested in an optical network infrastructure that provides them the ability to flexibly and efficiently support the introduction of new communication links generally between any pairs of nodes within the network as the need arises and with minimal restrictions from already deployed traffic. This maximizes the growth lifetime of the network and consequently the utility of the network infrastructure investment they have already made.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_8

8.1 Generalized Functions of an Optical Node

The base functionality of an optical network node is to direct each optical signal entering that node to its prescribed and respective exit direction from that node, which may include directing one or more signals to local transceivers to terminate those particular links or directing it to continue through the network node along a prescribed fiber route as depicted in Fig. 8.1. However, and as will be described in this chapter, the extent to which the optical channels can be routed independently and dynamically depends greatly upon the hardware and network management software used to implement the node and the network. The simplest implementations, in which channels are routed through the node and are terminated locally, is essentially permanently decided during the design and initial deployment of the network infrastructure. Such nodes are identified as nonreconfigurable. Intermediately advanced node implementations provide the ability to independently select during the deployment process how each new channel is routed through the node or whether it is locally terminated. This level of reconfigurable flexibility, while requiring greater hardware complexity within the node, provides the ability for the network operator to introduce new capacity onto the network when and where the demand arises, thereby extending the growth lifetime of the network. Such network nodes with the capability to reconfigure how signals are routed within the nodes through management software control are collectively referred to as reconfigurable optical add/drop multiplexer (ROADM) nodes. Finally, highly advanced ROADM node implementations provide the ability to independently and dynamically modify how deployed and active channels are routed through the node through network management software control and without any modifications to the physical configuration or interconnections within the node or network. This level of flexibility provides the network operator the ability to efficiently, reliably and perhaps automatically make adjustments and optimizations to the network while the network is operating in order to improve the efficiency of the network's utilization, to reroute traffic around faulted hardware or fibers, or to dynamically align the deployed capacity levels to time-dependent demand requirements, further extending the growth lifetime and increasing the overall service delivery efficiency of the network infrastructure.

This chapter will describe the dominant optical communication network ROADM node implementations, beginning with a general example of a nonreconfigurable node, progressing through examples of less complex and less flexible ROADM node implementations to the most flexible and dynamic ROADM node implementations. This will include highlighting which aspects of signal routing flexibility are enabled, how that flexibility is physically implemented, and what performance or device characteristics are required to ensure the desired minimum performance capabilities. This chapter will also describe how other important network node functions are implemented, such as optical amplification, channel presence detection, channel monitoring, channel protection and power equalization.

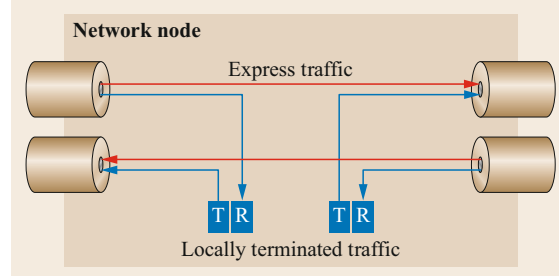


Fig. 8.1 General depiction of an optical network node (T = transmitter, R = receiver)

8.2 General Configuration and Functionality of a ROADM Node

Optical network nodes are generally located at the convergence of multiple pairs of fibers originating from one or more remote locations. These fiber pairs carry the optical wavelength signals that enter and leave the node along respective fibers of each pair. Generally each fiber carries signals propagating unidirectionally. Therefore each fiber pair comprises a bidirectional communication link along that respective physical route, each transporting signal traffic to and from adjacent nodes within the network. Each fiber pair converging upon a network node, and the equipment dedicated to managing the channels entering and exiting the node through those fibers, is described as a *degree* of that node. In a ring network, each node consists of a 2-degree node with each degree respectively connecting that node to the two adjacent nodes (Fig. 8.2).

The general function of a 2-degree node is to control and dictate if each of the optical wavelengths entering each degree will either traverse through the node, from one degree to the other and continue on to the next node within the network, or be routed to a local transceiver to terminate that link at that node. This wavelength signal routing can be implemented at the granularity of each independent wavelength or at a coarser granularity of groups of wavelength channels.

As shown in Fig. 8.2, 2-degree node implementations generally consist of separate hardware configurations comprising each degree. Each configuration includes both the hardware to manage the wavelengths entering/exiting the node through that respective degree, as well as the group of transceivers that terminate/originate selected wavelength traffic that enters/leaves through that same degree. This physical separation of hardware between each degree is intentional as it prevents failure of any hardware element within one degree from impacting the viability of the traffic terminating within the other degree. A typical approach for transporting critical, high-value traffic consists of transporting duplicate copies of that traffic along fully separate and physically diverse routes within the network. This is implemented by utilizing separate transceivers in each degree and routing the bidirectional signal pairs from each transceiver through the physically separate degrees as well as through diverse physical routes within the network to the remote location. Therefore, any single hardware or fiber failure will only impair one of these two (bidirectional) connections and the operational viability of the other (bidirectional) connection will not be impacted, enabling that critical traffic to continue to be transported to the remote location. In general, an electronic switch platform duplicates the traffic within each link endpoint node and provides a copy to both transponders. That same switch platform receives the inbound traffic from both transceivers and selects the healthier traffic to forward out of the node to the network user, providing nearly continuous connectivity despite any single hardware or fiber fault (Fig. 8.3).

Nodes with more than two degrees enable the creation of networks with more complex optical mesh topologies (Fig. 8.4). Such optical nodes are naturally more complicated to implement as wavelength signals entering each degree, if not terminated at a local transceiver, can be routed to one of the other multiple degrees to continue through the network. Similar to a 2-degree node, the hardware of each degree is implemented separately, with each degree containing a complement of transceivers adding/dropping wavelength traffic entering the node through that respective degree. Therefore, a failure in the functionality within

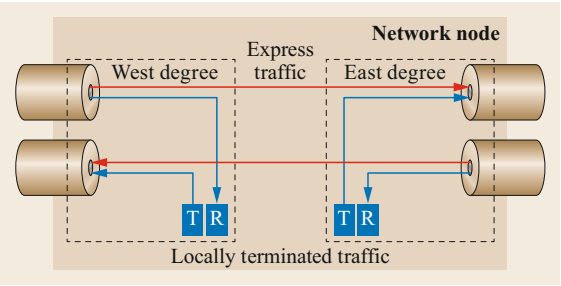


Fig. 8.2 General depiction of an optical network node highlighting East and West degree demarcations

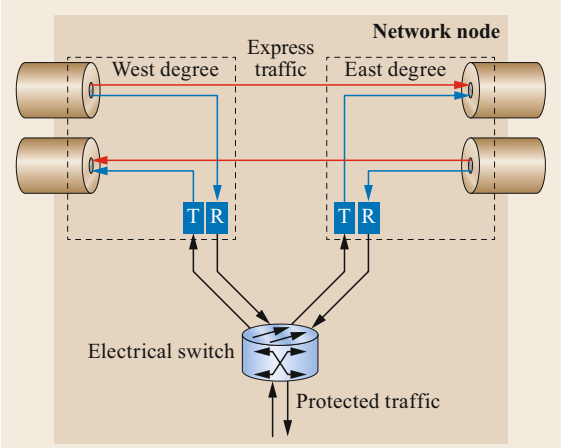


Fig. 8.3 General depiction of an optical network node implementing 1 + 1 protection using dual transceivers over diverse fiber routes

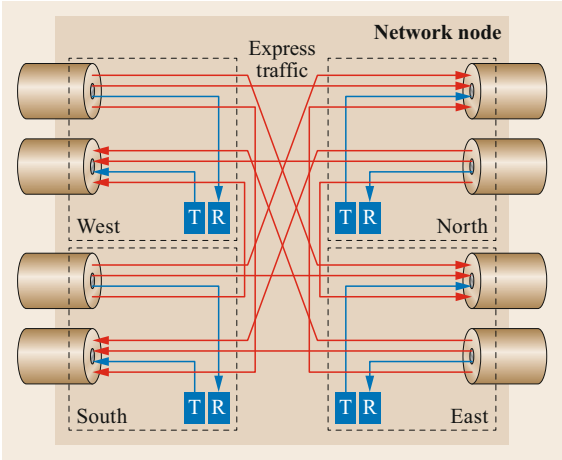


Fig. 8.4 General depiction of a 4-degree node with separate degrees

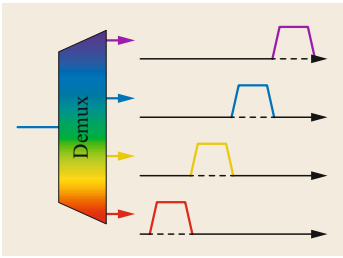


Fig. 8.5 General functionality of an optical demultiplexer (Demux)

any degree does not impact the viability of any of the other traffic within the node passing through any of the other degrees and thus, critical traffic can be diversely routed through the network and its viability will not be significantly impacted by a single fiber or hardware failure.

In order to extract a single wavelength channel from an ensemble of wavelengths (and conversely in order to insert a single channel into an ensemble of wavelengths), the hardware within each degree must be capable of optically demultiplexing (and multiplexing) each single desired wavelength in order to provide only that single extracted wavelength to a transceiver's receiver (Fig. 8.5). Conversely, each degree must be capable of optically multiplexing channels originating at multiple transceivers local to that degree together, along with other wavelength signals from other degrees, onto a single fiber in preparation to launch that ensemble onto the outbound transmission fiber.

As the transceivers are all located locally within each degree along with dedicated optical wavelength multiplexing and demultiplexing capabilities, each transceiver is said to be *directional* given each transceiver is inflexibly connected to the corresponding fiber route of that degree (Fig. 8.4).

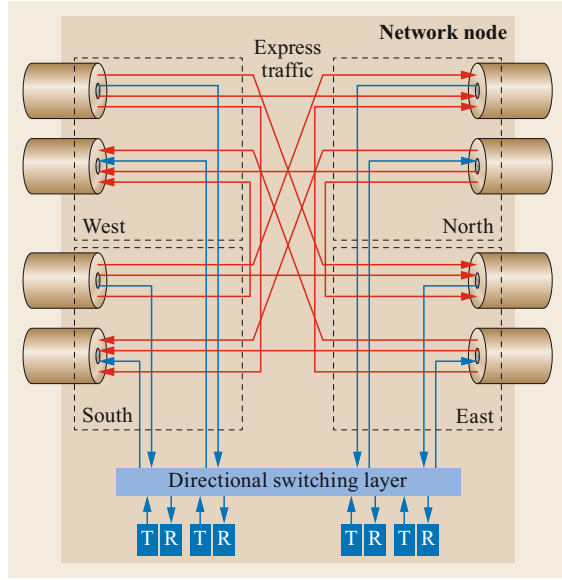


Fig. 8.6 General depiction of a 4-degree node with directionless transceivers

The flexibility of how wavelength channels are physically routed within the network can be enhanced by removing this fiber route inflexibility of which the transceiver's signal can utilize immediately adjacent to the termination node. This requires the transceivers to be physically located separately from the degrees, and an optical switching functionality is introduced between each of those transceivers and each of the node degrees (Fig. 8.6). In this configuration, each transceiver is said to be *directionless* as it can transmit/receive its respective wavelength channel through any degree, and thus fiber route of the node, by selecting the proper configuration of that intermediate switching functionality through the management control software.

When implementing directionless multiplexed/demultiplexed ports, typically multiple multiplexing/demultiplexing modules are deployed so that two diverse and independent paths can be established to carry critical traffic. With this configuration, any single hardware failure cannot impair both the performance and connectivity of the signals along these diverse routes.

The process of optically multiplexing/demultiplexing wavelength channels into/from the full group of channels is said to be *colored* if the multiplexing/demultiplexing add/drop ports that connect to each transceiver permanently support a single particular wavelength channel (Fig. 8.5). In this case, the transceiver, even if it possesses wavelength tunability, must be tuned to that specific channel of that add/drop port in which it is connected in order for that signal to pass through the multiplexer/demultiplexer.

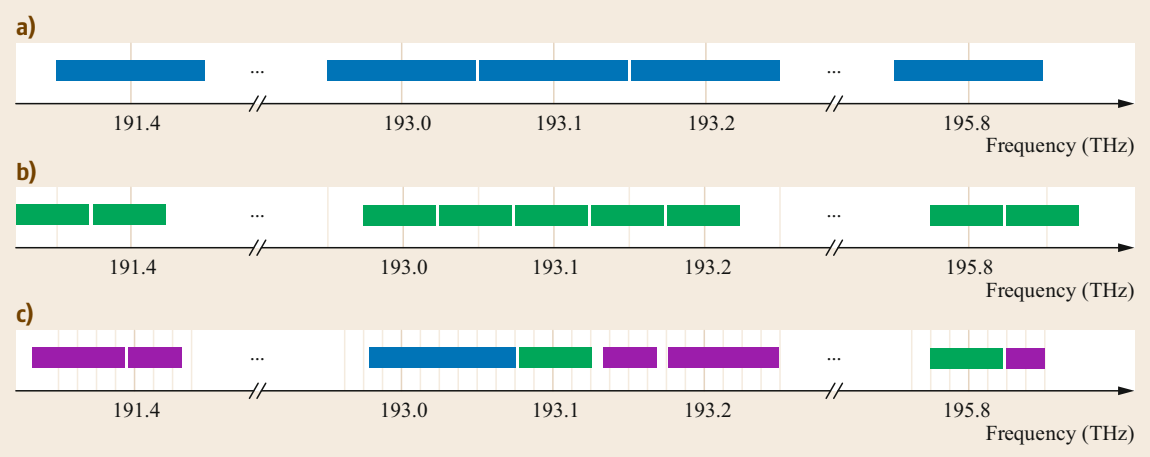


Fig. 8.7a–c Wavelength channel definition for **(a)** 100 GHz fixed wavelength grid, **(b)** 50 GHz fixed wavelength grid, and **(c)** an example of a flex spectrum channel assignment

Therefore, there is no flexibility to alter the operating wavelength of that transceiver without physically disconnecting it from the multiplexer/demultiplexer and reconnecting it to the add/drop port pair associated with the desired new operating wavelength channel. Given a manual intervention of disconnecting and reconnecting the transceiver in order to change the wavelength of operation takes time to manually plan and execute, and errors can be made, such modifications to in-service wavelengths are generally not an operational consideration.

The ability to flexibility modify the operating wavelength is provided by *colorless* multiplexers/demultiplexers, which allow the operating wavelength of the colorless add/drop ports to be specified to any available wavelength channel using the network management control plane. Therefore, the desired operating wavelength channel can be selected after the transceiver is deployed, and the operating wavelength can be modified entirely through the network management software without the need to physically disconnect and reconnect the transceiver.

For a multiplexing/demultiplexing module implementing both directionless and colorless multiplexing/demultiplexing, the module is said to be *contentionless* if the same wavelength channel can be provisioned between separate node degrees and independent respective transceivers (Fig. 8.6). Naturally, if a wavelength channel is already in operation within a given degree, a second independent signal at that same wavelength cannot be routed to that same degree even with a contentionless multiplexer/demultiplexer, as two signals at the same wavelength cannot be carried within a single node degree or fiber. Contentionless multiplexing/demultiplexing simply introduces additional flexibility that can be leveraged by a tunable transceiver, thereby providing greater flexibility to grow and optimize the network traffic.

Since the introduction of dense wavelength division multiplexing (DWDM), the available amplified spectrum has been segmented into optical channels defined by the International Telecommunication Union (ITU) spec G.694.1 [8.1] with channels spaced by 100 or 50 GHz and centered on defined optical frequencies (Fig. 8.7). For each defined channel, the actual usable bandwidth depends upon the quality and quantity of all passband filtering that signal experiences during multiplexing, network wavelength routing and demultiplexing, such that the effective bandwidth is always less than the full defined width of the channel. Despite this, the available bandwidth within a 50 GHz defined wavelength channel has historically been sufficient for optical signals with baud rates below roughly 35 GBd. However, as transceiver technology progresses and optical signal baud rates have continued to increase, the ability to implement wavelength channels with flexibly defined widths and central wavelengths in order to accommodate wider signal bandwidths has emerged (Fig. 8.7). This mode of highly flexible channel definition is commonly referred to *flexible spectrum* or *flexible grid*. The ITU has introduced a finer granularity frequency grid definition [8.1], which allows the definition of channel widths with a granularity of 12.5 GHz and central wavelengths tied to a defined frequency grid. However, some optical systems support channel edge definition with a finer granularity and with an arbitrary total width and center. Note that such systems are equivalently able to support the conventionally defined 50 GHz channel grid if desired. Implementing the flexible grid capability re-

quires that all wavelength-specific elements within the node support the necessary flexible spectrum characteristics, including wavelength multiplexing/demultiplexing, network wavelength management filtering, optical spectrum monitoring, transceiver tunable lasers and network management software.

8.3 General Performance Requirements of a ROADM Node

The general primary functions of a ROADM node are to manage wavelength routing and to amplify signal strengths for propagation and error-free reception. ROADM node design and comprising hardware components thus strive to accomplish these primary functions while imparting the least amount of degradation or distortion of the optical signals in order to enable their respective maximum total propagation distance and avoid costly reception and optical signal regeneration.

The process of optical amplification includes the addition of unwanted optical noise, which can impair the ability to successfully receive and extract the digital information encoded within a signal (Chap. 3). Therefore, typically within a ROADM node, all signals entering the node are immediately amplified so that the least amount of optical noise is added to the signal during that amplification (Fig. 8.8). Generally, following that amplification the wavelength signals are then independently directed to their respective destinations, such as expressed through the network to another degree, or directed towards a local transceiver to be terminated. The wavelength management function generally introduces significant loss to the signals such that, following that function, further amplification is required so that the signals have the desired power levels to exit a degree and be launched onto a transmission fiber (Fig. 8.8). Therefore, in order to minimize the noise added to the signals by this outbound amplification and extend the potential optical transmission range, minimizing the total insertion loss of the wavelength management function is desired and can have a significant influence upon the node architecture and device design, as will be described later in this chapter.

For network efficiency, it is highly desirable to be able to use any unoccupied wavelength channel between two end-point nodes independent of how that wavelength channel is utilized elsewhere in the network. This ability requires that the wavelength management devices provide sufficient blocking and isolation of signals from each other during the process of routing and filtering wavelength channels. Unwanted signal power improperly leaking from one area of the network to another and mixing with an active signal can distort that signal and limit its propagation distance [8.2]. Therefore, minimizing the potential sources of this interference is a critical consideration in the node architecture and in conjunction with the capabilities of the devices that implement that architecture. This will be apparent in the node architectures detailed in this chapter.

Other signal degrading effects must be managed and minimized as part of the ROADM node design and implementation. In particular, additional potential sources of distortion generally include minimizing the accumulation of polarization-dependent loss, polarization mode dispersion, chromatic dispersion and multipath interference. Mitigating these impairments typically involves leveraging optical devices within the node that are intentionally designed with minimal levels of these impairments.

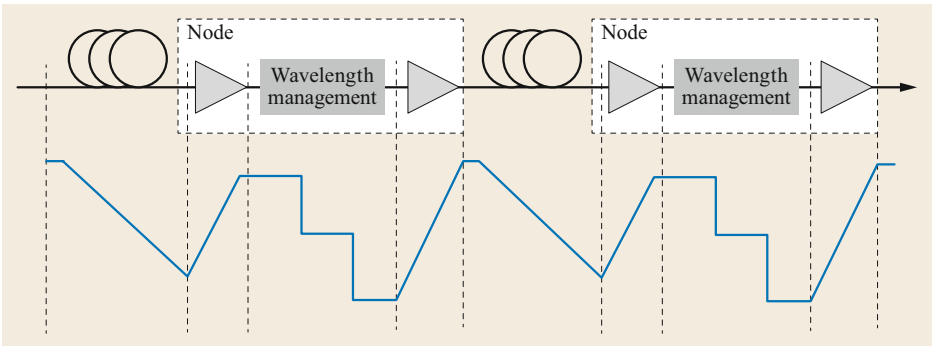


Fig. 8.8 Illustrative channel power (*blue line*) during propagation through fiber spans and optical network nodes

8.4 2-degree Fixed Optical Add/Drop Multiplexer (OADM) Node

The simplest implementation from a hardware perspective of optical add/drop functionality at a node junction is realized by a 2-degree fixed optical add/drop multiplexer (fixed OADM, or FOADM) configuration. It provides the minimum functionality to achieve local adding/dropping or express pass-through of wavelength traffic at that node. Any channels that are not added/dropped at the node are by definition expressed through the node. Historically, there have been a variety of implementations of a fixed OADM topology, but the three most common approaches are to utilize:

1. Individual wavelength multiplexing/demultiplexing filters placed in series, which extract/add the desired channel from/to the remainder of the channel spectrum that passes optically through the node (Fig. 8.9a).
2. Band demultiplexer/multiplexers placed in series, which extract/add a contiguous segment of wavelength channels from/to the remaining channel spectrum that passes optically through the node (Fig. 8.9b).
3. Full demultiplexers/multiplexers with local add/drop paths connected to colocated optical transceivers (Fig. 8.9c) and express wavelengths created by individual optical connections between same-channel demultiplexer outputs and multiplexer inputs.

The first approach utilizing individual add/drop filters relies upon thin-film filter (TFF) technology, which is inherently colored and tied to a fixed ITU grid. As mentioned, each TFF module is a three-port device with a common, express, and single-channel port. On the ingress side, the full spectrum of signals enters, and the desired channel is wavelength-demultiplexed to a receiver, while the rest of the spectrum is reflected and passed through (the reverse occurs for the add function). To realize additional add/drop channels, multiple TFF devices are cascaded in series (Fig. 8.9a). An obvious drawback of this approach is that the insertion loss to the express channels increases linearly with each TFF added in series, and so there is typically a practical upper bound for the number of devices that can be cascaded (i.e., the number of channels that can be added/dropped) at a node. Additionally, the insertion loss for each individual added/dropped channel is different. This is particularly problematic during multiplexing as the transmitted power levels into fiber of the multiplexed and expressed channels can be highly nonuniform.

The third approach using full multiplexer/demultiplexer devices alleviates some of these drawbacks (Fig. 8.9c). Any number of channels may be added/dropped at a given node without impacting the insertion loss of the add, drop, or express path. Express path channels are achieved by simply interconnecting demultiplexer outputs to multiplexer inputs with individual patch cables. Added/dropped channels are implemented by directly connecting local transceivers to the proper multiplex/demultiplex ports. As these devices are typically implemented with arrayed waveguide grating (AWG) devices, the insertion losses do not significantly depend upon the number of channels being locally added/dropped. Transmitted power levels into fiber are uniform, although some consideration may need to be made to equalize the add signals to the express signals.

These approaches have the advantage of being relatively low loss (particularly the first and second approaches when a small number of channels are added/dropped at a node), inexpensive, and compact. However, in addition to the drawbacks mentioned above, perhaps the largest disadvantage is that there is no flexibility in these approaches – the network topology must be fully defined at deployment. Should the desire to augment or reconfigure the add/drop capacity at the node arise, significant manual intervention is required, including manual rearrangement of optical interconnections within the node, and for the first and second approaches the addition of new hardware and full interruption of all expressed traffic. Alluded to above, matching channel power levels between express and add/drop channels is also problematic since the loss of individual channel paths depends on number of TFFs and the position of that channel's TFF within the cascade. This must be mitigated through a very manual process of placing appropriately valued attenuators on the transceiver transmitter based upon a measured spectral uniformity at each node. A transceiver equipped with a variable optical attenuator (VOA) on the transmitter can alleviate the complication of using manual attenuators, but the resulting power levels must still be set and manually measured by a trained technician. Lastly, inherent to these implementations is an inability to achieve any type of true mesh networking at the optical layer – ingress signals must be terminated, electronically switched, and regenerated to achieve degrees greater than two.

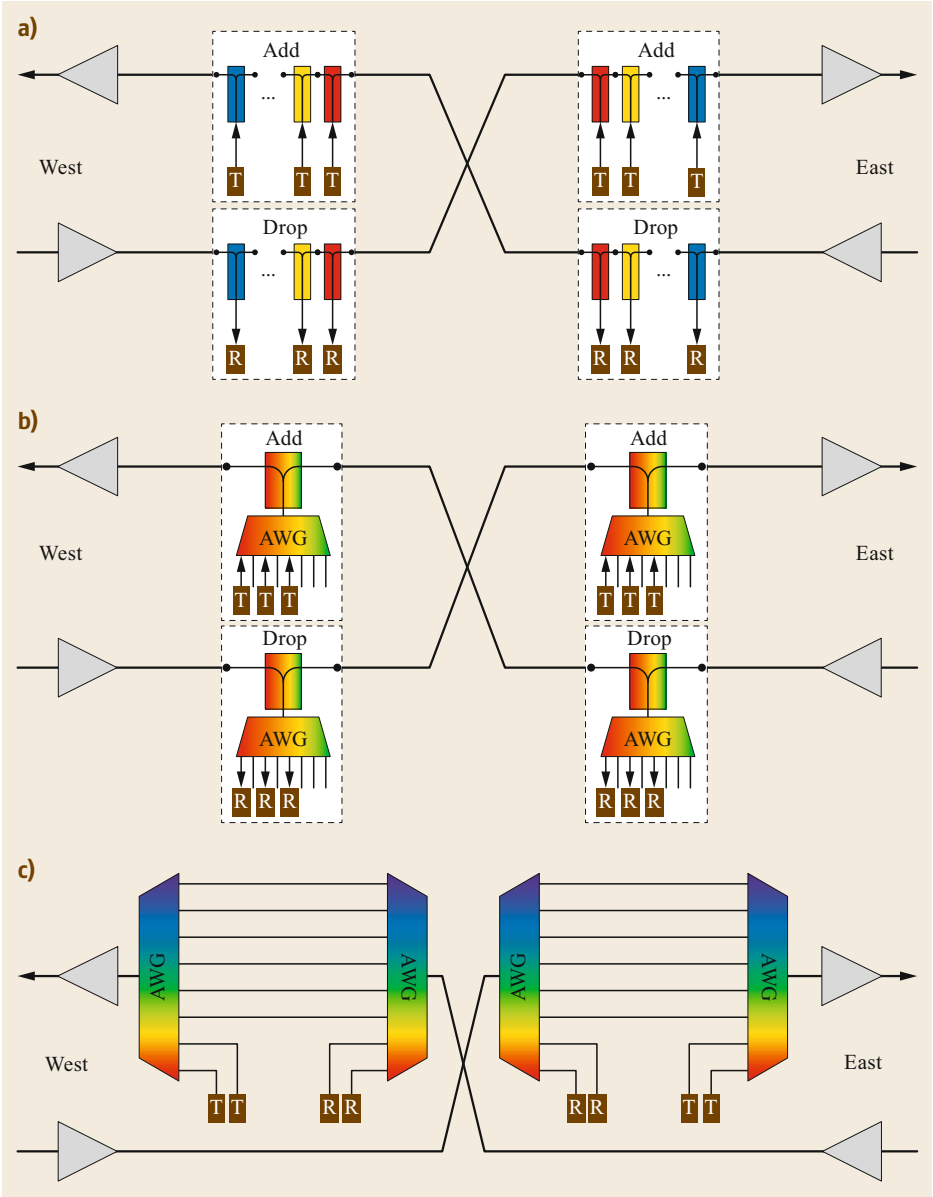


Fig. 8.9a–c FOADM implementations – **(a)** individual wavelength demultiplexers/multiplexers using TFFs, **(b)** band demultiplexer/multiplexer and **(c)** full spectrum demultiplexer/multiplexer with local add/drop (T = transmitter, R = receiver)

8.5 2-degree ROADM Node

In order to improve upon some of the drawbacks inherent to the fixed OADM architectures described above, as the enabling technologies became available optical network system suppliers began offering reconfigurable solutions. Early ROADM implementations address only 2-degree nodes and provide some automated ability to block individual channels from passing from one degree to the other. A copy of all channels entering the node are routed to a demultiplexing AWG so that if a channel is desired to be dropped at this node, a receiver may be connected to the proper drop port. If a channel is dropped, the signal for that channel is blocked from reaching the opposite degree, thus allowing that wavelength channel to be reused with an independent channel in the other degree.

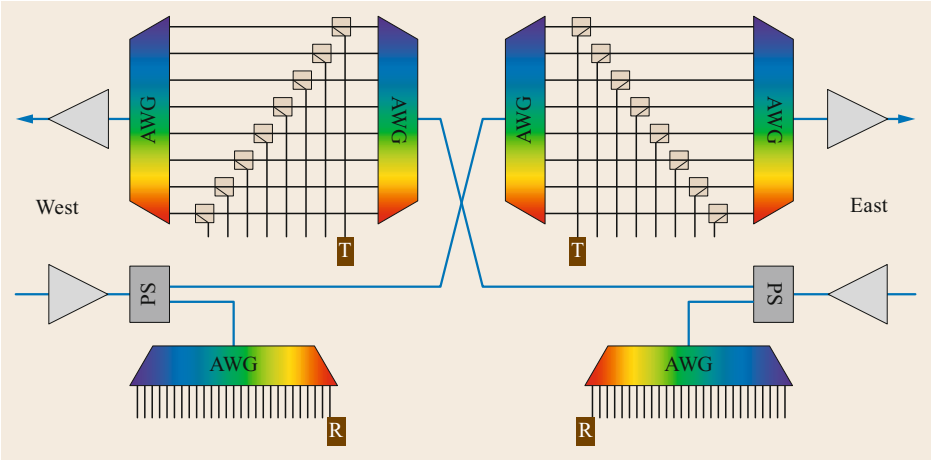


Fig. 8.10 iPLC-based 2-degree ROADM (PS = power splitter)

In general, there are two basic approaches taken, depending on the technology employed. One approach leverages integrated planar lightwave circuit (iPLC) technology and integrates a pair of AWGs and a 2×1 switch between each same wavelength pair of single-channel ports (Fig. 8.10). A third AWG could be integrated within each degree to provide the demultiplexing function. This approach addresses the primary shortcoming of the previous nonreconfigurable topologies in that every wavelength channel could independently be configured to be locally added/dropped or expressed entirely through the management software control without any physical intervention. This allows new channels to be flexibly provisioned quickly and dependably. In addition, this technology allows convenient integration of VOAs within each of the channel add paths for independent channel power equalization. This equalization can be fully automated with the inclusion of optical channel monitors (OCMs) with this configuration.

An alternative variant of a 2-degree ROADM architecture leverages a wavelength blocker (WB), a 2-port device with the ability to independently attenuate and block each wavelength channel. The general functionality of a WB is represented in Fig. 8.11 where all channels entering the device on a single input fiber are separated by wavelength channel, individually and independently attenuated or fully blocked, and then multiplexed to exit the device all on the same output fiber.

This node architecture is illustrated in Fig. 8.12 and is quite similar to the iPLC node architecture (Fig. 8.10). The WB is located on the optical connection between each of the two degrees such that if a wavelength channel is not provisioned to be expressed through the node, the WB fully attenuates that signal so that wavelength channel can be reused by an independent transceiver in the opposite degree (if desired). Similar to the iPLC approach, a separate AWG is used for demultiplexing all channels in each degree. However, with the WB architecture, a second AWG is needed to multiplex all locally added channels in each degree. Individual channel power equalization is accomplished using per-channel VOAs within the multiplexing AWG or by variable attenuation within the WB.

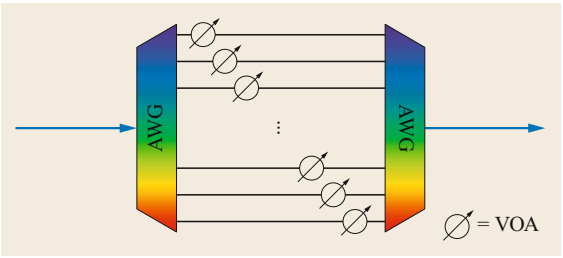


Fig. 8.11 Functional representation of a wavelength blocker (WB)

An enhancement to the 2-degree WB-based node architecture is implemented by replacing the WB and power combiner functionality with a 2×1 wavelength selective switch (WSS) device. A general representation of the functionality of a 2×1 WSS device is shown in Fig. 8.13 (note the optical functionality is fully equivalent independent of in which direction the light is passing) [8.3–6].

In this node configuration (Fig. 8.14), the node ingress signal is amplified and duplicated, with one copy routed to the opposite degree to support express traffic and the other sent to a demultiplexing AWG where the individual drop channels are connected to corresponding transceiver receivers. The express traffic enters one port of the two (or more) WSS ports while the output of an add multiplexer AWG is connected to the other WSS input port. The WSS device is then

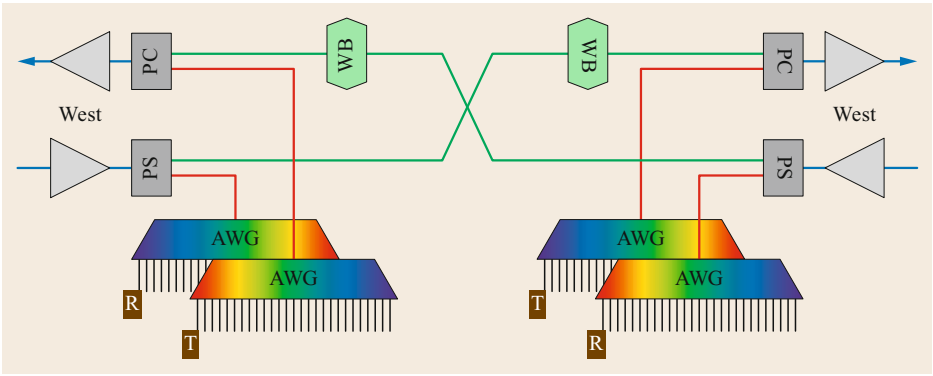


Fig. 8.12 WB-based 2-degree ROADM node (PS = power splitter, PC = power coupler, WB = wavelength blocker)

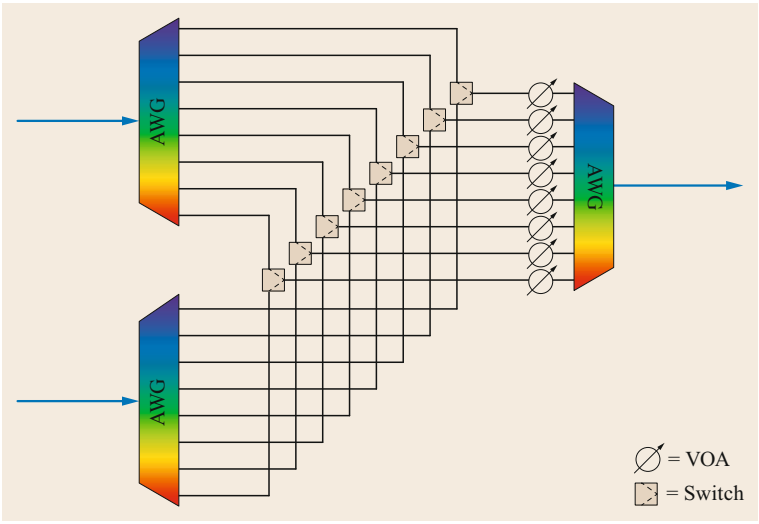


Fig. 8.13 Representation of the principle functionality of the wavelength selective switch (WSS). Note that practical WSS devices are not constructed from the elements as shown

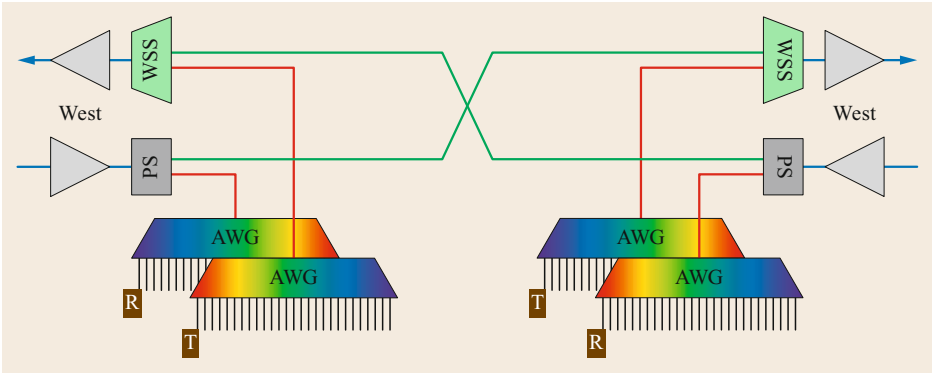


Fig. 8.14 General node architecture for a 2-degree WSS-based ROADM

configured to route any locally added signal from the add port of the WSS to the common port, which provides the collection of output signals to exit that degree. Similarly, it is also configured to route any selected express signals to the common output port. Since the express input of the WSS is receiving the full complement of channels entering the opposing node, the WSS is responsible for selecting only the desired express channels provisioned to pass through to the output, and blocking all others. For this reason, this configuration is known as broadcast-and-select.

Several advantages are immediately apparent compared to fixed OADM architectures. Foremost, this implementation gives the ability to add/drop wave-

lengths independently and programmatically, which in the former case, always requires early detailed topology planning, with very disruptive and error-prone requirements in making changes or augmentations. However, due to the colored and directional nature of this implementation, this flexibility can only be exploited during initial channel provisioning. Once a channel is deployed, generally any alteration requires significant manual intervention and subsequently such modifications are rarely implemented. Another advantage is the fact that add/express channels' power levels can be equalized in an automated fashion, as again, modern ROADM nodes are typically deployed with integrated OCM devices. Due to the broadcast nature of this configuration as implemented with the passive splitter, the express optical signals only encounter a single WSS bandpass per channel at each node. This reduces the amount of bandwidth narrowing the signal experiences, which leads to reduced transmission penalty at the receiver [8.7]. The presence of a colored multiplexer on the add path and a demultiplexer on the drop path minimize the effects of adjacent channel crosstalk and make this architecture ideally suited for direct-detect optical receivers.

8.6 2- to 8-degree Broadcast-and-select Colored and Directional ROADM Node

By utilizing WSSs with port counts beyond $N = 2$ as a building block, topologies with higher degrees of connectivity and flexibility are realized, enabling optical mesh networking. The next evolution of ROADM node architectures commonly featured 1×4, 1×5, 1×8 or 1×9 WSSs, providing up to eight additional degrees of connectivity at a given node. The configuration for the N-degree ROADM is identical to the 2-degree ROADM of the previous section, except now the broadcast power splitter on the drop path is a $1 \times N$ and the *select* WSS on the add path is an $N \times 1$, as pictured in Fig. 8.15. The additional splitter outputs and WSS inputs are used to interconnect other degrees. Local add and drop is achieved with multiplexer/demultiplexer units such as the AWGs discussed earlier. As with the 2-degree ROADM, this implementation is inherently colored with wavelength channels operating on the fixed ITU 50 or 100 GHz grid.

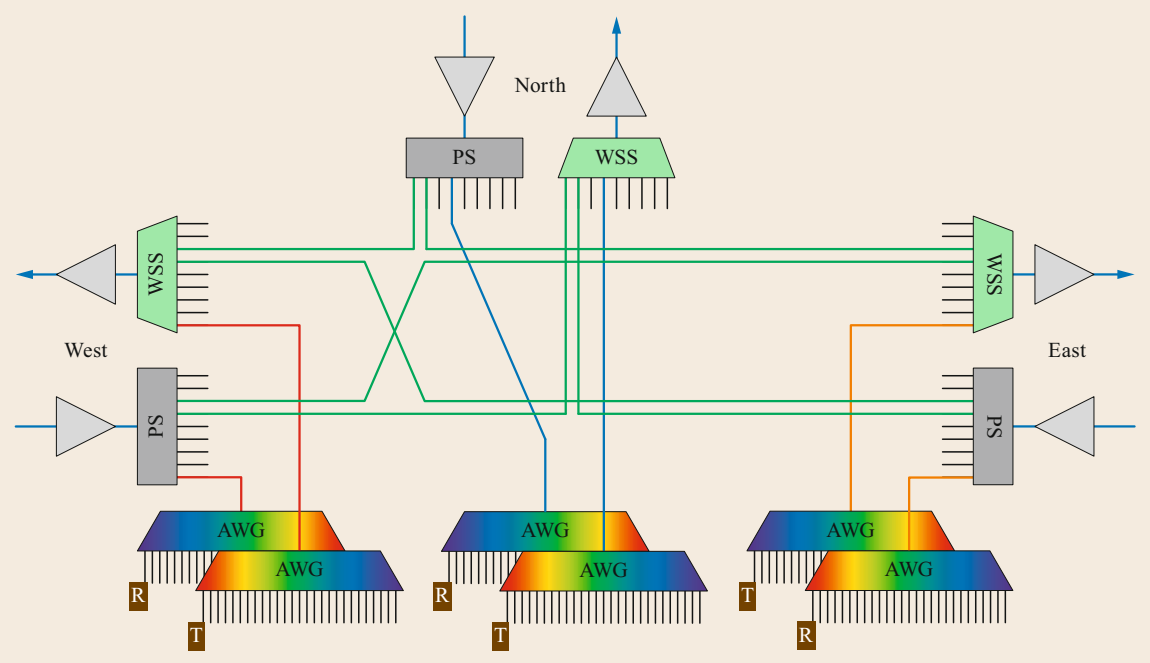


Fig. 8.15 General node architecture for a 3-degree broadcast-and-select node with AWG-based colored/directional add/drop

Fig. 8.15 shows a 3-degree application with local add/drop per degree. The *directional* nature of the topology is apparent, where a multiplexer/demultiplexer pair is located within each degree.

As with the 2-degree broadcast-and-select WSS-based ROADM, the user still has the flexibility to add/drop wavelengths independently, although due to the colored nature of the multiplexer/demultiplexer units, this flexibility can practically only be utilized during initial provisioning or in combination with manually reconnecting a transceiver to a different add/drop port. Additionally, intermediate ROADM nodes can, in principle, dynamically reroute wavelengths to adjacent degrees without any physical intervention. Although the corresponding wavelengths at the terminal nodes would need to be manually reconfigured to transmit/receive the rerouted channels, it alleviates the requirement for manual intervention at the intermediate sites. Other advantages inherited from the 2-degree broadcast-and-select WSS architecture include integrated power monitoring on the add and express paths, single WSS filter passed per node, and per-channel filtering provided by the colored multiplexer/demultiplexer on the add/drop paths.

One challenging aspect of any multidegree broadcast-and-select-based ROADM topology is that for a given direction, a single WSS device must provide sufficient optical isolation across all of its ports to prevent any optical distortion of a provisioned wavelength arising from interference from signal power from insufficiently suppressed signals at that same wavelength

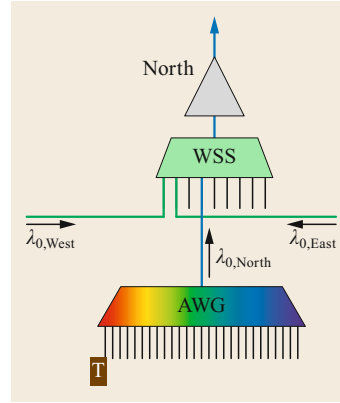


Fig. 8.16 Isolation requirements of broadcast-and-select architectures

channel [8.8]. This requirement can be understood using the system shown in Fig. 8.15. Assume an optical source with wavelength λ_0 is being added to the North direction. In principle, that same wavelength λ_0 could be propagating into the node from both the East and West degrees from adjacent nodes, and either expressed between these directions, or each dropped at the East and West terminal AWGs. Regardless, due to the broadcast nature of broadcast-and-select ROADMs' splitters, nominally powered signals with wavelength λ_0 are also present at two additional North WSS input ports ($\lambda_{0,East}$ and $\lambda_{0,West}$). This situation for the North WSS is captured in Fig. 8.16, which must be able to sufficiently block the unwanted λ_0 signals from the East and West degrees such that the λ_0 signal propagating northward experiences no interference [8.9].

8.7 2- to 8-degree Broadcast-and-select, Colorless and Directionless ROADM Node

The broadcast-and-select architecture shown in the previous section supports mesh network nodes and flexible wavelength configuration between degrees at each node. However, as indicated, all transceivers are located within a particular degree and consequently have no directional flexibility. Implementing directionless transceivers provides the network operator an additional level of flexibility to reprovision existing wavelengths along new routes, including modifying the route used directly adjacent to both the terminating nodes. Similarly, including colorless multiplexing/demultiplexing capability further increases reprovisioning flexibility by allowing the selection of a new wavelength channel. This route and wavelength reprovisioning flexibility provides a greater number of options for that signal to be routed between two network nodes and thus a great probability that a wavelength and route combination can be available despite the existing population of other channels already operating within the network.

To implement directionless multiplexing/demultiplexing, each transceiver signal must be able to be configured to enter/exit the node from any degree. Therefore, they cannot be physically located within any specific degree. One approach is to utilize the general broadcast-and-select degree structure as the switching stage between the array of transceivers and all of the degrees (as shown in Fig. 8.17). In this node construction, a complete copy of the incoming signals entering each degree is provided to the colorless and directionless (CD) multiplexer/demultiplexer module, in the same manner that a copy is also provided to all other node degrees. All these copies are provided to the M inputs of an $M \times 1$ WSS, which then only allows those wavelength channels intended to be demultiplexed and

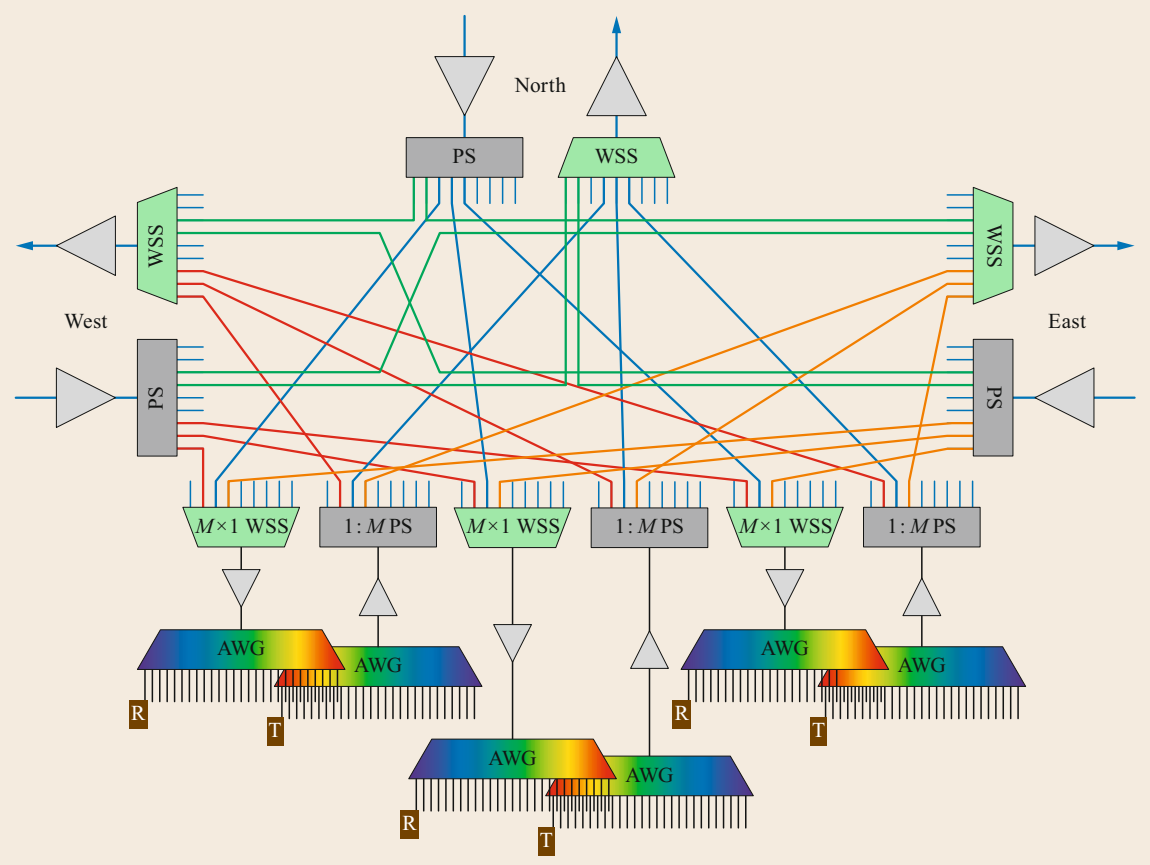


Fig. 8.17 General node architecture for a 3-degree broadcast-and-select node with AWG-based CD add/drop

received by the transceivers connected to that CD module. All other wavelength signals must be blocked by the WSS so that none interfere with those express wavelengths. Note this is the same function that the $M \times 1$ WSS performs in the node degrees [8.2, 5].

To demultiplex the signals and provide them to the individual transceivers, the common port of the $M \times 1$ WSS is connected to a demultiplexing element such as an AWG (Fig. 8.17) for colored demultiplexing or a $1 \times n$ WSS (Fig. 8.18) for filtered colorless demultiplexing. In both cases, the AWG or the $1 \times n$ WSS optically filter and suppress all wavelength signals except for the particular channel to be received by the respective receiver. This prevents any interference from other wavelength channels within receivers that are sensitive to total optical signal power.

With the advent and use of coherent transmission technology [8.10], colorless wavelength demultiplexing can be significantly simplified as coherent receivers have the inherent ability of limiting their sensitivity to optical signal power existing within a very narrow wavelength range [8.11, 12]. In general, coherent receivers operate by optically mixing the incoming signal (or collection of a multitude of signals) with a powerful, locally generated, narrow-linewidth continuous-wave signal and then limiting a detection to only the resulting baseband signal within a narrow radio frequency (RF) range determined by the baud rate of the encoded signal [8.13–15]. Thus, the coherent receiver's sensitivity predominantly lies within the narrow wavelength region about the wavelength of the respective local oscillator and signal power outside of this wavelength region is nominally ignored. Therefore, this capability allows the coherent receiver to select and suitably receive a single channel from a multitude of incident wavelength channels [8.16, 17]. This capability allows for the removal of optical filtering of unwanted channels at a receiver, thereby significantly simplifying the demultiplexing requirements. As shown in Fig. 8.19, the AWG or WSS element can be replaced with a $1 : n$ power splitter to provide colorless demultiplexing (assuming all receivers are suitable coherent receivers capable of isolating a desired channel from a multitude of n channels). Due to the combined insertion loss

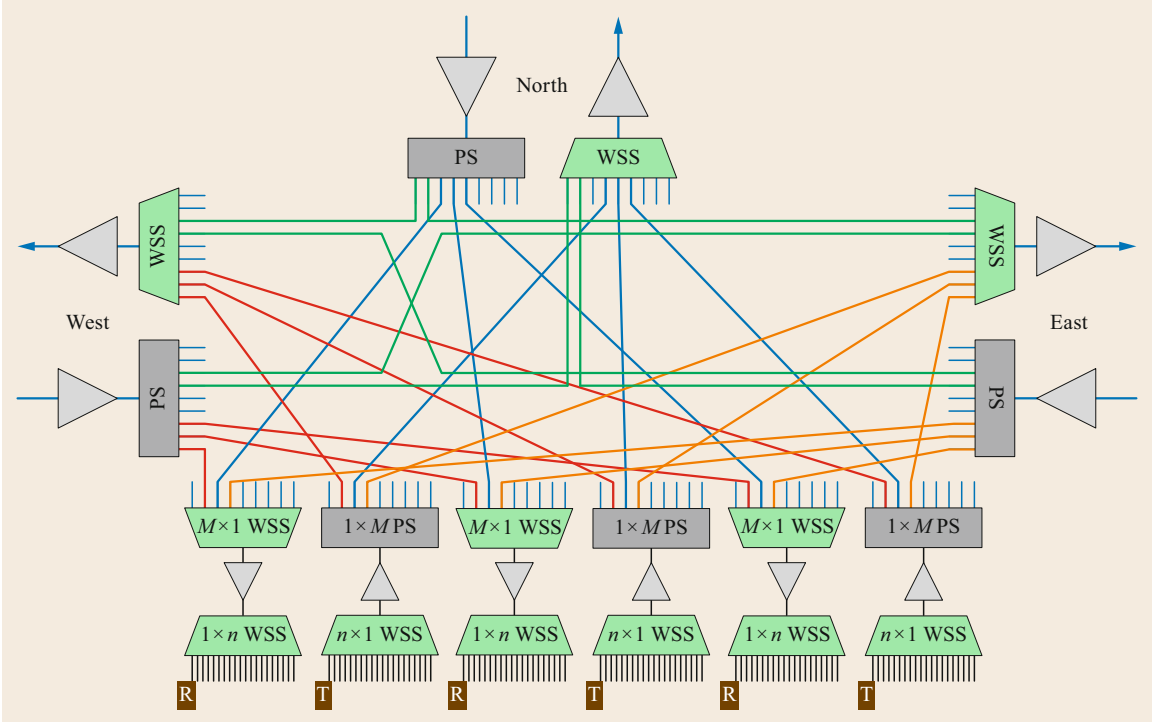


Fig. 8.18 General node architecture for a 3-degree broadcast-and-select node with WSS-based CD add/drop

of the WSS and the AWG, power splitter or WSS demultiplexer, optical amplification is generally required between the WSS and AWG, WSS or power splitter as shown in Figs. 8.17–8.19.

For multiplexing, a similar AWG structure is used for colored multiplexing of the individual signals. These combined signals are then split M ways with a copy routed to each of the degrees, much like the routing between node degrees (Fig. 8.17). The WSS within each degree then will pass only those channels which are provisioned to leave the node through that respective degree. All other channels must be sufficiently blocked to avoid any interference with other channels that are correctly exiting the WSS and the degree. Similar to the demultiplexing configuration, the insertion loss of the AWG and power splitter generally requires the inclusion of an optical amplifier.

To implement filtered colorless multiplexing, a $1 \times n$ WSS is used to passively multiplex n channels onto a single fiber. As with colored multiplexing, this combination of channels is then split M ways to each degree (Fig. 8.18). Alternatively, unfiltered colorless multiplexing is implemented using a $n:1$ power combiner to multiplex n channels together, which are then split M ways (Fig. 8.19).

In both multiplexing and demultiplexing portions of the CD structure, all the signals pass through a single fiber between the $1 \times M$ element facing the degrees and the $1 \times n$ element facing the transceivers. Therefore, this approach possesses wavelength contention within each CD structure such that only a single signal at any wavelength channel from any degree can be provisioned through each respective CD structure. Note that multiple CD structures can be deployed such that each channel can be deployed multiple times through independent degrees. However, when wavelength contention arises for a given transceiver, overcoming that contention requires that the transceiver be physically disconnected and reconnected to a different CD structure where that desired wavelength is available. Therefore, given this need for physical intervention, any wavelength contention is practically only avoidable during the deployment process of a new transceiver and wavelength contention is not avoidable during dynamic or automated channel restoration or reprovisioning activities.

Generally a minimum of two structures are implemented to enable two redundant paths for protected channels. However, as each structure requires a respective port on the WSS within every degree as well as an additional copy of the channels entering each node, significantly increasing the number of CD structures within a node possesses practical limitations.

Within the unfiltered colorless configurations shown in Fig. 8.19, the scale of the $1:n$ power splitter

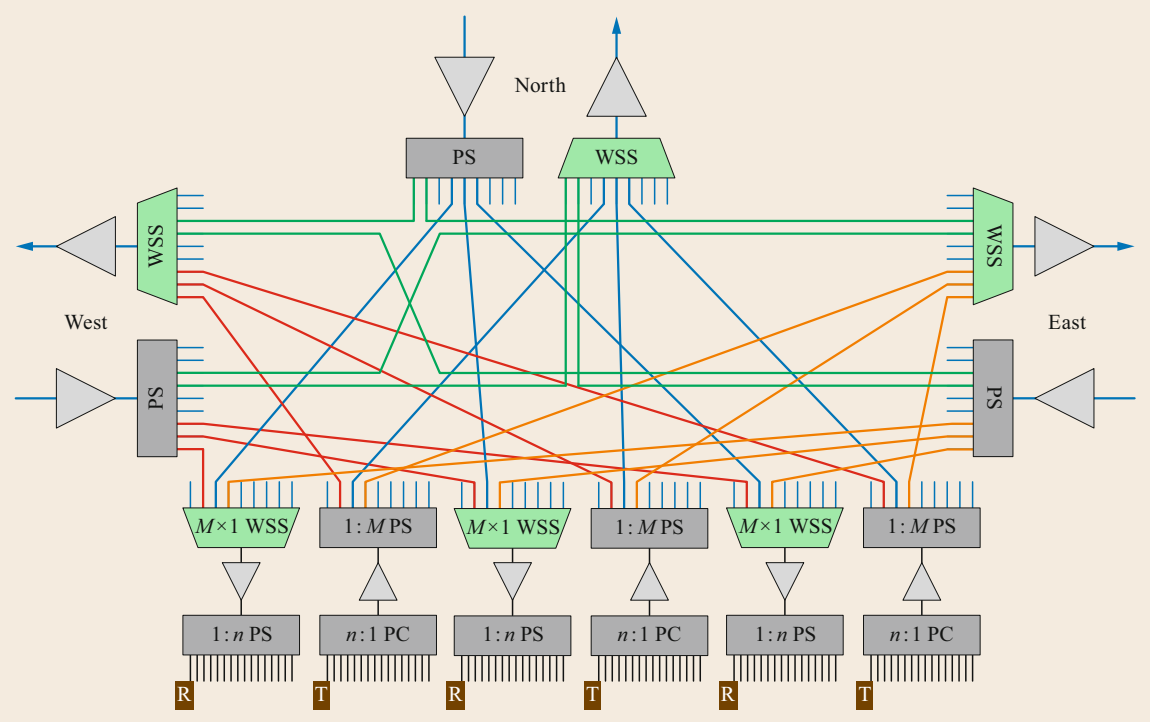


Fig. 8.19 General node architecture for a 3-degree broadcast-and-select node with power splitter and power coupler-based CD add/drop

and coupler determines the number of add/drop ports provided within each CD structure. The ability to increase the number of ports on these splitters/couplers generally depends upon three factors. The first is the need to sufficiently overcome the loss of these passive devices. The second is the limit of the number of channels the coherent receiver can tolerate while adequately isolating the desired channel (typically this limit is due more to limited power dynamic range during mild power transient events). The third limit is the accumulation of out-of-signal-band noise as the channels are combined by the passive coupler as illustrated in Fig. 8.20 [8.18]. This accumulated noise will overlap and interfere with signals and can degrade performance. The accumulation and impact of this noise also directly depends upon the amount of noise produced by the transceivers. In practice, the limit for the number of power splitting/combining ports is around 16 for both coherent receiver dynamic range as well as for noise accumulation within the combiner (assuming a transceiver emitting a signal with a 40 dB Tx_OSNR specification) [8.6].

For unfiltered colorless multiplexing/demultiplexing, the number of add/drop ports within a CD structure can be increased by implementing a second layer of wavelength management as shown in Fig. 8.21. In the

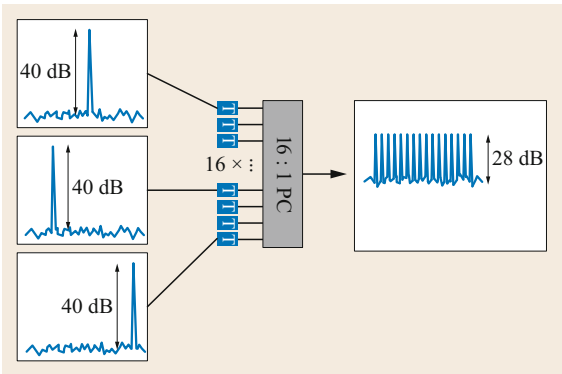


Fig. 8.20 Accumulation of out-of-band signal noise through a power coupler assuming transceivers with a 40 dB Tx_OSNR specification resulting in a potential for a 28 dB OSNR after the passive combination

demultiplexing direction, the common output of the $1 \times M$ WSS is connected to the common port of a second $1 \times n$ WSS. Each WSS output is then connected to a $1 : n$ power splitter. As the WSS is only directing those wavelength signals that are to be received following the power splitter, and blocking all others, the number of signals presented to each coherent receiver is limited to n. Therefore, the total number of add/drop ports

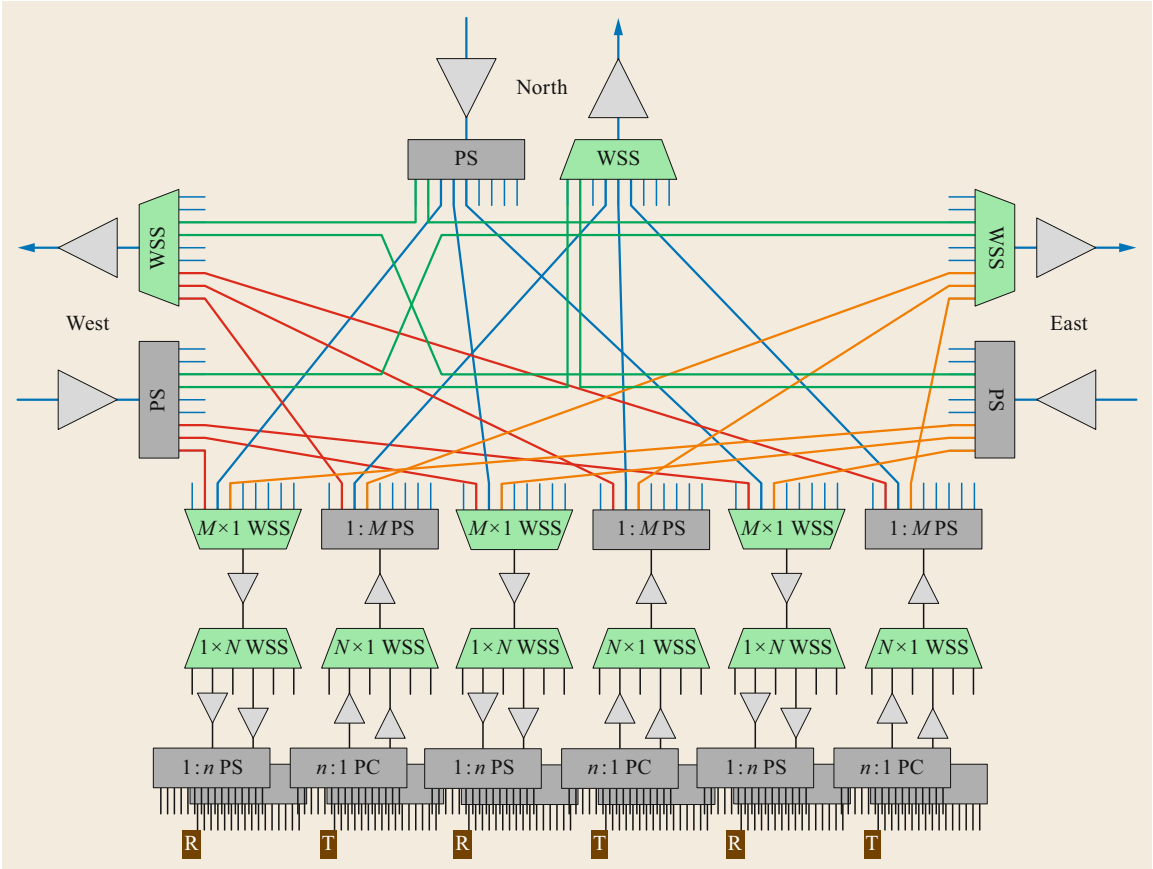


Fig. 8.21 General node architecture for a 3-degree broadcast-and-select node with WSS-based CD add/drop and expansion power splitter and power coupler demultiplexing and multiplexing

that can be supported within each CD structure can be increased N-fold, nominally without significant degradation in performance [8.19, 20]. A similar structure is used in the multiplexing direction (Fig. 8.21). Here the $1 \times N$ WSS is blocking any noise outside the active wavelength signal bands and thus practically eliminates any accumulation of noise outside of those active signal wavelengths bands.

These broadcast-and-select CD architectures have the advantage that they utilize the single $1 \times M$ WSS arrangement within each degree, a simpler arrangement generally with lower implementation cost, and they essentially reuse this same arrangement within the CD structure. Also, operators can deploy networks with a combination of non-CD multiplexing/demultiplexing in some nodes and CD multiplexing/demultiplexing in others.

However, these architectures require a dedicated WSS and splitter port for each CD structure within each degree, in addition to the WSS and splitter ports needed for each other node degree. Therefore, a 4-degree node with four CD structures requires a 1×7 WSS and a $1:7$ power splitter within each degree. An 8-degree node with maximum of eight CD structures requires a 1×15 WSS and a $1:15$ power splitter within each degree. Due to the finite signal isolation provided by practical WSS devices as well as the increased loss introduced by larger power splitters, performance degradation arising from accumulated interference from incompletely blocked signals [8.2] and increased amplification requirements respectively places practical constraints on the size of the WSS and power splitter that can be utilized within each degree. Thus, the number of degrees that can be implemented within a node using the broadcast-and-select approach with CD multiplexing/demultiplexing is practically limited to a number between four and eight, depending upon the desired design maximum number of CD structures [8.20].

8.8 4- to 16-degree Route-and-select, Colorless and Directionless ROADM Node

To address the need to support nodes with more than four degrees with full CD multiplexing/demultiplexing and without performance compromise due to signal distortion, a node architecture is introduced where the broadcasting function within each degree is replaced with the wavelength routing function of a WSS. This architecture is shown in Fig. 8.22 and is generally referred to as a route-and-select architecture. By using a WSS instead of a power splitter, only those wavelength channels which are provisioned to be routed to each respective degree are allowed to pass to that degree. Those that are not provisioned to a respective degree are blocked. Therefore, wavelength signals must pass through two WSS devices (one within the inbound degree and one within the outbound degree) before they are present on an outbound fiber where signal interference would occur and the blocking capability of both WSS devices is aggregated to prevent undesired signals from propagating where they should not. This combined blocking approach has the practical advantage that the isolation capability provided by each WSS need only be half of the total of what is needed to prevent accumulated signal distortion [8.18, 20]. Relaxing the required isolation within a WSS simplifies the ability to design WSS devices with higher port counts, smaller physical dimensions, as well as two independent WSS devices in a single package [8.20]. Additional advantages of this arrangement include reduced insertion loss through the routing WSS relative to a power splitter of the same port count (which reduces the amount of optical amplification needed, which improves performance) [8.20] and the provision of filtering of unwanted signals provided to the CD structure [8.20].

To implement CD demultiplexing within a route-and-select architecture with perhaps limited isolation within the WSSs in each degree, a broadcasting function or power-combing function within the CD add/drop structure may result in unwanted crosstalk between channels of the same wavelength. Therefore, a routing function is generally used within the CD structure as shown in Fig. 8.22. In this approach, the CD add/drop structure is not contentionless given the unitary fiber connection between the $M \times 1$ and $1 \times N$ WSS. Similar to the approach shown in Fig. 8.21, $1 : n$ power splitters and $n : 1$ power couplers are connected to the multichannel outputs of each of the N WSS ports to expand

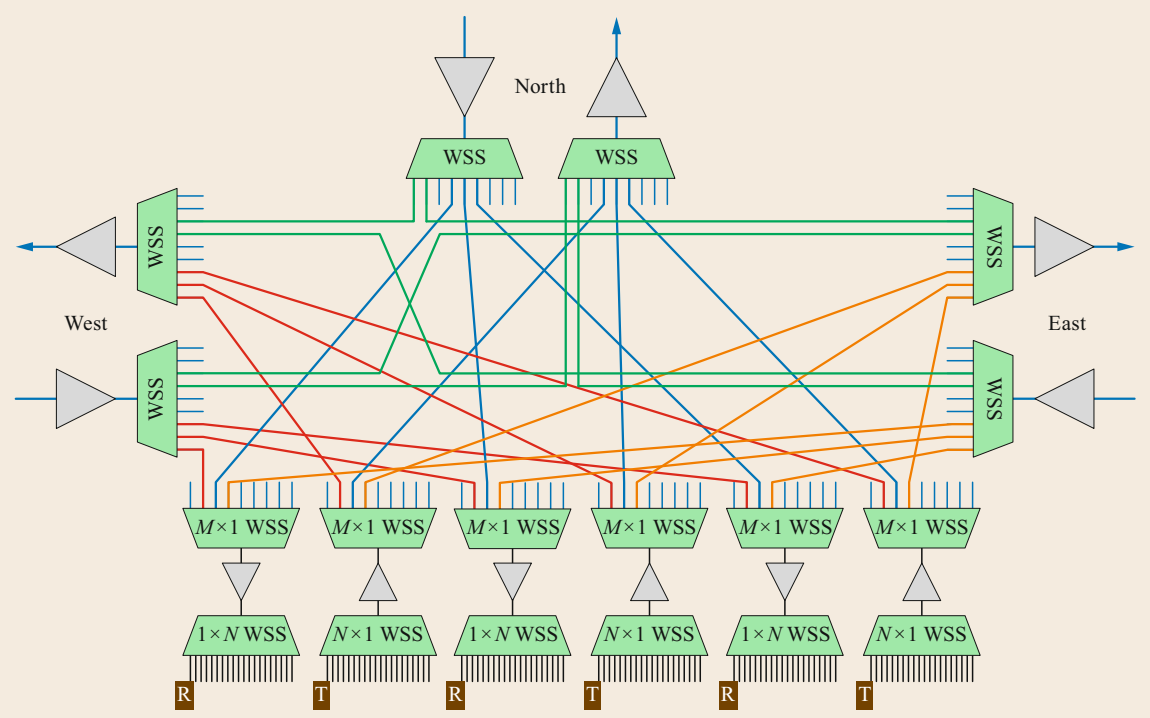


Fig. 8.22 General node architecture for a 3-degree route-and-select node with back-to-back $M \times 1$ and $1 \times N$ WSSs providing CD add/drop

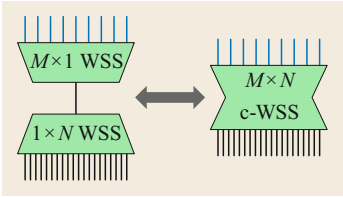


Fig. 8.23 Representation of the principle functionality of a wavelength-contentioned $M \times N$ WSS device ($M \times N$ c-WSS)

each WSS port into n ports supporting n transceivers, where each transceiver receives an independent channel.

The function of the back-to-back $M \times 1$ and $1 \times N$ WSSs (shown in Fig. 8.22) can be more efficiently accomplished by use of an integrated wavelength-contentioned $M \times N$ WSS device [8.21]. The functionality of a contentioned $M \times N$ WSS is shown in Fig. 8.23. The same device can function as a multiplexer or demultiplexer. The architecture utilizing the contentioned $M \times N$ WSS is shown in Fig. 8.24 [8.20]. This approach has the advantages that the contentioned $M \times N$ WSS has nominally the same physical form factor and insertion loss as a single $1 \times N$ WSS and it provides a more compact overall solution and removes the need for the additional amplifier located between the common ports of the $1 \times M$ and $1 \times N$ WSSs.

This architecture, with the availability of 1×20 WSSs within the degrees and 16×5 wavelength-contentioned WSSs within the CD structure enables the construction of practical 16-degree nodes, which support up to five CD structures, each of which can support the full complement of C-band channels (i.e., ninety-six 50 GHz wide wavelength channels). While this scale of node and add/drop is generally sufficient for the majority of network nodes, additional CD structures can be accommodated by adding a second layer of 1×20 WSSs attached to each degree WSS to provide additional ports in which to connect up to 20 additional CD structures. These expansion WSSs can be added after the initial five CD structures are deployed and without interrupting any service already deployed within the node, thus allowing the network operator the confidence to be able to expand any node to its maximum of 16 degrees but also with the flexibility to only need to deploy this expansion equipment once the need is identified for any given node.

Unfortunately, due to the construction of the wavelength-contentioned $M \times N$ WSS device (at time of this printing), when provisioning a connection between one of the M ports and one of the N ports, one or more unwanted connections between other M and other N ports can be made [8.22]. This characteristic is not a concern for the architecture shown in Fig. 8.25 because the degree WSS provides some blocking of all unwanted

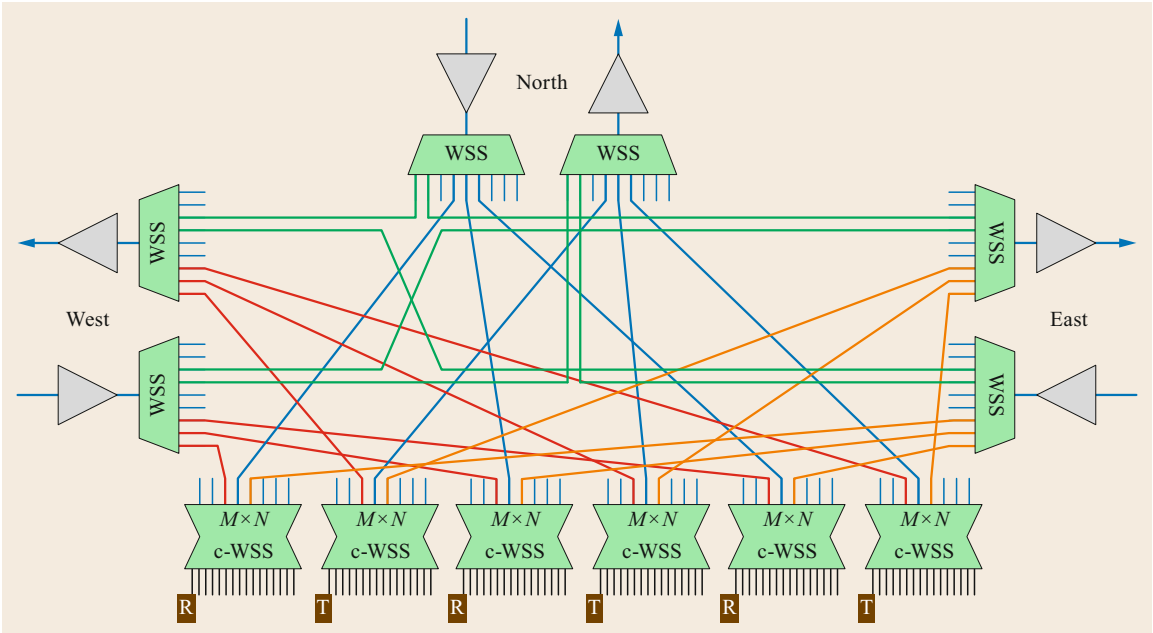


Fig. 8.24 General node architecture for a 3-degree route-and-select node with wavelength-contentioned $M \times N$ WSS-based CD add/drop

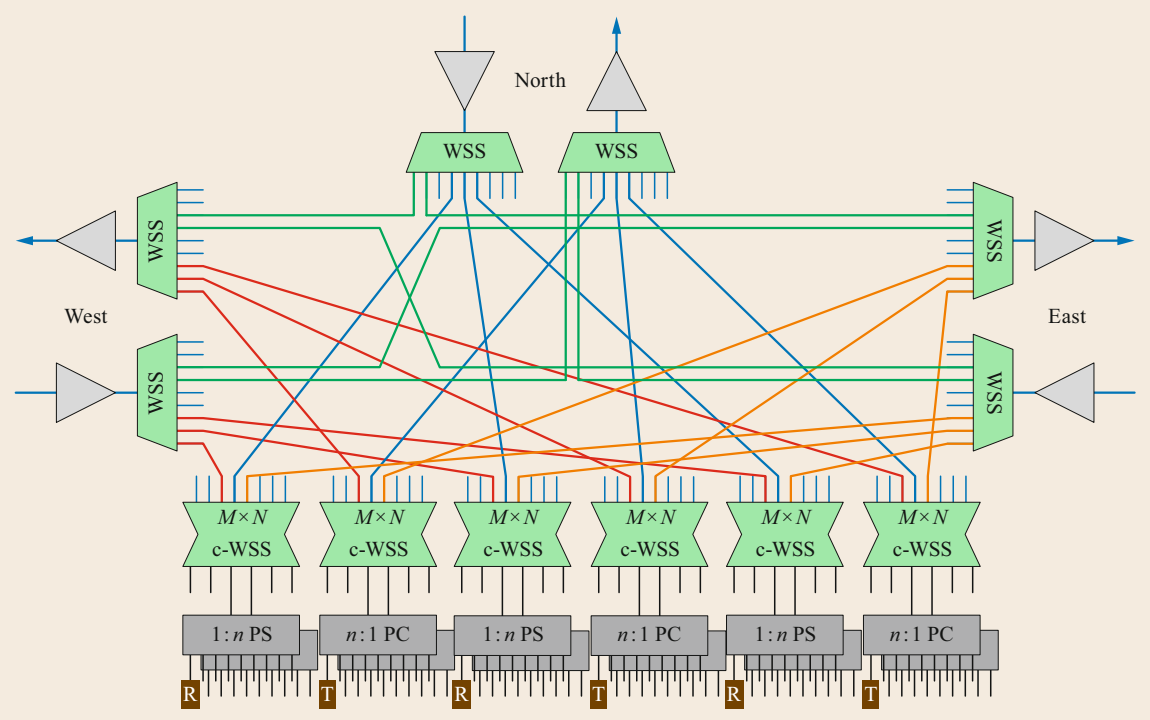


Fig. 8.25 General node architecture for a 3-degree route-and-select node with CD add/drop based upon a contentioned $M \times N$ WSS ($M \times N$ c-WSS) and expansion power splitter (PS) and power coupler (PC) demultiplexing and multiplexing

wavelength signals headed to the CD $M \times N$ WSS and if that mostly blocked, low-power signal were to reach a coherent transceiver it would not interfere with that receiver's ability to receive its desired signal. However, this characteristic does present a performance problem for using the wavelength-contentioned $M \times N$ WSS as a functional replacement for the back-to-back WSS architecture, given the potential for lower power signals to reach a direct detect receiver and have sufficient power level to potentially degrade its performance.

8.9 4- to 16-degree Route-and-select, Colorless, Directionless and Contentionless ROADM Node

The impact of wavelength contention within CD multiplexing/demultiplexing dominantly affects dynamic and automated wavelength reprovisioning and restoration activities and the level of the impact strongly depends upon the overall network architecture and the network utilization percentage [8.22, 23]. As network automation and optimization increases, the interest level is increasing in removing the multiplexing/demultiplexing wavelength-contention limitation simply to gain greater flexibility and utility from operators' ROADM networks. Therefore, networks providing colorless, directionless and contentionless (CDC) multiplexing/demultiplexing are currently increasingly being developed and deployed.

The first generation of CDC multiplexing/demultiplexing implementation utilizes the route-and-select architecture within the node degrees in conjunction with generally multiple subtending CDC multiplexing/demultiplexing structures incorporating a pair of $M \times N$ multicast switches (MCSs) (Fig. 8.26). The basic functionality of an $M \times N$ MCS is shown in Fig. 8.27. Each signal input into each of the M inputs is broadcast to N selection switches, each $M \times 1$ in dimension, such that any signal input into the $M \times N$ WSS can be routed to any of the N outputs without any wavelength contention. However, each of the N outputs will contain the full complement of wavelength signals that was directed to that input from the degree WSS as the MCS

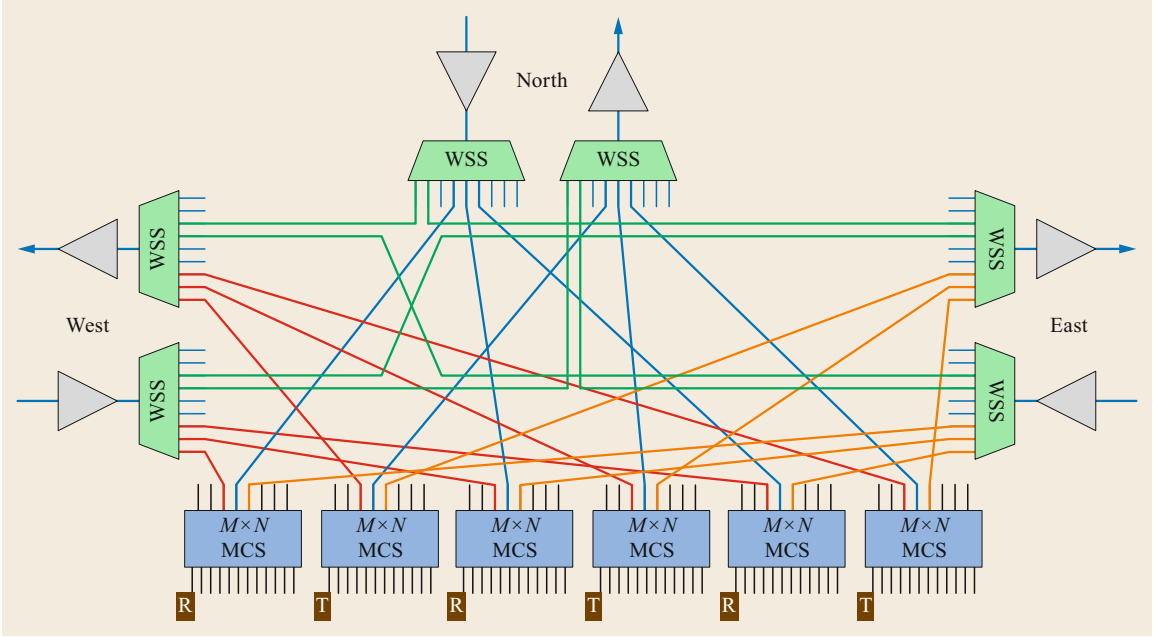


Fig. 8.26 General node architecture for a 3-degree route-and-select node with MCS-based CDC add/drop

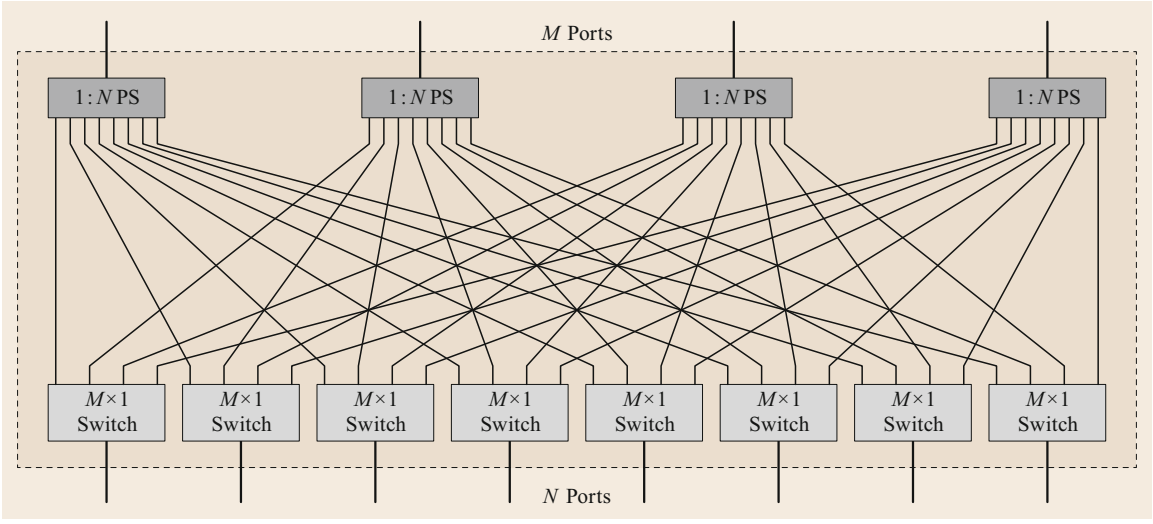


Fig. 8.27 Representation of the principle functionality of an $M \times N$ MCS device (PS = power splitter)

contains no wavelength filtering capabilities. Therefore, if multiple transceivers connected to the MCS receive separate independent wavelength channels from the same degree, all of those wavelength signals from that degree will be present at each respective receiver provisioned to receive one of those channels. The WSS within the degree performs the critical function of limiting the total number of wavelength signals directed to each MCS. Finally, unless individual tunable filters capable of isolating only the desired channel are implemented after the MCS, MCS-based demultiplexing is only compatible with coherent detection.

Multiplexing for CDC is accomplished with the same MCS structure, only with all optical signals propagating in the reverse direction (Fig. 8.26). Each $M \times 1$ switch directs the incoming signal to a respective $N : 1$ power combiner associated with the desired node degree where that signal is combined with all other signals directed to that same degree. As the MCS does not possess any wavelength filtering, all out-of-band noise

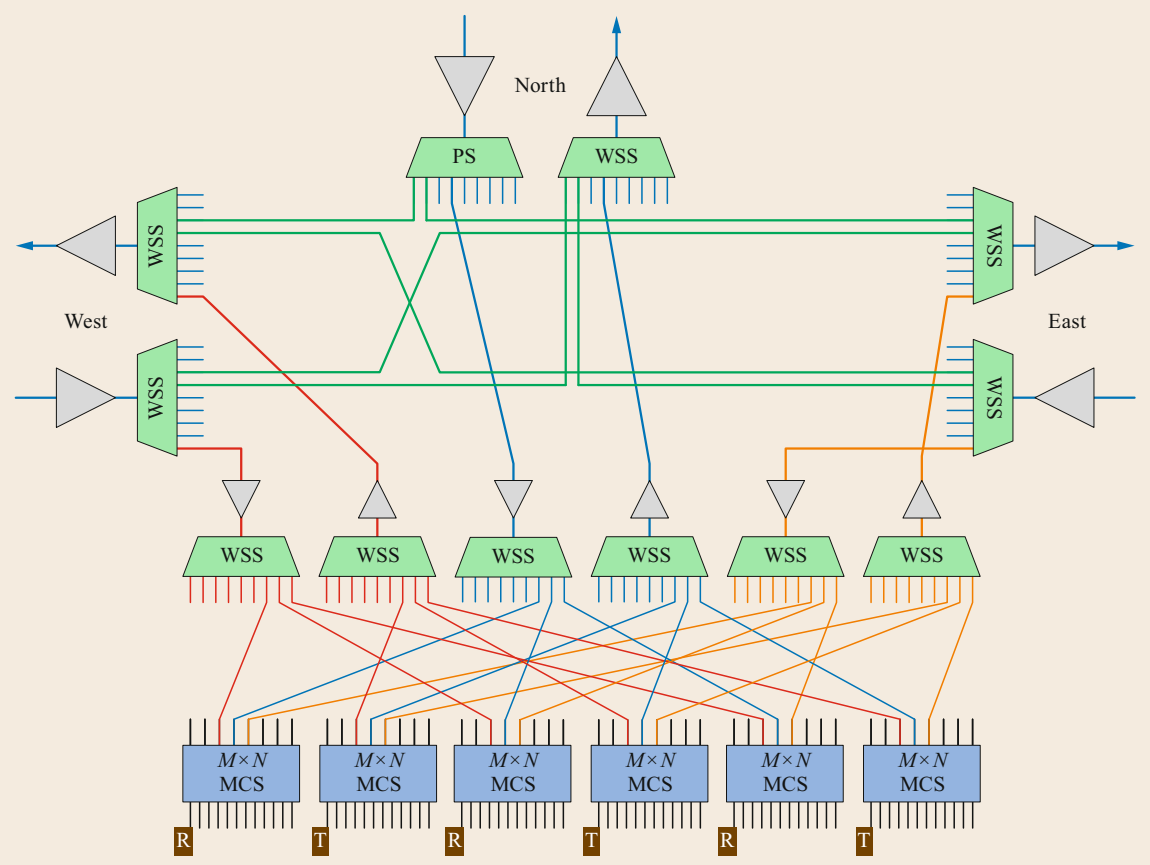


Fig. 8.28 General node architecture for a 3-degree route-and-select node with MCS-based CDC expansion WSS

generated by the transmitters will be combined within the MCS, which places a minimum requirement on the Tx_OSNR specification of the transmitters so that the reach performance of those aggregated signals is not meaningfully degraded. The impact of this noise aggregation from the transmitters as well as from the amplifier array is limited to only those channels from that respective MCS given the degree WSS will block all optical noise not spectrally located within the channel passbands of those active channels.

As the MCS construction is based upon power splitting/combining, the insertion loss through the device fundamentally depends upon N, the number of transceiver ports, with additional loss attributed to the implementation of the selection switch as well as additional device losses. Generally, for MCS devices where N is larger than four to eight, the total combined loss of the degree WSS and the MCS results in insufficient power at the respective receiver, thus requiring the inclusion of optical amplification located between the WSS and the MCS. Given the M input, this requires an M-wide optical amplifier array where each amplifier is capable of supplying enough output power to amplify N wavelength signals that can be distributed in any configuration across the C-band and must minimize gain excursion during any transient power fluctuation events of some or all channels.

In the multiplexing direction, a similar situation triggers the need for the inclusion of an M-wide optical amplifier array between the MCS and the degree WSSs, again with each multichannel amplifier capable of an output power sufficient to amplify N wavelength signals.

For each node, generally a minimum of two CDC structures is implemented to ensure redundancy for critical traffic. In order to increase the total number of wavelengths added/dropped within a node, additional CDC structures must be added, with each CDC structure requiring a connection to a port on every WSS within the node. Currently, MCS dimensions favor N not exceeding 16, given a larger number of ports could increase the MCS insertion loss such that an undesirably large output power is needed within the demultiplexer amplifier array, the maximum number of wavelength signals po-

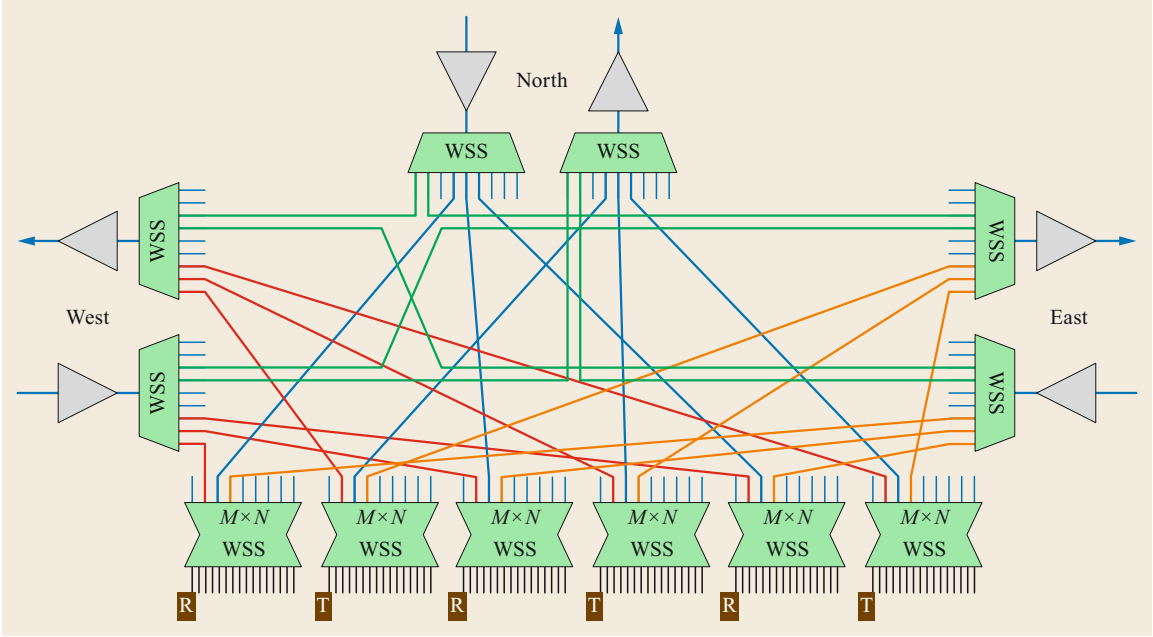


Fig. 8.29 General node architecture for a 3-degree route-and-select node with $M \times N$ WSS-based CDC add/drop

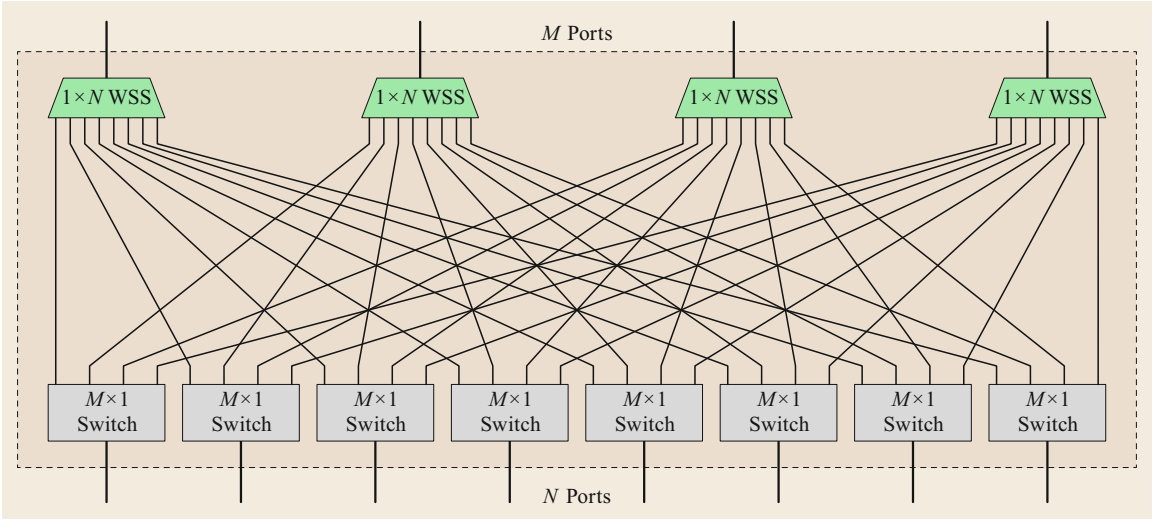


Fig. 8.30 Representation of the principle functionality of a contentionless $M \times N$ WSS device

tentially incident onto a coherent receiver is excessive, and/or the amount of out-of-band transmitter noise accumulated within the MCS is excessive.

To accommodate up to 16 degrees as well as a significant total number of add/drop ports within a degree, node WSSs with port counts exceeding 32 ports are utilized. Additional WSS ports for more CDC structures can be provided by deploying an expansion WSS in the same manner as illustrated in Fig. 8.28.

A subsequent alternative approach to CDC multiplexing/demultiplexing that address a number of the perceived shortcomings of the MCS implementation, particularly for larger nodes, is the wavelength-contentionless $M \times N$ WSS [8.24] (Fig. 8.29). This device has the ability to independently route any wavelength signal between any of the M input ports to any of the N output ports without any dependence upon how any of the other wavelength signals are routed (excluding routing two of the same wavelength signals to the same port) (Fig. 8.30).

As the contentionless $M \times N$ WSS device is based upon wavelength routing and not power splitting, the

insertion loss is not fundamentally dependent upon the port count on either the M or N dimension. Therefore, if the insertion loss of the device can be sufficiently low, no amplification array is necessary within the CDC structure, leading to a simplified CDC structure. With limited insertion loss dependence, both the degree (M) and add/drop port count (N) can be increased without a fundamental increase in insertion loss, allowing the number of degrees and add/drop ports supported by a contention to increase without requiring an amplification array. This scalability enables more ports per CDC structure and thus requires fewer node WSS ports for a given total number of add/drop ports.

The contentionless $M \times N$ WSS possesses flex-spectrum wavelength filtering such that only the wavelength bands routed to a port are allowed to pass with all other portions of the spectrum significantly blocked. This provides advantages in that only the desired channel need be routed to each receiver, which can improve the receiver's ability to receive that channel. Similarly, any out-of-band transmitter noise will be suppressed within the contentionless $M \times N$ WSS such that the Tx_OSNR specifications on the transceivers can be relaxed without performance degradation and independent of the number of add ports (N).

8.10 Reach and Performance Dependencies

Reduction of insertion loss is a primary goal of any optical network design. The attenuation of signals passing through a ROADM node is compensated by amplifiers, which increase signal power at the cost of added noise (Chap. 3). Accumulation of noise throughout the optical link leads to a reduction in optical signal to noise ratio (OSNR) and limits the maximum propagation distance of optical signals. Hence, there is a desire to reduce the loss and the amplification requirements throughout an optical network.

The loss of a ROADM node depends on the node architecture and the optical components that are utilized. The loss of a power splitter and coupler scales with the number of ports, whereas the loss of a WSS generally does not increase linearly with port count. However, a WSS has a higher excess insertion loss than a coupler. Therefore, a 16-degree route-and-select ROADM node employing 20-port WSS modules can possess a higher total loss than 2-degree ROADMs using a simple splitter and a 2×1 WSS architecture. However, that same 20-port route-and-select node can have less total loss than an 8-degree broadcast-and-select node. Additionally, the ROADM performs wavelength channel power equalization by reducing the power on all wavelengths to the lowest channel power within the spectrum. The difference in power between the average channel power and the lowest channel power, and the accuracy in power setting, should also be considered as contributors to the total node loss [8.25].

In general, on either side of the ROADM's broadcast and/or switching elements sits an inbound erbium-doped fiber amplifier (EDFA) (commonly referred to as a preamplifier) and an outbound EDFA (referred to as a postamplifier) (Fig. 8.31). The purpose of the preamplifier is to compensate for the fiber loss of the preceding span, typically with the gain, noise figure, and maximum output power optimized to compensate for high losses and channel power targets suitable for the ROADM's drop structure. The postamplifier compensates for the node loss, including equalization gain margins, with gain, noise figure, and power optimized to achieve channel power targets well suited for fiber propagation with minimal nonlinear impact. For long-haul networks, the preamplifier may be augmented with counterpropagating Raman in order to reduce the effective noise figure of the node (Chap. 3).

In typical metro networks, the intranode loss is usually higher than the fiber span loss because of the limited distances needed to interconnect a metropolitan area. Most of the OSNR degradation is induced by the postamplifier, which in turn is a function of the ROADM loss. Maximum reach is therefore typically more dependent upon the total number of nodes than it is on fiber span lengths. In contrast, in long-haul networks, amplifier-only non-ROADM nodes occur with much greater frequency, and OSNR degradation is dominated by the fiber loss-compensating amplifier-only nodes. Maximum reach is therefore largely related to length of fiber spans and fiber type than it is to the total number of ROADM nodes.

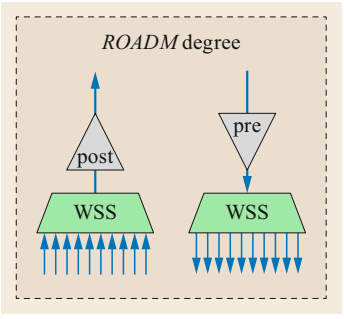


Fig. 8.31 Pre- and postamplifier positioning at a ROADM node

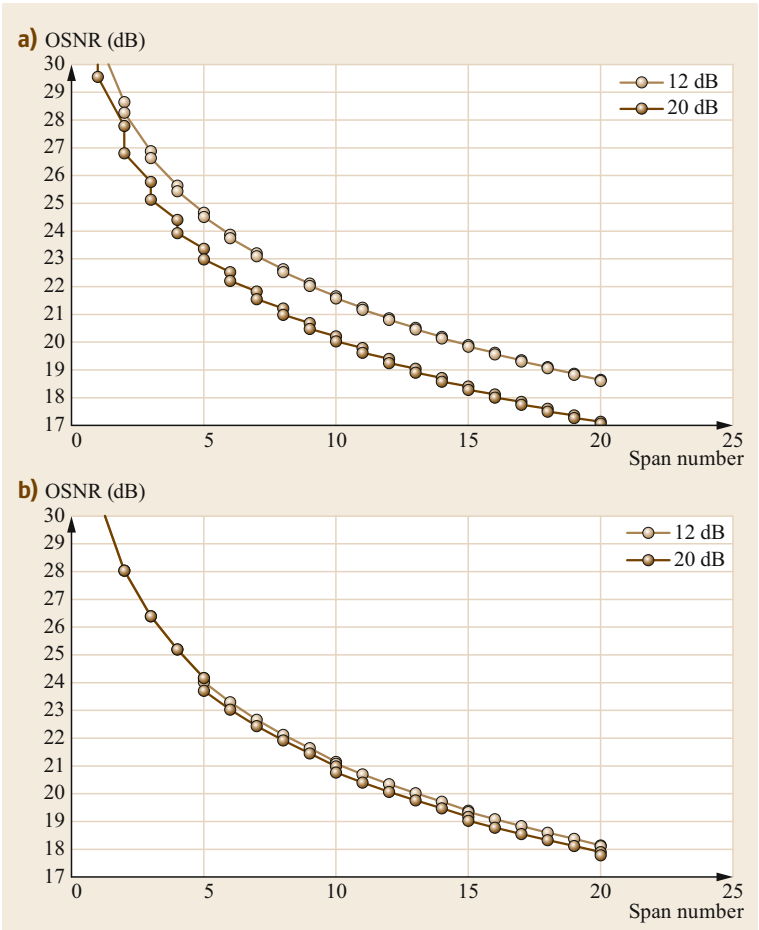


Fig. 8.32a,b Examples of OSNR degradation for **(a)** metro and **(b)** long-haul representative networks

To highlight the impact of ROADM design (particularly insertion loss) on overall system performance, a case study of a 96-channel, highly meshed metro network with ROADMs after every span is compared to one of a long-haul network with fewer, infrequent ROADMs placed primarily for periodic channel power equalization. Each example network is analyzed for two cases of total ROADM loss: 12 and 20 dB. The 12 dB node could correspond to a simple 2-degree ROADM, while the 20 dB could represent a higher-loss broadcast-and-select 8-degree ROADM. Each node is assumed to have an EDFA preamplifier and an EDFA postamplifier to compensate fiber span and node loss, as discussed above. The amplifiers are assumed to have noise figure values that are a function of gain; the preamplifiers optimized for lower noise figures at higher gain values, and higher available gain range, than the postamplifiers. The fiber span loss is assumed to be 16 dB for the metro example and 24 dB for the long haul.

The reduction in OSNR with span number is illustrated in Fig. 8.32a, b for the metro and long-haul cases respectively. As expected, a much larger ROADM loss dependency on the delivered OSNR can be seen in the metro case (a) compared to the long-haul case (b). After 20 nodes, the delivered OSNR for the metro example is about 1.5 dB lower for ROADMs with 20 dB loss versus 12 dB loss (about a 30% reduction). Viewed another way, if an optical transceiver requires an OSNR of 19 dB for error-free operation (e.g., 200 Gb/s 16QAM [8.26]), with 12 dB ROADMs the signal could achieve transmission over 19 spans, but with 20 dB ROADMs, it would be limited to 13 spans. In contrast, because the ROADMs are so infrequent, in the long-haul network the OSNR difference after 20 spans is only about 0.3 dB between the 12 and 20 dB cases. Considering the same 19 dB OSNR example, a reduction in maximum reach of only one span (from 16 to 15 spans) is observed.

8.11 WSS Bandpass Filtering and Implications

Another impairment requiring consideration in ROADM-enabled networks originates from the filtering of the optical signals as they pass through multiple WSS devices. As signals independently propagate through the network, neighboring channels may be added or dropped at nodes along a signal's route to its destination. As these neighboring channels get added or dropped, an optical bandpass filtering effect is applied to the signal of interest. An example of such a filter shape can be seen in Fig. 8.33a, which shows the WSS insertion loss as a function of frequency, assuming neighboring channels have been filtered out on either side of the center channel. As more ROADM nodes are cascaded for a given signal, there is the potential for the signal to experience an increasing amount of total filtering with each successive ROADM node. Figure 8.33b shows the cascaded filter bandwidth as a function of number of WSSs traversed. When considering that a signal passes through two WSSs in a route-and-select ROADM node, the number of potential WSSs traversed can grow quite large. This cascading of optical bandpass filtering can lead to narrowing of the transmitted optical signal – an impairment difficult to equalize within the receiver. Particularly for modern optical networks, which employ high-bitrate optical sources with Nyquist signaling (allowing channel spacing at intervals slightly higher than the baud rate), this is problematic and can lead to high transmission penalties after relatively few devices [8.27, 28].

The implications of this potential distortion source as transceiver technologies move towards higher baud rates and subsequently wider optical signals has motivated ROADM node designers to adopt statistical modeling of bandpass narrowing in engineering specifications and design rules, and WSS designers to develop higher-resolution WSS devices with sharper filter passband shapes.

During earlier ROADM node designs and deployments, state-of-the-art transponder technology generally employed on-off keying at 10 Gb/s, and chan-

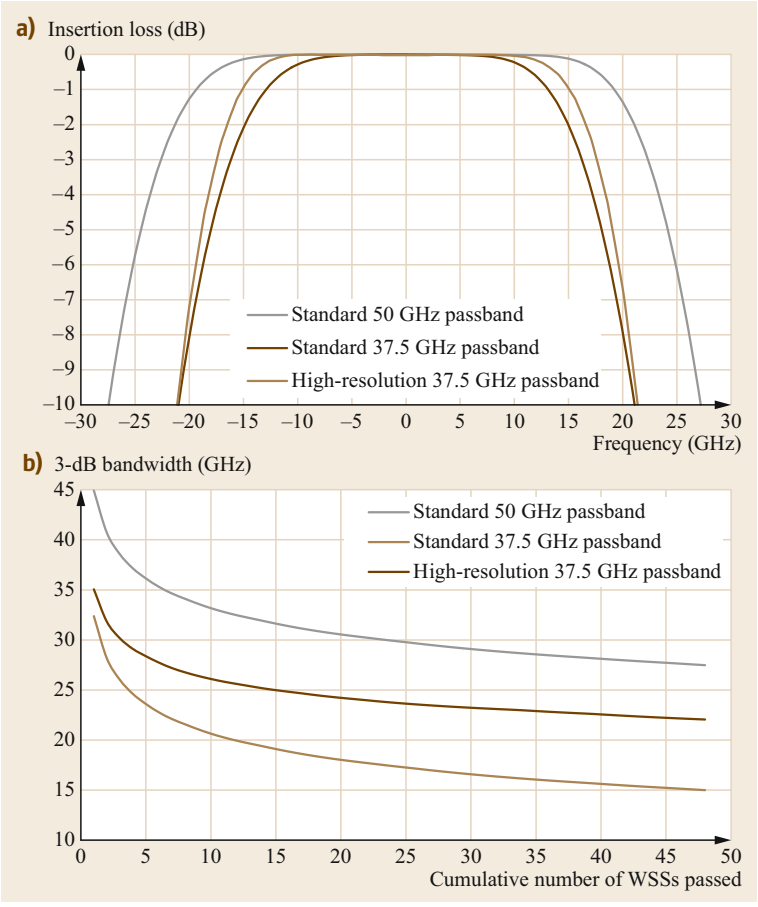


Fig. 8.33 **(a)** Example insertion loss channel passband for 50 GHz channel of a standard quality WSS and channel passbands of a 37.5 GHz channel for a standard (*blue*) and high-resolution (*green*) WSS. **(b)** Cumulative channel passband after a cascade of a given number of respective WSSs

nel spacing was typically 50 or 100 GHz such that the spectral occupancy of the signals was much narrower than the cumulative passband filtering created by the cascaded WSSs. Specifying WSS devices in terms of a simple center frequency and minimum passband width was typical, and penalties incurred due to filtering were negligible. As transceiver signal bandwidths have approached that of the channel spacing, WSS device manufacturers' and optical network system suppliers' handling of specifications and engineering rules evolved accordingly. Rather than using the minimum specifications for the WSS passbands to determine the cumulative WSS passband width, given the large number of WSSs a signal potentially experiences, a statistical approach has widely been adopted to model the cumulative passband width of a cascade of ROADM nodes. It has been found that for WSSs with sharper passband filter shapes, the cumulative bandpass filter width decreases sublinearly and the net bandwidth variation in a ROADM cascade (due to variations among individual WSSs in the cascade) decreases rapidly with increasing number of ROADMs [8.7]. This has allowed ROADM networks designs to accommodate higher baud rate signals without costly hardware, network topology, or reach implications.

The second implication has been the general improvement in WSS device technology providing higher-resolution WSS designs, which enable sharper filter shapes without significant compromises in physical size or cost. The curves in Fig. 8.33 labeled *high-resolution* show the impact of such design improvements.

8.12 Performance Monitoring and Wavelength Connection Validation

In the discussions above, it was assumed that the ability to perform monitoring of the incoming and outgoing optical signals is provided within the ROADM node. Optical signal monitoring is a critical capability as it provides both the information needed for the system to automatically control various automated control features (such as channel power equalization), and provides validation and visibility into the network characteristics at remote locations, allowing technicians such visibility entirely though the management software.

8.12.1 Optical Channel Monitoring

The ability to measure optical power is essential to the ability to perform channel presence detection, per-channel and spectral power equalization, spectral pre-emphasis, early detection of performance degradation, and fault troubleshooting. A ROADM node design very typically implements optical channel monitoring capability using low-percentage optical taps on all node ingress and egress ports providing a measurement of power versus frequency of the optical signals that enter and leave the node (similar to what an externally connected optical spectrum analyzer would provide). Early OCMs provided measurements of channel power levels integrated over the full 50 or 100 GHz channels. However, with the advent of flexible grid systems and Nyquist signaling, OCM requirements have evolved to provide subchannel granularity measurements (e.g., sub-6.25 GHz resolution) and report power spectral density (PSD, typically dBm/Hz). Given the measurements from the OCM are generally the primary information source from which automated functional control is determined, more frequent measurements can accelerate the frequency in which these controls can operate. This can translate into more rapid wavelength provisioning, reprovisioning following a fault, and more effective power transient suppression. OCM scanning speeds on the order of hundreds of milliseconds are typically currently desired for these applications.

8.12.2 Connection Validation and Wavelength Activation

Due to the large number of physical connections that are required between discrete elements or modules within an advanced ROADM node, operators desire the ability to automatically detect and validate that connections have been properly made, ensuring that the ROADM node is fibered properly and with the expected losses. A simple form of validation includes using low-percentage optical taps connected to a photodiode at strategic locations within the node, which can detect a low dither frequency unique to each transceiver. Another approach can include injecting a dedicated test signal at a reserved wavelength and then verifying that signal is detected by filtered tap-photodiodes at the proper locations within the node. A third approach leverages an extra test port on the WSSs to implement an optical loopback function for a prescribed wavelength channel in order to provide transceiver connection validation within various domains of the ROADM node, including within the CDC multiplexing/demul-

tiplexing as well as within the full node degree. This aids in troubleshooting initial deployments and turn-up, in identifying sources of network faults during ongoing operations and allows for self-calibration of optical transponders before turn-up to ensure optimal performance.

8.13 Looking Forward

The chapter above describes node architectures generally aimed at supporting up to 16 degrees, whereas the industry is currently developing designs to support beyond 16 degrees. Similarly, current network implementations are optimized around supporting transceivers that operate a single-wavelength signal within a channel. To keep pace with the needed increase in capacities [8.29], it is expected that the industry will eventually shift to transceivers that utilize multiple wavelength signals tightly spaced together so that they are treated as a single, wide channel [8.30]. While requiring the ability to support flexible spectrum, which most modern ROADM networks already support, these channels will possess higher total power levels due to the presence of multiple wavelengths. This is likely to force some modifications to existing CD and CDC networks, which may be engineered for lower total power levels per channel. This impact may not drive strong modifications to the ROADM node degree architecture, but may drive design choices where amplification is avoided given these higher power levels, such as using the contentionless $M \times N$ WSS for CDC multiplexing/demultiplexing (Fig. 8.29), which does not incorporate amplification in comparison to the MCS-based approach, which does (Fig. 8.26).

References

8.1 International Telecommunication Union: ITU Recommendation, ITU-T G.691, 2.0 edn. (2/2012)

8.2 B.C. Collings, F. Heismann, C. Reimer: Dependence of the transmission impairment on the WSS port isolation spectral profile in 50 GHz ROADM networks with 43 Gb/s NRZ-ADPSK signals. In: *Proc. Opt. Fiber Commun. Conf.* (2009), Paper OThJ3

8.3 O. Solgaard, J. P. Heritage, A. R. Bhattarai: Multiwavelength cross-connect optical switch, US Patent 6097859A (1998)

8.4 T.A. Strasser, J.L. Wagener: Wavelength-selective switches for ROADM applications, IEEE J. Sel. Top. Quantum Electron. **16**(5), 1150–1157 (2010)

8.5 P. Wall, P. Colbourne, C. Reimer, S. McLaughlin: WSS switching engine technologies. In: *Proc. Opt. Fiber Commun. Conf.* (2008), Paper OWC1

8.6 G. Baxter, S. Frisken, D. Abakoumov, H. Zhou, I. Clarke, A. Bartos, S. Poole: Highly programmable wavelength selective switch based on liquid crystal on silicon switching elements. In: *Proc. Opt. Fiber Commun. Conf.* (2006), Paper OTuF2

8.7 F. Heismann: System requirements for WSS filter shape in cascaded ROADM networks. In: *Proc. Opt. Fiber Commun. Conf.* (2010), Paper OThR1

8.8 T. Zami, B. Lavigne, E. Balmefrezol, M. Lefrancois, H. Mardoyan: Comparative study for crosstalk created in 50 GHz-spaced wavelength selective switch for various modulation formats at 43 Gbit/s. In: *Proc. Eur. Conf. Opt. Commun.* (2006), Paper We3.P.81

8.9 M. Filer, S. Tibuleac: Generalized weighted crosstalk for DWDM systems with cascaded wavelength-selective switches, Opt. Express **20**(16), 17620–17631 (2012)

8.10 H. Sun, K.T. Wu, K. Roberts: Real-time measurements of a 40 Gb/s coherent system, Opt. Express **16**(2), 873–879 (2008)

8.11 L. Kazovsky: Multichannel coherent optical communications systems, J. Lightw. Technol. **5**(8), 1095–1102 (1987)

8.12 E.-J. Bachus, R.-P. Braun, C. Caspar, E. Grossmann, H. Foisel, K. Heimes, H. Lamping, B. Strebel, F.-J. Westphal: Ten-channel coherent optical fibre transmission, Electron. Lett. **22**(19), 1002–1003 (1986)

8.13 D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, K. Kikuchi: Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation, J. Lightw. Technol. **24**(1), 12–21 (2006)

8.14 S.J. Savory: Digital filters for coherent optical receivers, Opt. Express **16**(2), 804–817 (2008)

8.15 M.G. Taylor: Phase estimation methods for optical coherent detection using digital signal processing, J. Lightw. Technol. **27**(7), 901–914 (2009)

8.16 L.E. Nelson, S.L. Woodward, S. Foo, M. Moyer, D.J.S. Beckett, M. O'Sullivan, P.D. Magill: Detection of a single 40 Gb/s polarization-multiplexed QPSK channel with a real-time intradyne receiver in the presence of multiple coincident WDM channels, J. Lightw. Technol. **28**(20), 2933–2943 (2010)

8.17 B. Zhang, C. Malouin, T.J. Schmidt: Towards full band colorless reception with coherent balanced receivers, Opt. Express **20**(9), 10339–10352 (2012)

8.18 R. Younce, J. Larikova, Y. Wang: Engineering colorless/directionless/contentionless metro/regional networks for 400G+. In: *Proc. Opt. Fiber Commun. Conf.* (2013), Paper OTh4B.5

8.19 P.D. Colbourne, B.C. Collings: ROADM switching technologies. In: *Proc. Opt. Fiber Commun. Conf.* (2011), Paper OTuD1
8.20 B.C. Collings: Advanced ROADM technologies and architectures. In: *Proc. Opt. Fiber Commun. Conf.* (2015), Paper Tu3D.3
8.21 M. Mathews, S. McGowan, P. Roorda, S. McLaughin, S. Daniels, A. Somani, B. Collings: Demonstration of a novel twin 10x10 WSS for application in a flexible spectrum, colorless and directionless ROADM. In: *Proc. Asia Commun. Photon. Conf.* (2013), Paper AF4E.1
8.22 R. Younce, S. Gringeri, Y. Wang, J. Larikova: Contentionless and near contentionless blocking performance and economics for all coherent metro/regional networks. In: *Proc. Opt. Fiber Commun. Conf.* (2014), Paper W4B.5
8.23 M.D. Feuer, S.L. Woodward, P. Palacharia, X. Wang, I. Kim, D. Bihon: Intra-node contention in dynamic photonic networks, J. Lightw. Technol. **29**(4), 529–535 (2011)
8.24 P. Colbourne: US patent US8045854 B2 (2009)
8.25 S. Tibuleac, M. Filer: Transmission impairments in DWDM networks with reconfigurable optical add-drop multiplexers, J. Lightw. Technol. **28**(4), 557–568 (2010)
8.26 G. Bosco, V. Curri, A. Carena, P. Poggiolini, F. Forghieri: On the performance of Nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers, J. Lightw. Technol. **29**(1), 53–61 (2011)
8.27 Y.-T. Hsueh, A. Stark, C. Liu, T. Detwiler, S. Tibuleac, M. Filer, G.-K. Chang, S.E. Ralph: Passband narrowing and crosstalk impairments in ROADM-enabled 100G DWDM networks, J. Lightw. Technol. **30**(24), 3980–3986 (2012)
8.28 M. Filer, S. Tibuleac: Cascaded ROADM tolerance of mQAM optical signals employing Nyquist shaping. In: *Proc. IEEE Photon. Conf.* (2014) pp. 268–269
8.29 Cisco: Cisco visual networking index: Forecast and methodology, 2016–2021, https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html (2017)
8.30 X. Liu, S. Chandrasekhar: Monolithic coherent receiver based on 120-degree optical hybrids on silicon. In: *Proc. Opt. Fiber Commun. Conf.* (2014), Paper W1H.5

Brandon Collings

Lumentum
San Jose, CA, USA
brandon.collings@lumentum.com

Brandon is the Chief Technology Officer for Lumentum, responsible for defining and driving the product and technology roadmaps and strategies. Prior to JDSU, he held optical network design and development positions at Ciena and Internet Photonics and was a Member of the Technical Staff at Bell Labs, performing research on advanced optical communication network designs and nonlinear performance impairments. He holds a PhD in Electrical Engineering from Princeton University.

Mark Filer

Microsoft Corporation
Redmond, WA, USA
mark.filer@microsoft.com

Mark is currently Optical Network Architect at Microsoft, working on next-generation long-haul and metro DCI optical solutions. Prior to that, he spent over 14 years in optical R&D at ADVA Optical Networking in Atlanta, GA. He has published research in the areas of long-haul transmission, ROADM network architectures, and system impairments focused on nonlinear effects and crosstalk. He holds B.Sc. and M.Sc. degrees in electrical engineering from the Georgia Institute of Technology in Atlanta, GA.

9. Fiber Nonlinearity and Optical System Performance

Alberto Bononi, Ronen Dar, Marco Secondini, Paolo Serena, Pierluigi Poggiolini

This chapter aims to provide a comprehensive picture of the impact of fiber nonlinear effects on modern coherent wavelength division multiplexing (WDM) systems' performance. First, the main nonlinearity models currently available are introduced and discussed in depth. Then, various specific aspects are addressed, such as the interplay of polarization mode dispersion (PMD)/polarization dependent loss (PDL) and nonlinearity, or the dependence of nonlinear effects on modulation format. The important topic of nonlinear effects mitigation is then dealt with. Finally, system performance metrics and capacity are discussed extensively, as to how they are fundamentally influenced and limited by fiber nonlinearity.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_9

Optical propagation within an optical fiber involves complex linear and nonlinear interactions among traveling optical signals, which manifest themselves as intersymbol and interchannel interference among the digital data. To shed light on such effects, a large number of analytical, numerical, and experimental investigations have been carried out. Scientific modeling is an essential tool for any investigation as a way to make it easier to understand and reproduce experimentally observed results. Models help system designers in predicting the behavior of an optical network and show their greatest effectiveness in their ability to isolate subsets of test conditions that are not easily obtainable otherwise.

The main aim of this chapter is to review the most important models for performance prediction in optical fibers, to point out their strengths and weaknesses, and to provide both theoretical background and practical indications on their use.

9.1 Modeling the Optical Fiber

Any model captures part of the reality at a given abstraction level, and hence each model can always be generalized to a more complete model able to explain the previous one, at the cost of increased complexity. Usually, in science the most complete models are local models, i.e., general descriptions of what happens at each space coordinate. Such models are usually written in terms of differential equations. In many cases, such equations cannot be solved exactly, nevertheless they contain all the information about the evolution of the signal of interest. In optical fiber propagation, the reference model is the nonlinear Schrödinger equation (NLSE), which is a partial differential equation that describes the local behavior of a propagating electromagnetic wave [9.1]. In optical communications, an electromagnetic wave is typically written as $A(z,t) \times \exp[\mathrm{j}(\omega t - \beta z)]$. Such notation describes a wave of amplitude $|A|$, phase $\arg[A]$, propagating at frequency $f = \omega/(2\pi)$ at phase velocity ω/β in the z-direction; $A\,(z,t)$ is assumed to be *slowly varying*, in both time t and distance z, with respect to the faster oscillating exponential factor. Henceforth, for compactness, we will omit to explicitly indicate the t and z-dependence of A.

In its simplest form, the NLSE can be written as [9.1]

$$\frac{\partial A}{\partial z} = -\frac{\alpha_\mathrm{p}}{2}A + \mathrm{j}\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - \mathrm{j}\gamma\,|A|^2 A\,. \tag{9.1}$$

The first term on the right-hand side in (9.1) causes the wave to experience a *power loss* at a rate set by α_p. The small value of α_P in silica compared to copper cable has been one of the greatest achievements of optical communications, celebrated by awarding the Nobel prize to the pioneering studies made by Sir Charles Kao.

The second term in (9.1) expresses dispersion in the time domain t. To understand the implications of dispersion we should first generalize the basic idea of the plane wave. The reader familiar with the Fourier transform should not be surprised that the true electromagnetic field propagating within an optical fiber can be described by a superposition of waves. However, such waves travel at different speeds with a group velocity dispersion (GVD) proportional to β_2. This is the implication of the second term on the right-hand side of (9.1), which is usually best understood in the frequency domain as a term in quadrature with the signal and proportional to the square of the frequency. As a result, GVD alone on a global scale manifests as a pure phase shift in the frequency domain. GVD makes the different waves making up the signal reach the receiver at different times, so that their sum is different than what it was at the transmitter. The result is signal distortion, generally referred to as intersymbol interference at the receiver. This problem was eventually solved by means of optical or electronic dispersion compensation.

The third term in (9.1) is a nonlinear term, i.e., a contribution for which the superposition principle no longer holds. It is called the nonlinear Kerr effect [9.1, 2]. As a consequence, because of nonlinearity, the traveling waves interact. We observe that nonlinearity is weighed by a factor $|A|^2$, hence proportional to the *power* of the propagating signal. We thus expect the nonlinear effect to be relevant only at high power. This observation is reminiscent of the idea of the Taylor series in mathematics, where any well-behaved function is linear in a small neighborhood of a given value. In this case, we could say that the fiber is linear close to the zero value of the amplitude A. It is worth noting, though, that from a naive similarity with the Taylor series, one would expect a nonlinear quadratic term, whereas the nonlinear Kerr term in (9.1) is *cubic* in A. The reason for the cubic behavior is strictly related to the odd symmetry of the silica molecule [9.1].

The nonlinear Kerr term in (9.1) is a memoryless effect since it impacts the propagating signal at the present time only. This is clearly unrealistic, since any physical effect has some memory. However, this assumption is justified by the extremely small time lag

over which nonlinearity occurs, of the order of femtoseconds.

There is a very interesting duality between the nonlinear Kerr effect in the time domain and dispersion in the frequency domain. For instance, GVD alone does not make different frequencies interact, as for any linear time-invariant effect, while it creates intersymbol interference in the time domain; the opposite occurs for nonlinearity, where different times do not interact, while different frequencies do. Moreover, while GVD acts in quadrature with the signal in the frequency domain, the same occurs for the nonlinear Kerr effect, but in the time domain. Hence, for similar arguments, both effects manifest as phase rotations: GVD in the frequency domain and the nonlinear Kerr effect in the time domain. The latter is generally called self-phase modulation (SPM).

Modeling nonlinearity has been one of the greatest challenges of the last decades because of the complexity of the problem. Thanks to the similarity between the propagation of optical waves in optical fibers and that of water waves in the sea, it is helpful to visualize nonlinearity by the waves' interaction in a heavy sea. The chaotic behavior of such interaction itself explains the difficulty of undoing it at the receiver, as well as that of the difficulty of a full statistical description.

Linear and nonlinear effects are experienced both by the information signal but also by the unavoidably propagating noise, generally introduced by optical amplifiers in the form of amplified spontaneous emission (ASE).

To understand the implications of ASE let us ignore nonlinear effects for a moment. At the receiver we know how to compensate for dispersion, thus we are left with the problem of noise cumulated along the link. The solution appears simple: increase power to make noise negligible with respect to the information signal. Figure 9.1 sketches this idea by showing the typical behavior of the receiver signal-to-noise ratio (SNR) versus signal power, at the output of a fiber transmission link. The SNR is a typical quality-of-transmission parameter, as we will discuss more in detail later in this chapter. In the absence of nonlinear effects, the SNR would grow indefinitely for increasing power, see the dashed line *linear asymptote* in Fig. 9.1. However, this ceases to hold when nonlinearity enters the game. The typical behavior of SNR versus power is depicted in Fig. 9.1 by a solid line. In the figure, we observe that an *optimal power* exists, corresponding to the maximum achievable SNR. Beyond the optimal power the SNR drops, with a slope which is actually steeper than the slope of increase at small power [9.3, 4].

We thus recognize two asymptotic behaviors at small/high powers, which in agreement with the previous discussion define what we call the ASE-limited region and the nonlinearity-limited region, respectively.

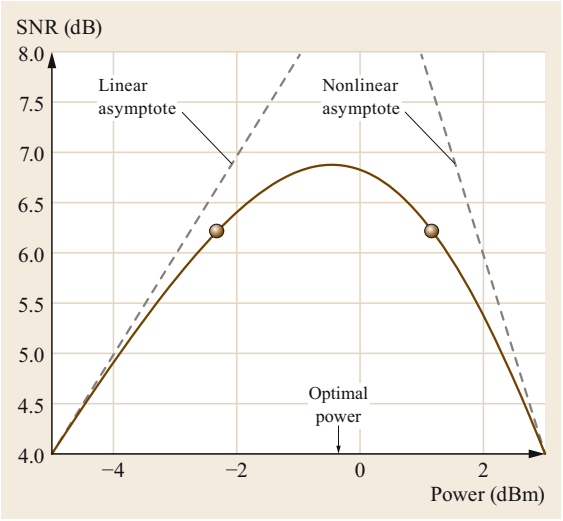


Fig. 9.1 Example of behavior SNR versus power after a fiber-optic link

The optimal power of Fig. 9.1 refers to a link of a given length. We may ask what happens by increasing the link length. In this scenario, both the linear and the nonlinear asymptotes run lower, because of more accumulated ASE and nonlinearity along transmission. If we fix a reference SNR, which may be the threshold of the forward-error correcting code under use, we can identify two power values, one in the ASE-dominated regime and one in the nonlinearity-dominated regime. Such values are sketched in Fig. 9.1 at a fixed distance, and in Fig. 9.2 at all distances up to the maximum distance, or *maximum reach* (MR), at which they merge into a single power value.

In this chapter, we discuss some models that are able to describe such important features as the optimal power and the reach with good approximation. Of course, their knowledge is of great importance in the study, design, and even realtime management of optical networks, a topic that will be covered in Sect. 9.10.3.

The NLSE can be further generalized. So far, we have described the electromagnetic wave in terms of its amplitude, frequency, and propagation constant. However, if this wave collides with an electron along propagation, such an electron starts oscillating around a preferred direction. The information about this direction has not been accounted for, so far. We thus need to accept that our signal wave is described by a *vector*. When such a direction matters for the investigation of propagation, we say that the optical fiber shows *birefringence*. Such vectors can be fully described by two components. The scalar impairments, such as loss and

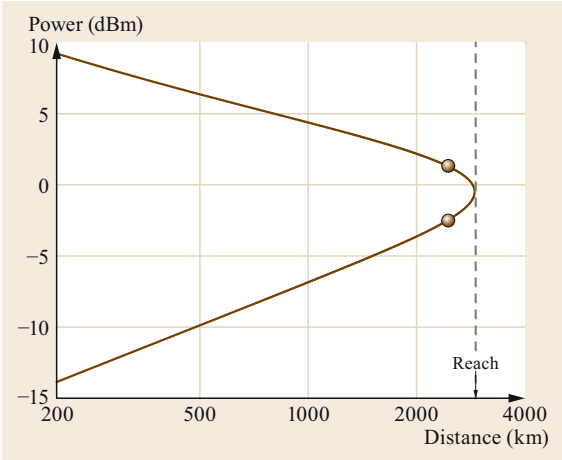


Fig. 9.2 Example of contour levels at fixed SNR. The lower part is set by accumulation of ASE, and the higher part is set by accumulation of nonlinearity. The reach is the maximum distance for the given SNR

dispersion, become matrices in the presence of birefringence.

The implications of the vector nature of the waves will be described more in detail in Sect. 9.9.3. In essence, the nonlinear term of the NLSE becomes more complex [9.2, 5, 6]. However, it is possible to simplify it with reasonable approximations, based on the following argument.

The two components of the wave, from now on referred to as *polarizations*, experience different propagation constants. Such a difference is the main manifestation of fiber birefringence. Because of it, the two polarizations periodically phase shift and return in phase after the so-called *beat length* [9.7]. For typical telecom fibers, the beat length is of the order of meters, a length scale usually much smaller than the length scale over which nonlinearity induces significant changes in the wave. We are, thus, justified to average out such fast phase variations.

The result yields an isotropic nonlinear effect, which turns the scalar nonlinear term $\gamma |A|^2 A$ into $\gamma(8/9)||\boldsymbol{A}||^2\boldsymbol{A}$, where $\boldsymbol{A} = \left[A_x, A_y\right]$ is the vector containing the amplitudes of the two polarizations in the fiber, while $||\boldsymbol{A}||^2$ is the overall power of the wave. The factor $8/9$ is the net result of the averaging. The corresponding equation is referred to as the *Manakov equation* [9.8, 9].

The Manakov equation is currently the most widely used signal-propagation equation, the starting point for the derivation of most analytical models quantifying the system impact of nonlinearity. This chapter will mostly focus on such analytical models. However, another approach is also possible for system study and design.

It is based on the brute-force numerical integration of the signal propagation equation, be it the NLSE, the Manakov equation, or more complex ones. While analytical modeling has recently gained widespread adoption, numerical integration is, and will likely be, a mainstay of system studies. One of the main reasons is that, as mentioned above, all analytical models are based on assumptions and approximations, and must be validated. System simulation, by means of direct numerical integration of the propagation equations and Monte-Carlo system performance estimation, provides an indispensable tool for model validation. In addition, nonconventional system setups may fall outside the scope of models and necessarily need numerical integration. Because of the key role of system simulations, the next section is devoted to this topic.

9.2 Fiber Propagation: Numerical Methods

The NLSE (9.1) introduced in Sect. 9.1 has no general closed-form solution. While some approximated analytical solutions (discussed in Sect. 9.3.1) are available and are widely used for modeling purposes, the accurate simulation of optical fiber systems largely relies on numerical methods.

Over the years, several different numerical methods have been developed to solve the NLSE and its generalizations, based on different approaches for the discretization of the time and space variables, such as classical explicit and implicit finite difference methods and pseudospectral methods (see, for instance, [9.1, 10] and references therein). Among them, the split-step Fourier method (SSFM) [9.11, 12] has largely prevailed due to its simplicity, flexibility (it can be easily extended to include other propagation effects and system elements), numerical stability, and computational efficiency [9.10], becoming the method of reference in the optical fiber community [9.1].

The key idea behind the SSFM is that of dividing the propagation along a link of length L into M steps of length $h = L/M$, over which the effects of the linear and nonlinear parts of the equation are split (hence, the attribute *split step*). Equation (9.1) is rewritten in the simple general form

$$\frac{\partial A}{\partial z} = (\mathcal{L} + \mathcal{N})A \,, \tag{9.2}$$

where the operators $\mathcal{L} = \mathrm{j}(\beta_2/2)(\partial^2/\partial t^2)$ and $\mathcal{N} = -\alpha_\mathrm{p}/2 - \mathrm{j}\gamma\,|A|^2$ represent the linear and nonlinear parts of the equation, respectively [9.13]. (Although fiber attenuation is a linear effect and should naturally be included in the linear operator $\mathcal{L}$, moving it into the nonlinear operator $\mathcal{N}$ makes the resulting SSFM more accurate [9.13].)

The propagation over a step of length h can be formally approximated as

$$A(z+h,t) \cong \exp\left[\int_z^{z+h} \mathcal{N}(\zeta)\mathrm{d}\zeta\right] \exp(h\mathcal{L})A(z,t)\ . \tag{9.3}$$

The approximation (9.3) is particularly convenient, as the separate effects of the linear and nonlinear operators can be easily expressed in closed form, the former in the frequency domain (hence the attribute *Fourier*) and the latter in the time domain.

The algorithm operates on a discrete-time representation of the propagating waveform $A(z,t)$. Assuming a sampling time T_s and a total duration $T = NT_\mathrm{s}$, the signal at distance $z = kh$ (i.e., after k propagation steps) is represented by the vector $\boldsymbol{A}_k = (A_{k,1}, \ldots, A_{k,N})$, with $A_{k,i} = A(kh, iT_\mathrm{s})$. Transformations from time to the frequency domain and back are based on the direct and inverse discrete Fourier transforms, which are efficiently computed by the fast Fourier transform (FFT) algorithm. This entails assuming periodic boundary conditions both in time and frequency, with periods T and $B = 1/T_\mathrm{s}$, respectively. Therefore, when selecting T and B, special care must be taken to ensure that they are sufficiently wider than the actual continuous-time signal duration and (two-sided) bandwidth, respectively, to avoid time and frequency aliasing [9.14]. In the basic implementation of the algorithm, the input vector $\boldsymbol{A}_0$ is propagated through the M fiber sections according to the scheme in Fig. 9.3. In particular, the propagation through the k-th section of the fiber consists of the following four operations:

1. Computation of the FFT of $\boldsymbol{A}_{k-1}$ to obtain the frequency-domain vector $\tilde{\boldsymbol{A}}_k = (\tilde{A}_{k,1}, \ldots, \tilde{A}_{k,N})$.
2. A linear step to compute the vector $\tilde{\boldsymbol{A}}'_k = (\tilde{A}'_{k,1}, \ldots, \tilde{A}'_{k,N})$ according to

$$\tilde{A}'_{k,i} = \tilde{A}_{k,i}\mathrm{e}^{-\mathrm{j}2\pi^2\beta_2 f_i^2 h}, \quad i = 1, \ldots, N\,, \tag{9.4}$$

   where f_i is the frequency of the i-th component
3. Computation of the inverse FFT (IFFT) of $\tilde{\boldsymbol{A}}'_k$ to obtain the time-domain vector $\boldsymbol{A}'_k = (A'_{k,1}, \ldots, A'_{k,N})$.
4. Nonlinear step to compute the vector $\boldsymbol{A}_k = (A_{k,1}, \ldots, A_{k,N})$ according to

$$A_{k,i} = A'_{k,i}\mathrm{e}^{-\alpha_\mathrm{p}h/2}\mathrm{e}^{-\mathrm{j}\gamma h_\mathrm{eff}|A'_{k,i}|^2}, \quad i = 1, \ldots, N\,, \tag{9.5}$$

   where $h_\mathrm{eff} = (1 - \mathrm{e}^{-\alpha_\mathrm{p}h})/\alpha_\mathrm{p}$ is the effective length of the step, which accounts for the dependence of the nonlinear operator on the propagation variable due to the attenuation.

The algorithm complexity is $O(N \log N)$ for FFTs and $O(N)$ for (9.4) and (9.5). The actual number of operations depends on the specific implementation of the algorithm and will be briefly discussed in Sect. 9.11.1. The algorithm can be extended to the Manakov equation for dual-polarization signals by replacing signal samples with two-element vector samples (each element representing a polarization component). All the operations described should then be applied to both polarizations, interpreting the squared modulus $|A'_{k,i}|^2$ in (9.5) as the squared norm $\|A'_{k,i}\|^2$ of the vector sample.

Since the linear and nonlinear operators in (9.2) do not commute, (9.3) is only an approximation. However, an argument based on the Baker–Campbell–Hausdorff formula ensures that the local error induced by (9.3) is of the order of h^2, such that the global error accumulated at the output of the link is of the order of h, i.e., it decreases linearly with the number of steps M. A more accurate approximation can be obtained by using (9.3) for half the step and applying the operators in reverse order for the other half, such that error terms of the order of h^2 cancel out. This results in a symmetric version of the algorithm that is based on the approximation

$$A(z+h) \cong \exp\left(\frac{h}{2}\mathcal{L}\right) \exp\left[\int_z^{z+h} \mathcal{N}(\zeta)\mathrm{d}\zeta\right] \times \exp\left(\frac{h}{2}\mathcal{L}\right) A(z) \tag{9.6}$$

and has local and global errors of the order of h^3 and h^2, respectively. The adjacent linear steps of two consecutive sections can be merged into a single linear step of length h. Therefore, the symmetric algorithm is almost equal to the basic one depicted in Fig. 9.3—but for halving the length of the first linear step and for adding one more linear steps of length $h/2$ after the last section—and achieves a higher accuracy with a negligible increase of complexity.

The accuracy and complexity of the algorithm can be traded off by selecting the step size h, which can

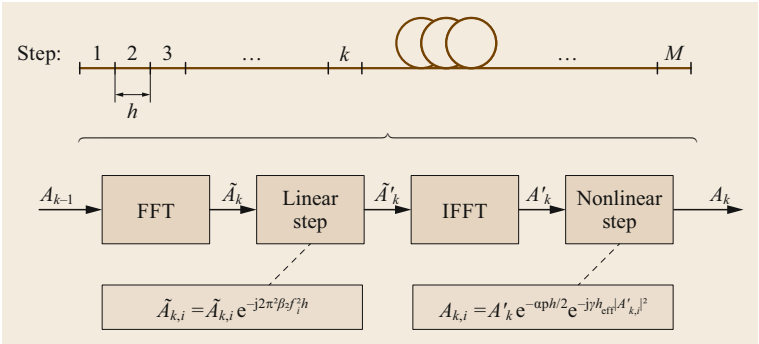


Fig. 9.3 Basic implementation of the SSFM algorithm

also be adaptively changed along the link to increase the accuracy for a fixed complexity. Some possible criteria for the selection of the step size, as well as a strategy for its adaptation based on the local error, are discussed and compared in [9.15]. A simple strategy for the adaptation of the step size based on the global error is presented in [9.16].

When processing a very long sequence of samples, the overlap-and-save technique is typically employed to minimize the complexity of the algorithm and keep its latency within reasonable limits [9.14]. The input sequence of samples is thus divided into several overlapping blocks of length N, which are separately processed according to the algorithm in Fig. 9.3. The output sequence is eventually reconstructed by concatenating the output blocks and discarding the overlapping samples.

The number of overlapping samples should be at least equal to the overall memory N_{m} of the fiber-optic channel, which depends on the maximum dispersion experienced during propagation. In a dispersion-uncompensated link of total length L, assuming a constant dispersion parameter β_2, the channel memory can be approximated as $N_{\text{m}} \cong 2\pi|\beta_2|B^2L$ [9.17].

9.3 Fiber Propagation: Analytical Perturbation Modeling

The absence of closed-form solutions of the NLSE, except for peculiar cases, stimulated the search for approximated analytical solutions. Among the plethora of methods that mathematics can provide, perturbation theory [9.18, 19] has been the most widely used in approximating the NLSE.

9.3.1 Perturbation Methods

The basic idea of perturbation theory is to identify a term in a differential equation that is expected to be a small disturbance to the exact solution of the equation, i.e., it is a perturbation. Most of the time, the small perturbative term is weighed by a parameter, which represents a key variable for perturbation theory. If such a parameter is not clearly visible, one can always multiply the perturbative term by a coefficient ϵ, apply the perturbation theory, and finally set $\epsilon = 1$.

Once the small perturbation and its related parameter ϵ, which is based on physical intuition, are identified, one can use math to properly approximate the solution for small ϵ. The idea is to write the unknown exact solution of the differential equation as a Taylor series in ϵ. In the NLSE case, we identify the nonlinear Kerr effect as the small perturbation and γ as the corresponding small parameter [9.20–22]. We, thus, search the solution of the NLSE as a Taylor series with respect to γ

$$A(z,t) \cong A_0(z,t) + A_1(z,t)\gamma + \dots \,.$$

We do not know $A(z,t)$, but due to perturbation theory we can evaluate $A_n(z,t)$ for all integers $n \geq 1$. They can be evaluated recursively by solving the following differential equations [9.18]

$$\frac{\partial A_n}{\partial z} = \mathcal{L}A_n + \mathcal{N}(A_{n-1}) \,. \tag{9.7}$$

At a closer look we note that such differential equations are linear, with a forcing term $\mathcal{N}(A_{n-1})$. Each of them can, thus, be solved exactly once we know the solution at the previous order.

Although we can truncate the Taylor series up to any order, all A_n with $n \geq 2$ are extremely complex and, with very few exceptions, the NLSE has typically been approximated to first order in γ. Incidentally, the NLSE is a well-behaved differential equation and setting $\gamma =$

0 does not cause divergence problems, so that we say that we have a regular perturbation (RP) solution [9.18]. We are then mainly interested in RP1, i.e., in RP up to γ^1.

The first-order perturbative term in the frequency domain takes the following general expression [9.19, 20, 23, 24]

$$\tilde{A}_1(z,f) = \iint_{-\infty}^{\infty} \eta(f_1,f_2,f) \times \tilde{A}(0,f+f_1)\tilde{A}(0,f+f_2)\tilde{A}^*(0,f+f_1+f_2)\mathrm{d}f_1\mathrm{d}f_2 \,, \tag{9.8}$$

where $\eta(f_1,f_2,f)$ is the fiber kernel and contains all the main properties of the link. From a numerical point of view, the RP solution (9.8) does not yield significant speed-up with respect to the numerical solution provided by SSFM. However, its main interest lies in providing an end-to-end analytical solution that can be statistically manipulated, as we will see in the following sections. Alternatively, the RP1 solution can be expressed as a propagating perturbative term along the distance, the implications of which will be discussed in Sect. 9.6.

For the NLSE, the RP has been showed to coincide with the Volterra series expansion [9.23]. Nevertheless, the physical intuition behind RP is of great help in suggesting ways to improve the solution. First, it has been shown that even in the presence of dispersion part of the perturbation is in quadrature with the unperturbed signal. Therefore, besides observing amplitude distortion at the receiver, we do have phase distortion that we generally categorize as SPM and XPM. Even if SPM and XPM are small phase distortions, they may induce a significant low-frequency contribution that is usually removed by the carrier-phase estimator (CPE) in the coherent detection, or by the photodiode in the direct detection. The basic RP solution (9.8) is *unaware* of such observations, and, thus, its basic expression (9.8) often fails to be satisfactorily accurate. Special tricks to remove such problems are, thus, necessary to improve the range of validity of the RP1 approximation.

A first improvement is to search the perturbation in the reference system of the *average nonlinear phase rotation*, yielding the enhanced RP (eRP) method [9.23]. For first-order eRP, the approximated solution of the NLSE is

$$A(z,t) \cong (A_0 + \gamma A_1)\, \mathrm{e}^{-\mathrm{j}\Phi_{\mathrm{NL}}} \,,$$

where Φ_{NL} is the average nonlinear phase cumulated along the link. For instance, for a single span of effective length L_{eff} and nonlinear length L_{NL}, it is $\Phi_{\mathrm{NL}} = \gamma L_{\mathrm{NL}}/L_{\mathrm{eff}}$. This approach corresponds to perturbing a modified NLSE with average power P removed in the nonlinear term, i.e., with $\mathcal{N}(A) = -\mathrm{j}\gamma(|A|^2 - P)A$ [9.25]. In Sects. 9.4, 9.5, and 9.6, we will calculate and discuss the variance of the perturbation in the eRP reference system.

Another improvement to RP is the logarithmic perturbation (LP), where the perturbation is searched in a logarithmic domain [9.26]. At first order for $A_0 \neq 0$, we have

$$A(z,t) \cong A_0 \mathrm{e}^{\gamma \frac{A_1}{A_0}} \,.$$

LP coincides with the exact solution of NLSE in the limit of zero dispersive effect, where the nonlinear Kerr effect is a pure phase rotation. With GVD the exponent in LP is a complex number, hence it does not manifest itself just as a phase rotation. This class of models will be discussed in Sect. 9.7.

9.4 From FWM to GN and EGN Models

In this section, two of the most well-known and widely used nonlinearity models, the GN and EGN model, will be introduced by deriving them from the basic Kerr-related nonlinear fiber effect called four-wave mixing (FWM). Historically, this is one of the ways in which these models have been derived, and perhaps it is one of the most intuitive. Interestingly, this FWM-based derivation can be shown to be equivalent to the first-order eRP solution described in [9.27] Sect. 9.3.1.

In the end, although some of the details will have to be skipped, the reader should gain a rather detailed appreciation of the derivation procedure, of the inner structure of these models, and a general understanding of how fiber nonlinear effects can be described in terms of the interaction of many discrete signal spectral components with one-another.

The GN model was actually derived several times, in slightly different versions, over the years. The earliest instance dates back to 1993 [9.28], where an FWM-based derivation was used. Though limited to single-polarization, ideal-lossless fiber and a rectangular overall WDM spectrum, the equations in [9.28] essentially coincide with those of the incoherent GN model or the iGN model (Sect. 9.4.7).

Later, in 2003, it was shown in [9.29] that results similar to [9.28] could also be derived using a different perturbation approach, earlier proposed in [9.30]. Similar equations were also independently derived in [9.24] using truncated Volterra series in the frequency domain, as introduced in [9.20].

Subsequently (2008–2010), a derivation approach based again on FWM was used in [9.31, 32] to derive a GN model version limited to optical OFDM. A similar approach was independently exploited in 2010–2011 to address not just OFDM, but all WDM systems [9.33–35]. The derivation yielded the frequency-continuous, dual-polarization integral GN model formula, which is currently considered the reference formula for the model (called GNRF, or GN model reference formula). The name *GN model* was first introduced in [9.34].

Detailed rederivations of the GN model were published in [9.27, 36], based on different methods. Both independently confirmed the GN model equations and provided substantial extensions and insightful generalizations. Various follow-up papers have been published since, providing further extensions, as well as closed-form approximate solutions of the integral GN model formula, for instance, [9.37–41].

More recently, based on a better understanding of the features and limitations of the GN model, an improved model called the *enhanced* GN model (or EGN model) has been proposed. For references and a detailed discussion of the EGN model, see Sect. 9.4.8.

9.4.1 Kerr Nonlinearity and Spectral Line Signals

We consider a signal $E(f)$, made up of Q spectral *lines*, ideally Dirac's deltas, written as

$$E(f) = \sum_{q=1}^{Q} \zeta_q \delta(f - f_q) \,, \tag{9.9}$$

where the ζ_qs are the complex amplitudes of each one of the lines. On such spectrally delta-like signals, fiber Kerr nonlinearity (Sect. 9.1) acts mainly through the FWM effect.

The literature on FWM is vast. A comprehensive treatment can be found in [9.1]. In the next few paragraphs, we are going to recall only what is strictly needed for our purposes.

When $E(f)$ is propagated along a stretch of optical fiber, then *each* set of *three* out of the Q signal lines in (9.9), say those located at frequencies f_k, f_m and f_n, generates a FWM contribution in the form of a *fourth* spectral line at frequency

$$f_{\mathrm{FWM}}^{(i)} = f_k + f_m - f_n \,, \tag{9.10}$$

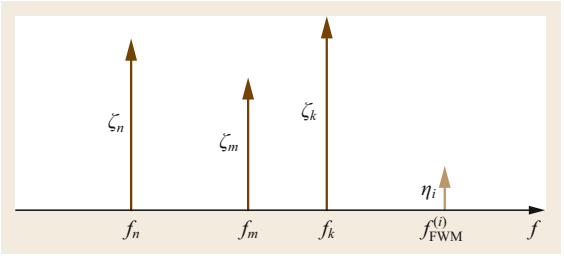


Fig. 9.4 Generation of a FWM spectral line at frequency $f_{\mathrm{FWM}}^{(i)} = (f_m - f_n + f_k)$ due to Kerr nonlinearity acting upon the signal spectral lines at frequencies f_m, f_n, and f_k, called *pumps*

as shown in Fig. 9.4. Note that k, m, n run each independently over $\{1 \dots Q\}$ and, therefore, there are Q^3 possible triples (k, m, n). We use the index i to identify each one of these triples, through the notation $(k, m, n)_i$, where i runs over $1 \dots Q^3$. A given triple $(k, m, n)_i$ may actually generate a FWM line at the same frequency as the FWM line generated by another triple $(k', m', n')_{i'}$. That is, it can be that: $f_{\mathrm{FWM}}^{(i)} = f_{\mathrm{FWM}}^{(i')}$. The reason is that it is possible that $(f_k + f_m - f_n) = (f_{k'} + f_{m'} - f_{n'})$, even though $(k, m, n)_i \neq (k', m', n')_{i'}$.

We now write the FWM generation formulas for a single stretch of fiber. An FWM spectral line generated by the triple $(k, m, n)_i$ at a specific frequency $f_{\mathrm{FWM}}^{(i)}$ can be written as

$$E_i(z, f) = \eta_i(z) \delta(f - f_{\mathrm{FWM}}^{(i)}) \,. \tag{9.11}$$

Its complex amplitude η_i grows along the fiber, depending on several parameters, including the power of the signal spectral lines, their relative spectral position, fiber dispersion, loss, and the strength of the Kerr nonlinearity of that fiber.

Significantly, such a growth law takes on a relatively simple form, provided that the so-called undepleted pump assumption (UPA) can be made. Such terminology stems from the fact that the signal lines in (9.9) are called *pumps* according to jargon inherited from physics. The UPA consists of this. Due to energy conservation, the FWM-generated spectral lines draw their growing power from the *pumps*, whose power is, therefore, *depleted*. However, if the overall power of the generated FWM lines is small versus the pumps' power, then the pumps can be assumed unperturbed, or *undepleted*. It can be shown that the UPA is equivalent to what is called a *first-order regular perturbation* (RP1) approach to finding the solution of the NLSE (Sect. 9.3.1).

We assume that the FWM amplitude value at the start of the fiber, $z = 0$, is equal to zero, so that $\eta_i(0) = 0$. Then, using the single-polarization NLSE, under the

UPA, the FWM line amplitude η_i can be shown to be

$$\eta_i(z) = -\mathrm{j}\zeta_m\zeta_n^*\zeta_k \times \mathrm{e}^{\int_0^z \kappa(f_{\mathrm{FWM}}^{(i)},z')\mathrm{d}z'} \times \int_0^z \gamma(z')\mathrm{e}^{\int_0^{z'}[-\kappa(f_{\mathrm{FWM}}^{(i)},z'')+\kappa(f_m,z'')+\kappa^*(f_n,z'')+\kappa(f_k,z'')]\mathrm{d}z''}\mathrm{d}z' , \tag{9.12}$$

where

$$\kappa(f,z) = -\mathrm{j}\beta(f,z) - \alpha(f,z) + g(f,z) . \tag{9.13}$$

Note that this equation can handle all-order dispersion through $\beta(f,z)$, as well as frequency-dependent distributed gain or loss along the fiber through $g(f,z)$ and $\alpha(f,z)$, respectively, g and α both being positive numbers by definition.

The overall FWM created by the signal of (9.9) at the location z along the fiber is actually given by the superposition of all the individually generated FWM contributions, that is,

$$E_{\mathrm{FWM}}(z,f) = \sum_i \eta_i(z)\delta(f - f_{\mathrm{FWM}}^{(i)}) . \tag{9.14}$$

Focusing on the quite typical case of a *homogenous* stretch of fiber, whereby γ is a constant versus the spatial coordinate z, one can rewrite (9.12) in the compact form

$$\eta_i(z) = -\mathrm{j}\gamma\zeta_k\zeta_m\zeta_n^* H(f_{\mathrm{FWM}}^{(i)},z)L_{\mathrm{eff}}(z,\bar{f}_i)\mu(z,\bar{f}_i) , \tag{9.15}$$

where $\bar{f}_i = (f_k,f_m,f_n)$ is shorthand for the i-th triple of frequencies from (9.4) generating a FWM line at $f_{\mathrm{FWM}}^{(i)}$ as per (9.10), and

$$H(f,z) = \mathrm{e}^{\int_0^z \kappa(f,z')\mathrm{d}z'} , \tag{9.16}$$

is the *linear fiber transfer function* at a generic frequency f, from the fiber start till the length z. In addition,

$$L_{\mathrm{eff}}(z,\bar{f}_i) = \int_0^z \mathrm{e}^{\int_0^{z'}\mathrm{Re}\{-\kappa(f_{\mathrm{FWM}}^{(i)},z'')+\kappa(f_m,z'')+\kappa^*(f_n,z'')+\kappa(f_k,z'')\}\mathrm{d}z''}\mathrm{d}z' , \tag{9.17}$$

$$\mu(z,\bar{f}_i) = L_{\mathrm{eff}}^{-1}(z,\bar{f}_i) \times \int_0^z \mathrm{e}^{\int_0^{z'}[-\kappa(f_{\mathrm{FWM}}^{(i)},z'')+\kappa(f_m,z'')+\kappa^*(f_n,z'')+\kappa(f_k,z'')]\mathrm{d}z''}\mathrm{d}z' , \tag{9.18}$$

where $\mathrm{Re}\{\cdot\}$ means the *real part*.

Equation (9.15) shows that the amplitude of the i-th generated FWM line, η_i, is proportional to the factors L_{eff} and μ. The former is called the *effective length* of the fiber. It is a *real and positive* number, with units of length. It can be thought of as the length of the fiber as *felt* by FWM generation. In a perfectly lossless fiber, it coincides with the actual physical length of the fiber, that is, $L_{\mathrm{eff}} = z$. In a lossy fiber, it is shorter than z. Note that when *both* loss and distributed gain are present, the result of (9.17) can be either shorter or longer than z, depending on their balance.

The quantity μ is the so-called *FWM efficiency*, and it accounts for the critically important effect of dispersion on FWM generation. If there is no dispersion, i.e., if $\beta(f)$ is a constant versus f, then it can be immediately seen from (9.18) that $\mu = 1$. If *any* amount of dispersion is present instead, then $|\mu| < 1$. This shows that dispersion is a mitigator of FWM and, in general, of nonlinearity, an important concept that we will return to later in the chapter.

Equations (9.15)–(9.18) constitute an analytical solution to the NLSE in the FWM framework, potentially providing a powerful tool for the derivation of nonlinearity models. In practice, for this to be the case, two conditions must be met: (i) that the propagation problem is within the validity range of RP1, and (ii) that the signal of interest, which is sent into the fiber, can be described as a set of spectral lines like (9.9), so that the FWM formalism can be applied.

As for (i), this condition is typically well verified in modern fiber telecom systems. We will later actually transition from RP1 to eRP1 (see Sect. 9.3.1 for the definition of eRP), which further extends the range of validity of the solution. Regarding (ii), the question is whether actual data signals just can be put in the form (9.9). In the following, we briefly address this latter aspect.

Actual physical transmitted signals (henceforth *data signals*), which we will denote by $E_{\mathrm{S}}(t)$, have the following features: they are finite-energy, limited-peak-power signals of finite bandwidth $[f_l,f_h]$ and finite duration $[t_l,t_h]$. This is, of course, true also for WDM signals consisting of multiple channels. In the WDM case, $[f_l,f_h]$ would represent the overall WDM comb optical bandwidth. Signals with these features can always be Fourier transformed. Formally, $E_{\mathrm{S}}(f) = \mathrm{F}\{E_{\mathrm{S}}(t)\}$, where $\mathrm{F}\{\cdot\}$ is the Fourier transform. In general, though, $E_{\mathrm{S}}(f)$ *would not be* in the form of (9.9); it could contain some spectral lines but would typically consist mainly of a continuous spectrum. However, given *any* signal with the above features as $E_{\mathrm{S}}(t)$, the *periodic signal* built by simply repeating it at regular inter-

vals T_0

$$E_{\mathrm{S}}^{\mathrm{per}}(t) = \sum_{q=-\infty}^{\infty} E_{\mathrm{S}}(t - qT_0) \tag{9.19}$$

acquires a *pure line spectrum* whose expression is

$$E_{\mathrm{S}}^{\mathrm{per}}(f) = \sum_{q=q_l}^{q_h} \zeta_q \, \delta(f - qf_0) \, , \tag{9.20}$$

where $f_0 = 1/T_0$ and $\zeta_q = f_0 E_{\mathrm{S}}(qf_0)$. The summation index q would ideally run over $[-\infty, \infty]$ but $E_{\mathrm{S}}(f)$ is band limited by assumption to $[f_l, f_h]$. This curtails q within $q_l = \min\{q : qf_0 > f_l\}$ and $q_h = \max\{q : qf_0 < f_h\}$. In conclusion, the signal of (9.20) is of the form (9.9) and, therefore, the FWM formalism can be applied to it.

Concerns may arise whether the periodization performed in (9.19) may lead to undesired consequences or loss of generality. There are subtleties to be heeded, such as signal continuity issues at the edges of each period, but the answer is that no loss of generality is incurred.

Note that (9.20) is actually more constrained than (9.9), because the spectral lines in (9.20) are not arbitrarily located in frequency, but rather lie on a *grid of fixed pitch* f_0. Due to this, without any loss of generality, we can from now on identify the frequencies f_k, f_m, and f_n in (9.10) as kf_0, mf_0, and nf_0, respectively. When this is inserted into (9.10), it can be seen that all generated FWM contributions $f_{\mathrm{FWM}}^{(i)}$ forcedly lie *on the same frequency grid* qf_0 where the periodized signal spectral lines also lie, with q

$$q = (k + m - n) \, . \tag{9.21}$$

For notational compactness, we then define $\{i\}_q$ as the set of all the indices i that identify triples $(k, m, n)_i$ that produce an FWM line *at the same frequency* qf_0, according to (9.10). In formulas

$$\{i\}_q = \{i \, : \, f_{\mathrm{FWM}}^{(i)} = qf_0\} \, . \tag{9.22}$$

This could be rephrased by saying that the index q identifies the class of equivalence $\{i\}_q$ of all possible triples $(k, m, n)_i$ that produce an FWM spectral line at the *same frequency* qf_0. Assuming that we are looking at a single span of fiber of length L_s, then the total signal, including both data signal and FWM, at the end of the span is given by

$$E_{\mathrm{tot}}(f, L_{\mathrm{s}}) = E_{\mathrm{S}}(f, L_{\mathrm{s}}) + E_{\mathrm{FWM}}(f, L_{\mathrm{s}}) \, , \tag{9.23}$$

where

$$E_{\mathrm{S}}(f, L_{\mathrm{s}}) = \sum_{q=q_l}^{q_h} \zeta_q(L_{\mathrm{s}}) \delta(f - qf_0) \, , \tag{9.24}$$

$$E_{\mathrm{FWM}}(f, L_{\mathrm{s}}) = \sum_{q} \sum_{\{i\}_q} \eta_i(L_{\mathrm{s}}) \delta(f - qf_0) \, . \tag{9.25}$$

The factors $\zeta_q(L_{\mathrm{s}})$ are the data signal spectral line amplitudes from (9.20), linearly propagated to the end of the span according to

$$\zeta_q(L_{\mathrm{s}}) = \zeta_q \times H(f_q, L_s) \, , \tag{9.26}$$

and H is the fiber linear transfer function defined in (9.16).

9.4.2 Multispan Links

Extending the previous equations to *multiple spans* requires that the UPA be augmented by two related assumptions.

One is that the total FWM at the end of a multispan link can be expressed as the sum of the FWM generated in each generic n-th span. The other is that the signal that generates FWM in the n-th span consists of just the data signal spectral lines injected at the input of the overall link, then linearly propagated to the input of the n-th span.

These assumptions neglect the fact that, at the input of each successive span, besides the data signal, there is also the FWM produced in the previous spans, as well as ASE noise. This approximation is viable provided that, at any point along the link, the data signal power can be assumed to be much larger than the accumulated FWM and ASE power. In practice, this is typically the case because otherwise, if the data signal power was comparable to either the ASE or FWM power, the data signal would be too degraded for successful reception to take place. This argument fails only for systems that can operate at quite low optical signal-to-noise ratio (OSNR) values. Indicatively, no significant effect of co-propagating ASE or FWM are observed down to about 10 dB OSNR [9.42, 43]. A clear impact can be observed at about 5 dB OSNR where, indicatively, about 10% maximum reach decrease is found between considering co-propagating ASE and FWM versus not considering them. In this section, we neglect this aspect, which is discussed in more detail in Sect. 9.9.1.

Analytically, (9.15) can then be modified as follows, to provide η_i at the end of the link (i.e., at L_{end}), which

we call η_i^{e}

$$\eta_i^{\mathrm{e}} = -\mathrm{j}\zeta_k\zeta_m\zeta_n^* \sum_{n_\mathrm{s}=1}^{N_\mathrm{s}} \mathbf{H}(\bar{f}_i; 1, n_\mathrm{s}-1) \times \gamma(n_\mathrm{s}) L_\mathrm{eff}(\bar{f}_i; n_\mathrm{s}) \mu(\bar{f}_i; n_\mathrm{s}) H(f_\mathrm{FWM}^{(i)}; n_\mathrm{s}, N_\mathrm{s}) \,, \tag{9.27}$$

where the fiber transfer function $H(f_\mathrm{FWM}^{(i)}; n_\mathrm{s}, N_\mathrm{s})$ linearly propagates the i-th FWM line from where it is generated, i.e., the n_s-th span, to the last span N_s; H can be found by cascading its single-span version (9.16) along the link. The factor $\mathbf{H}(\bar{f}_i; 1, n_\mathrm{s}-1)$ is shorthand for the product of the three transfer functions that propagate linearly the three pumps from the first span to the end of the $(n_\mathrm{s}-1)$-th span, that is,

$$\mathbf{H}(\bar{f}_i; 1, n_\mathrm{s}-1) = H(f_k; 1, n_\mathrm{s}-1) H(f_m; 1, n_\mathrm{s}-1) H^*(f_n; 1, n_\mathrm{s}-1) \,, \tag{9.28}$$

The notation $L_\mathrm{eff}(\bar{f}_i; n_\mathrm{s})$ indicates the effective length calculated for the n_s-th span alone, relative to the i-th triple $(k, m, n)_i$. It is found by using (9.17) in the local span length coordinate, which runs from 0 to the span length $L_\mathrm{s}^{(n_\mathrm{s})}$, and setting $z = L_\mathrm{s}^{(n_\mathrm{s})}$. Similarly for $\mu(\bar{f}_i; n_\mathrm{s})$ with (9.18). Finally, the summation in (9.27) sums the overall FWM produced in each span, then linearly propagated to the end of the link.

We could then use (9.22)–(9.25) to express the overall data and FWM signals at the end of the multispan link. We will do this in Sect. 9.4.4 after dealing with dual-polarization propagation.

9.4.3 Dual Polarization

Actual fibers carry two polarizations. Nonlinear propagation in dual polarization (DP) obeys a set of two stochastically-coupled nonlinear differential equations, called dual-polarization NLSE (DP-NLSE) [9.1]. The stochastic nature of the coupling is due to the random birefringence of the fiber. However, as pointed out in Sect. 9.1, in typical telecom fibers, it is possible to average the DP-NLSE over the spatial evolution of random birefringence, obtaining the so-called Manakov equation (ME) [9.8, 44, 45]. The ME provides a very effective means of modeling DP nonlinear propagation and is very widely used in analytical modeling and simulations.

To address DP-FWM, we first redefine the *pump* amplitudes ζ_q in (9.9) as DP quantities

$$\zeta_q = \zeta_{\hat{x},q}\hat{x} + \zeta_{\hat{y},q}\hat{y} \,, \tag{9.29}$$

where $\zeta_{\hat{x},q}, \zeta_{\hat{y},q}$ are the $\hat{x}$ and $\hat{y}$ pump components, respectively. Then, according to the ME, quite remarkably all of the previously shown formulas remain *unchanged*, with the exception of the substitution of the scalar quantity $\zeta_k\zeta_m\zeta_n^*$, in (9.27), with the Jones vector (in the $\hat{x}$ and $\hat{y}$ basis) appearing in square brackets below

$$\eta_i^{\mathrm{e}} = -\mathrm{j}\frac{8}{9} \begin{bmatrix} (\zeta_k^{\hat{x}}\zeta_m^{\hat{x}}\zeta_n^{\hat{x}*} + \zeta_k^{\hat{x}}\zeta_m^{\hat{y}}\zeta_n^{\hat{y}*}) \\ (\zeta_k^{\hat{y}}\zeta_m^{\hat{y}}\zeta_n^{\hat{y}*} + \zeta_k^{\hat{y}}\zeta_m^{\hat{x}}\zeta_n^{\hat{x}*}) \end{bmatrix} \sum_{n_\mathrm{s}=1}^{N_\mathrm{s}} L_\mathrm{eff}(\bar{f}_i; n_\mathrm{s}) \times \gamma(n_\mathrm{s}) \mu(\bar{f}_i; n_\mathrm{s}) \mathbf{H}(\bar{f}_i; 1, n_\mathrm{s}-1) H(f_\mathrm{FWM}^{(i)}; n_\mathrm{s}, N_\mathrm{s}) \,. \tag{9.30}$$

Note, in particular, that L_eff, μ, $\mathbf{H}$, and H are *scalar quantities* and are unaffected by DP. Of course, as a result of the above substitution, the FWM amplitudes η_i^{e} acquire a DP nature too, that is, $\eta_i^{\mathrm{e}} = \eta_{\hat{x},i}^{\mathrm{e}}\hat{x} + \eta_{\hat{y},i}^{\mathrm{e}}\hat{y}$.

9.4.4 The Statistical Modeling Approach

The shown analytical FWM formalism lends itself to the derivation of *statistical models*, which are potentially more powerful design/analysis tools than brute-force numerical NLSE or ME integration and Monte-Carlo performance estimation.

We focus on the *end of the link*, identified as L_end. In agreement with system-related literature, we call the Kerr-induced nonlinear disturbance at this location *nonlinear interference*, or NLI, replacing the *FWM* acronym, that is,

$$E_\mathrm{NLI}(f) = E_\mathrm{FWM}(f, L_\mathrm{end}) \,. \tag{9.31}$$

The formula for E_NLI at L_end is then, similarly to (9.25),

$$E_\mathrm{NLI}(f) = \sum_q \sum_{\{i\}_q} \eta_i^{\mathrm{e}} \delta(f - qf_0) \,. \tag{9.32}$$

In essence, the *statistical approach* (SA) consists of viewing the η_i^{e} as random variables (RVs), so that the overall NLI disturbance field (9.32) becomes a random process (RPR), whose statistical features can be studied to assess its impact on data signal detection. Note that the η_i^{e} are *not* RVs if transmission of a *specific* fully-defined data signal takes place. They become RVs if transmission of *random* data signals is assumed. Then, by statistical manipulation, one can extract the average features of NLI over *all possible* data signals, which is the great advantage of the SA over any Monte-Carlo strategy.

In this section, we focus on the GN/EGN class of models. These models aim at finding the *power spectral density* (PSD) of $E_\mathrm{NLI}(f)$, which we will call $G_\mathrm{NLI}(f)$

(typically in [W/Hz]). The reason for the focus on this quantity is that the impact of NLI noise on system performance, typically expressed in terms of a degraded estimated OSNR, can be found based on $G_{\mathrm{NLI}}(f)$. Other indicators can then be estimated based on such OSNR, such as BER, MI, and GMI, to different accuracy.

For convenience, we first define a new RV that is the sum of all the FWM amplitudes *contributing to the same* NLI *line at frequency* (qf_0)

$$\nu_q^{\mathrm{e}} = \sum_{\{i\}_q} \eta_i^{\mathrm{e}} \,. \tag{9.33}$$

Given (9.32) and (9.33), $G_{\mathrm{NLI}}(f)$ is then directly found as

$$G_{\mathrm{NLI}}(f) = \sum_q \mathrm{E}\{|\,\nu_q^{\mathrm{e}}\,|^2\}\delta(f-qf_0) \,. \tag{9.34}$$

To carry out the calculation of $\mathrm{E}\{|\,\nu_q^{\mathrm{e}}\,|^2\}$, with reference to (9.30) and (9.33), we define the *data signal factor*

$$\mathbf{S}_i = \begin{bmatrix} (\zeta_k^{\hat{x}}\zeta_m^{\hat{x}}\zeta_n^{\hat{x}*} + \zeta_k^{\hat{x}}\zeta_m^{\hat{y}}\zeta_n^{\hat{y}*}) \\ (\zeta_k^{\hat{y}}\zeta_m^{\hat{y}}\zeta_n^{\hat{x}*} + \zeta_k^{\hat{y}}\zeta_m^{\hat{x}}\zeta_n^{\hat{x}*}) \end{bmatrix} \tag{9.35}$$

and the *link factor*

$$\begin{aligned} \mathrm{LK}_i = -\mathrm{j}\frac{8}{9}\sum_{n_{\mathrm{s}}=1}^{N_{\mathrm{s}}} & \gamma(n_{\mathrm{s}})L_{\mathrm{eff}}(\bar{f}_i;n_{\mathrm{s}})\mu(\bar{f}_i;n_{\mathrm{s}}) \\ & \times \mathbf{H}(\bar{f}_i;1,n_{\mathrm{s}}-1)H(f_{\mathrm{FWM}}^{(i)};n_{\mathrm{s}},N_{\mathrm{s}}) \,. \end{aligned} \tag{9.36}$$

This way, we can compactly write $\nu_q^{\mathrm{e}} = \sum_{\{i\}_q} \mathrm{S}_i\,\mathrm{LK}_i$. We then have

$$\begin{aligned} \mathrm{E}\{|\nu_q^{\mathrm{e}}|^2\} &= \mathrm{E}\left\{\left(\sum_{\{i\}_q}\mathrm{S}_i\,\mathrm{LK}_i\right)^{\mathsf{T}*}\sum_{\{i'\}_q}\mathrm{S}_{i'}\mathrm{LK}_{i'}\right\} \\ &= \sum_{\{i\}_q}\sum_{\{i'\}_q}\mathrm{E}\{\mathrm{S}_i^{\mathsf{T}*}\mathrm{S}_{i'}\}\mathrm{LK}_i^*\mathrm{LK}_{i'} \,, \end{aligned} \tag{9.37}$$

where the superscript T means *transpose*. Note that the link factor, apart from normalizations, is the same quantity appearing as η in (9.8), which provides a clue that the FWM formalism with the UPA is equivalent to the RP1 approach (Sect. 9.3.1).

Equation (9.37) is at the core of both the GN and EGN models. It shows that the SA reduces to the evaluation of the expectation $\mathrm{E}\{\mathrm{S}_i^{\mathsf{T}*}\mathrm{S}_{i'}\}$, since $\mathrm{LK}_i\mathrm{LK}_{i'}^*$ is completely deterministic and is even available in closed-form in various relevant cases. Note also from (9.35) that S_i *depends only on the data signal at the input of the fiber* and not on the link features, which are instead confined within LK_i.

Despite its apparent compactness, (9.37) does not typically generate simple results. The reason is that the RVs ζ_q, which are the amplitudes of the overall WDM data signal spectral lines in (9.20), interact nonlinearly in (9.37) giving rise to complex sixth-order moments, which generate a large number of terms. Only under drastic simplifying assumptions or special circumstances (9.37) and, as a result, the NLI spectrum (9.34), take on simple forms. This is the case of the GN model, which we introduce next.

9.4.5 The GN Model

The complexity hidden within (9.37) simplifies drastically provided that certain approximations are made. Such approximations give rise to the so-called GN model. They are as follows.

It is assumed that the ζ_qs in (9.9) are *statistically independent* (SI) *zero-mean complex circular jointly-Gaussian* RV. As for their variances $\sigma_{\zeta_q}^2$, they directly relate to the average PSD of the periodized data signal of (9.20), which is given by

$$G_{\mathrm{S}}^{\mathrm{per}}(f) = \sum_{q=q_l}^{q_h} \sigma_{\zeta_q}^2\,\delta(f-q\,f_0) \,. \tag{9.38}$$

Then, a logical choice is to assume that the $\sigma_{\zeta_q}^2$ are such that $G_{\mathrm{S}}^{\mathrm{per}}(f)$ tends to follow the average PSD of the *aperiodic* data signal, $G_{\mathrm{S}}(f)$. This is obtained by imposing

$$\sigma_{\zeta_q}^2 = f_0 \times G_{\mathrm{S}}(q\,f_0) \,, \tag{9.39}$$

where the multiplying constant f_0 adjusts dimensions and also ensures signal power consistency. This way, $G_{\mathrm{S}}^{\mathrm{per}}(f)$ and $G_{\mathrm{S}}(f)$ relate as in the pictorial example of Fig. 9.5.

A key point is whether the main assumptions made so far are reasonable and accurate. As has already been stated, the signal periodization of (9.19) is of no concern. In addition, in Sect. 9.4.5, we will actually show that it can be removed. The PSD matching of (9.39), depicted in Fig. 9.5, is not problematic either. The zero-mean complex circular Gaussian distribution assumption for each one of the ζ_q, *individually*, is also actually quite accurate. What, in general, is *not* accurate, is the assumption that the ζ_q are SI. In fact, assuming conventional memoryless modulation, with independent symbols and standard PM-QAM constellations, the ζ_q *are not independent* within each single WDM channel. Also, their dependence is not straightforward, since,

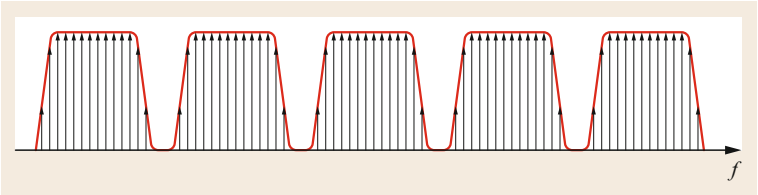


Fig. 9.5 *Solid red line*: an example of the PSD $G_{\mathrm{S}}(f)$ of a five-channel aperiodic WDM transmission signal. *Black arrows*: Dirac's deltas making up the PSD $G_{\mathrm{S}}^{\mathrm{per}}(f)$ of the periodized transmission signal (9.38), according to (9.39)

even though they are individually Gaussian, they are not jointly-Gaussian.

The GN model *neglects* the statistical dependence of the ζ_q within each WDM channel, and this approximation can be expected to generate some error, which we will discuss later on. On the other hand, it allows for extremely simple handling of the average $\mathrm{E}\{\mathrm{S}_i^{\mathsf{T}*}\mathrm{S}_{i'}\}$. The very compact result, which can be obtained through Isserlis's theorem, is

$$\begin{aligned}\mathrm{E}\{\mathrm{S}_i^{\mathsf{T}*}\mathrm{S}_{i'}\} &= \frac{3}{4}\sigma_{\zeta_k}^2\sigma_{\zeta_m}^2\sigma_{\zeta_n}^2\delta_{ii'} \\ &= \frac{3}{4}f_0^3 G_{\mathrm{S}}(kf_0)G_{\mathrm{S}}(mf_0)G_{\mathrm{S}}(nf_0)\delta_{ii'}\ . \end{aligned} \tag{9.40}$$

It should be mentioned that (9.40) is valid when the three indices k, m, n of the i-th triple are *all different*. Using it when this is not the case introduces an approximation, which is, however, of vanishing impact for decreasing f_0. We consequently assume that a small-enough f_0 is chosen. Interestingly, it can be shown that by introducing such approximation, and assuming a small-enough f_0 to make it accurate, the result tends to the better eRP perturbation approach rather than RP. See Sect. 9.3.1 for details on eRP versus RP.

Once (9.40) is plugged into (9.37), Kronecker's delta $\delta_{ii'}$ reduces the double summation to a single one. As a result, the power carried by the NLI spectral line at frequency (qf_0) is given by

$$\begin{aligned}\mathrm{E}\{|\nu_q^{\mathrm{e}}|^2\} = &\frac{16}{27}f_0^3 \\ &\times \sum_{\{i\}_q} G_{\mathrm{S}}(kf_0)G_{\mathrm{S}}(mf_0)G_{\mathrm{S}}(nf_0)\,|\mathrm{LK}_i|^2\ . \end{aligned} \tag{9.41}$$

We must now make explicit the implicit indexing system used in (9.41), defined in (9.22), and write equivalently but *explicitly*

$$\begin{aligned}\mathrm{E}\{|\nu_q^{\mathrm{e}}|^2\} = &\frac{16}{27}f_0^3 \sum_{k=q_l}^{q_h}\sum_{m=q_l}^{q_h} \left|\mathrm{LK}_{km(k+m-q)}\right|^2 \\ &\times G_{\mathrm{S}}(kf_0)G_{\mathrm{S}}(mf_0)G_{\mathrm{S}}([k+m-q]f_0)\ , \end{aligned} \tag{9.42}$$

where the relation $n = (k+m-q)$, derived from (9.21), was used. The link factor LK_{kmn} is now written as

$$\begin{aligned}\mathrm{LK}_{kmn} = -\mathrm{j}\sum_{n_{\mathrm{s}}=1}^{N_{\mathrm{s}}} &\gamma(n_{\mathrm{s}})L_{\mathrm{eff}}(kf_0, mf_0, nf_0; n_{\mathrm{s}}) \\ &\times \mu(kf_0, mf_0, nf_0; n_{\mathrm{s}})H(qf_0; n_{\mathrm{s}}, N_{\mathrm{s}}) \\ &\times \mathbf{H}(kf_0, mf_0, nf_0; 1, n_{\mathrm{s}}-1)\ . \end{aligned} \tag{9.43}$$

Equations (9.42) and (9.43), together with (9.34), provide *a general frequency-discrete* GN *model equation set*. These formulas could be directly used to calculate the PSD of NLI by executing the indicated summations numerically. However, a *frequency-continuous* set of formulas can be derived from them.

Continuous Frequency

It is possible to recast the GN model in integral form by going from discrete to continuous frequency. This is also useful because in certain cases it allows analytical closed-form solutions to be obtained.

The procedure consists of lengthening the periodicity T_0 of the data signal in (9.20). In the limit of $T_0 \to \infty$, that is, of the data signal effectively becoming aperiodic and infinitely extended in time, then also $f_0 \to 0$, and through rather straightforward mathematical manipulation, the *frequency continuous* GN *model* is found. The PSD of NLI is given by

$$\begin{aligned}G_{\mathrm{NLI}}(f) = \frac{16}{27}\int_{f_l}^{f_h}\int_{f_l}^{f_h} &G_{\mathrm{S}}(f_1)G_{\mathrm{S}}(f_2)G_{\mathrm{S}}(f_1+f_2-f) \\ &\times |\mathrm{LK}(f_1,f_2,f_1+f_2-f)|^2\,\mathrm{d}f_1\mathrm{d}f_2\ , \end{aligned} \tag{9.44}$$

$$\begin{aligned}\mathrm{LK}(f_1,f_2,f_3) = -\mathrm{j}\sum_{n_{\mathrm{s}}=1}^{N_{\mathrm{s}}} &\gamma(n_{\mathrm{s}})L_{\mathrm{eff}}[(f_1,f_2,f_3); n_{\mathrm{s}}] \\ &\times \mu[(f_1,f_2,f_3); n_{\mathrm{s}}] \\ &\times H(f_1+f_2-f_3; n_{\mathrm{s}}, N_{\mathrm{s}}) \\ &\times \mathbf{H}[(f_1,f_2,f_3); 1, n_{\mathrm{s}}-1]\ . \end{aligned} \tag{9.45}$$

For the readers' convenience, below we replace the shorthand forms appearing in (9.45) with their definitions, and include certain elements, such as lumped gain/loss and possible dispersion-compensating units (DCUs) that have not been explicitly introduced so far.

The result is

$$\begin{aligned}\mathrm{LK}(f_1,f_2,f_3) &= -\mathrm{j}\sum_{n_s=1}^{N_s}\gamma_{n_s}\\ &\times\int_0^{L_s^{(n_s)}} \mathrm{e}^{\int_0^{z'}[\kappa_{n_s}(f_1,z'')+\kappa_{n_s}^*(f_3,z'')+\kappa_{n_s}(f_2,z'')-\kappa_{n_s}(f_1+f_2-f_3,z'')]\mathrm{d}z''}\mathrm{d}z'\\ &\times\prod_{p=n_s}^{N_s}\Gamma_p^{1/2}(f_1+f_2-f_3)\mathrm{e}^{\int_0^{L_s^{(p)}}\kappa_p(f_1+f_2-f_3,z)\mathrm{d}z}\\ &\quad\times\mathrm{e}^{-\mathrm{j}\beta_{\mathrm{DCU}}^{(p)}(f_1+f_2-f_3)}\\ &\times\prod_{p=1}^{n_s-1}[\Gamma_p(f_1)\Gamma_p(f_2)\Gamma_p(f_3)]^{\frac{1}{2}}\\ &\quad\times\mathrm{e}^{\int_0^{L_s^{(p)}}[\kappa_p(f_1,z)+\kappa_p(f_2,z)+\kappa_p^*(f_3,z)]\mathrm{d}z}\\ &\quad\times\mathrm{e}^{-\mathrm{j}[\beta_{\mathrm{DCU}}^{(p)}(f_1)+\beta_{\mathrm{DCU}}^{(p)}(f_2)-\beta_{\mathrm{DCU}}^{(p)}(f_3)]}\,, \end{aligned}\tag{9.46}$$

where Γ_p is lumped power gain/loss and $\beta_{\mathrm{DCU}}^{(p)}(f)$ is the effect of a DCU element, both located at the end of the p-th span. The quantity κ_p is the complex propagation constant, defined in (9.13), for the p-th span. Equation (9.46) is quite general, as it can handle both distributed and lumped loss/gain, also frequency dependent, as well as the possible presence of DCUs.

9.4.6 Noteworthy Link Factors

In this section, we focus on closed forms or simplified forms for the LK factor, which can be found for various cases of interest. LK is a function of the frequencies of the three *pumps* generating a FWM contribution. This dependence can be cast in the discrete framework of (9.42)–(9.43) through the indices k,m,n or in the continuous framework of (9.44)–(9.45) using frequencies f_1,f_2,f_3. We use the latter, but it is possible to convert from one to the other through the direct substitutions: $f_1=kf_0, f_2=mf_0, f_3=nf_0$.

Transparent and Uniform Link with Lumped Amplification

A *transparent span* is such that its optical gain equals its loss exactly, from the input of that span to the input of the next span. A *transparent link* is one where every span is transparent. A link is *uniform if all spans are identical*.

Under the assumptions of a transparent and uniform link, (9.45) takes on the closed form

$$\begin{aligned}\mathrm{LK}(f_1,f_2,f_3) &= -\mathrm{j}N_s\gamma L_{\mathrm{eff}}(f_1,f_2,f_3)\mu(f_1,f_2,f_3)\\ &\times\mathrm{e}^{-\mathrm{j}N_sL_s\beta(f_1+f_2-f_3)}\mathrm{e}^{\mathrm{j}\frac{\beta_{123}}{2}(N_s-1)L_s}D\left(L_s\frac{\beta_{123}}{2},N_s\right),\end{aligned}\tag{9.47}$$

where the dependence of L_{eff} and μ on n_s was removed, since they are identical for all spans, and we use the shorthand

$$\beta_{123}=\beta(f_1+f_2-f_3)-\beta(f_1)-\beta(f_2)+\beta(f_3)\,;\tag{9.48}$$

$D(x,N)=[\sin(Nx/2)]/[N\sin(x/2)]$ is the periodic-sinc Dirichlet function of order N [9.46].

If we further assume that the fiber loss coefficient α is frequency independent and that lumped (nondistributed) frequency-independent amplification is used, with the amplifier placed at the end of each span, then L_{eff} and μ simplify to

$$L_{\mathrm{eff}}=\frac{1-\mathrm{e}^{-2\alpha L_s}}{2\alpha}\,,\tag{9.49}$$

$$\mu(f_1,f_2,f_3)=\frac{1}{L_{\mathrm{eff}}}\times\frac{1-\mathrm{e}^{-2\alpha L_s}\mathrm{e}^{\mathrm{j}\beta_{123}L_s}}{2\alpha-\mathrm{j}\beta_{123}}\,.\tag{9.50}$$

Note that quite often β is accounted for through the so-called β_2 and β_3 coefficients, as

$$\beta(f)=2\pi^2\beta_2f^2+\frac{4}{3}\pi^3\beta_3f^3\,.\tag{9.51}$$

If so, (9.48) is rewritten as

$$\beta_{123}=4\pi^2(f_1-f_3)(f_2-f_3)[\beta_2+\pi\beta_3(f_1+f_2)]\,.\tag{9.52}$$

The Uniform Lossless Link

Lossless links are often used in research papers as a limiting case of ideal distributed amplification. A uniform lossless link with no DCUs can actually be viewed as a single span, whose length is that of the whole link. Using (9.17), (9.18), with $\alpha=0$, $g=0$ and integration length $z=L_{\mathrm{tot}}$, where L_{tot} is the whole link length, we have that $L_{\mathrm{eff}}=L_{\mathrm{tot}}$ and

$$\mu(f_1,f_2,f_3)=\mathrm{e}^{\mathrm{j}\beta_{123}\frac{L_{\mathrm{tot}}}{2}}\mathrm{Sinc}\left(\frac{\beta_{123}L_{\mathrm{tot}}}{2}\right),\tag{9.53}$$

where $\mathrm{Sinc}(x)=\sin(x)/x$. Using these results, the link factor becomes

$$\mathrm{LK}(f_1,f_2,f_3)=-\mathrm{j}\gamma L_{\mathrm{tot}}\mu(f_1,f_2,f_3)\mathrm{e}^{-\mathrm{j}\beta(f_1+f_2-f_3)L_{\mathrm{tot}}}\,.\tag{9.54}$$

If β is expressed as (9.51), then (9.52) allows us to modify (9.53)–(9.54) accordingly.

Note that there are substantial differences between the link factor of a uniform and transparent lossy multispan link (9.47) and a lossless one (9.54). Indeed, nonlinearity presents strongly different features in the two setups, so a lossless link should never be used as an approximation of a lossy one in practical system studies.

The Uniform Transparent Link with Undepleted Backward-Pumped Raman Amplification

Raman amplification has become a key staple of modern coherent terrestrial links, where it can provide equivalent noise-figure *gains* on the order of 4–6 dB. Equation (9.45) is capable of handling any distributed gain profile, and, therefore, Raman amplification too, but it would typically require numerical integration of (9.17) and (9.18), and, in general, the evaluation of LK would be quite involved.

A closed-form expression for LK is available provided that the following simplifying assumptions and approximations are made: the link is uniform and transparent; gain is due to a single counter-propagating Raman pump; the Raman pump is *undepleted*; Raman gain and fiber loss are frequency independent. Note that the Raman gain need not balance exactly fiber loss; however, we assume transparency, so either lumped loss or lumped gain are assumed to be present at the end of each span to bring the total span gain/loss to balance, from which $|H(f;n_s,N_s)|=1$, for any n_s.

The distributed (field) gain due to a single undepleted backward-propagating Raman pump at a location z within a span whose local spatial coordinate is $z\in[0,L_s]$, is

$$g(z)=C_R P_{R_0}\frac{e^{2\alpha_R z}}{2}\,, \tag{9.55}$$

where C_R is the Raman gain coefficient $1/(\mathrm{W\,km})$, α_R $(1/\mathrm{km})$ is the (field) attenuation of the fiber at the Raman pump frequency, and P_{R_0} is the power of the pump at $z=0$, i.e., at the start of the span. Note that, since the pump is injected at the end of the span and propagates backwards, its power profile actually grows exponentially in the forward direction, as shown by (9.55). The calculations are fully reported in [9.37], and the relevant results are

$$L_{\mathrm{eff}}=\frac{e^{-\frac{C_R P_{R_0}}{2\alpha_R}}}{2\alpha_R}\times\left(-\frac{C_R P_{R_0}}{2\alpha_R}\right)^{\frac{\alpha}{\alpha_R}}\times\left[\Gamma\left(-\frac{\alpha}{\alpha_R},-\frac{C_R P_{R_0}}{2\alpha_R}\right)-\Gamma\left(-\frac{\alpha}{\alpha_R},-\frac{C_R P_R}{2\alpha_R}\right)\right], \tag{9.56}$$

$$\mu(f_1,f_2,f_3)=\frac{e^{-\frac{C_R P_{R_0}}{2\alpha_R}}}{2\alpha_R L_{\mathrm{eff}}}\left(-\frac{C_R P_{R_0}}{2\alpha_R}\right)^{\frac{2\alpha-j\beta_{123}}{2\alpha_R}}\times\left[\Gamma\left(-\frac{2\alpha-j\beta_{123}}{2\alpha_R},-\frac{C_R P_{R_0}}{2\alpha_R}\right)-\Gamma\left(-\frac{2\alpha-j\beta_{123}}{2\alpha_R},-\frac{C_R P_R}{2\alpha_R}\right)\right], \tag{9.57}$$

where $P_R=P_{R_0}e^{2\alpha_R L_s}$ is the power of the pump as it is injected at the end of the span, and

$$\Gamma(x,y)=\int_y^\infty w^{x-1}e^{-w}\,dw$$

is the *upper incomplete* Gamma function. Here too, if β is expressed as (9.51), then (9.52) allows us to modify (9.57) accordingly. The overall expression of LK is still (9.47).

9.4.7 The Incoherent GN Model

Among the assumptions leading to the GN model, there is the premise that the NLI, generated by FWM in each span, adds up at the *field* level at the end of the link, that is,

$$E_{\mathrm{NLI}}^{\mathrm{end}}(f)=\sum_{n_s=1}^{N_s}E_{\mathrm{NLI}}^{(n_s)}(f)\,. \tag{9.58}$$

Equations (9.44) and (9.45) were derived accordingly. A possible alternative assumption is that NLI adds up in *power* at the end of the link, that is,

$$G_{\mathrm{NLI}}^{\mathrm{end}}(f)=\sum_{n_s=1}^{N_s}G_{\mathrm{NLI}}^{(n_s)}(f)\,. \tag{9.59}$$

This assumption is called the *incoherent accumulation approximation* or IAA. The GN model version that uses it is the *incoherent* GN *model*, or iGN model. The iGN model equations are (9.44) and the following (replacing (9.45))

$$|\mathrm{LK}(f_1,f_2,f_3)|^2=\sum_{n_s=1}^{N_s}|\gamma(n_s)L_{\mathrm{eff}}((f_1,f_2,f_3);n_s)\mu((f_1,f_2,f_3);n_s)|^2\times|H(f_1+f_2-f_3;n_s,N_s)\mathbf{H}((f_1,f_2,f_3);1,n_s-1)|^2\,. \tag{9.60}$$

The IAA is a rather drastic approximation. On the other hand, (9.59) brings about substantial simplification, which is greatly advantageous in the practical use of the GN model. There are three main aspects in which it helps.

First, if the GN model is evaluated by numerical integration of (9.44) and (9.45), the $|\mathrm{LK}|^2$ factor typically turns out to be very rapidly oscillating versus the integration variables, requiring a fine integration grid or

sophisticated integration strategies [9.37]. This can be appreciated by looking, for instance, at a closed-form LK, such as for the transparent and uniform link, where the Dirichlet function in (9.47) has a rapidly oscillatory behavior. Such behavior is due to the phase interference of the NLI contributions from different spans, occurring at the end of the link (see Fig. 21 in [9.37]). If the IAA is made, then this interference does not occur and, in the example in question, the Dirichlet function factor disappears altogether.

Secondly, the simplified analytical form of the LK factor makes it easier to find fully closed-form solutions for the overall NLI PSD $G_{\mathrm{NLI}}(f)$, see Sect. 9.4.10.

Thirdly, (9.59) makes it easy to assess NLI within complex reconfigurable networks, because the computation of NLI within each span, as well as the optimization of certain aspects such as launch power, become dependent only on span-local features. NLI is then easily propagated from that span on, by accounting only for loss and gain along the path. This was recognized early on and has developed, among others, into the LOGO network optimization and control strategy [9.47–49], which is currently in use in physical-layer-aware optical networking products of some vendors.

Of course, the above advantages would be moot if the IAA induced too large errors. However, despite the drastic nature of the IAA, the results of the iGN versus the GN model are typically quite close. In fact, in many cases, the iGN model appears more accurate than the GN model, due to the fortuitous circumstance that the error due to the Gaussian data signal assumption made by the GN model partially cancels out the error induced by the IAA. This aspect was studied in detail in [9.41] (Sects. III-D and V-D) and in [9.50]. It is also dealt with in the next section (Sect. 9.4.7).

GN Model Accuracy and Validity Range

For theoretical reasons,the discussion of which can be found in the rich literature on the topic, such as [9.25, 51–55], the GN model approximations tend to work well and produce accurate results when:

(1) The symbol rate is large
(2) Fiber dispersion is large
(3) Loss per span is not too small
(4) Amplification is prevalently lumped
(5) The number of traversed spans is sufficiently large.

In agreement with the above, the GN model has shown excellent predictive capability in ultra-long-haul (ULH) submarine transmission over PSCF at large symbol rates. Various level of inaccuracy have instead been found, for instance, over very low dispersion fibers, at low symbol rates, with short reach (1–3 spans). Note, though, that the model is rather robust in the sense that, typically, several of the above conditions must be simultaneously or severely violated in order to see large errors in NLI prediction by the model. Also, the GN model error is *always* towards underestimating system performance, i.e., it is always conservative. Such underestimation error is typically in the range −5 to −20% of maximum reach.

Several dedicated GN model test and validation studies have been carried out over the years, both simulative and experimental. Examples of the experimental ones are [9.49, 56–62], where the GN model consistently proved quite accurate. Two of these experiments were specifically designed to test the model over different fiber types [9.56, 60]. In particular, the latter addressed seven quite different fiber types. A massive ULH experiment [9.57] showed a very good accuracy of GN model predictions for a 106-channel system at 30 GBaud channel rate, with PM-16QAM transmission, over 10 290 km, with and without ideal DBP.

Regarding simulative tests, a great many papers have been published. A recent broad study [9.50] addressed a very wide range of scenarios, including five QAM formats (from PM-QPSK to PM-64QAM) at 32 GBaud, three fibers (NZ-DSF, SMF, PSCF), three channel spacings (33.6, 27.5 and 50 GHz), and two span lengths (60 and 100 km). Mixed Raman/EDFA amplification was also addressed. The range of maximum reach (MR) probed spanned from 200 to 10 000 km. Criteria (1)–(5) above were confirmed, and a gradual error increase was found when such criteria were departed from. Along similar lines, two newer studies [9.63, 64] have explored the realm of high symbol rates (from 64 to 512 GBaud), showing quite compellingly that, as symbol rates go up, the GN model becomes more accurate. We will come back to this aspect after introducing the EGN model in the next section.

Here, in Fig. 9.6a, we provide an original set of results where we focused on 15-channel systems operating at the next-generation industry-standard symbol-rate of 64 GBaud. The test was run on either single-mode-fiber (SMF) or a challenging nonzero dispersion-shifted-fiber (NZDSF) with high nonlinearity coefficient and very low dispersion. Six QAM formats were tested, from PM-8QAM to PM-256QAM, with three different spacings. The detailed system data are reported in the figure caption. MR at optimum launch power is shown, found based on reaching a target minimum generalized mutual information (GMI) equal to 87% of each format entropy, corresponding to about 15% FEC coding overhead. For instance, the entropy of PM-32QAM is 10 bit/symbol, and the target GMI was set to $0.87 \times 10 = 8.7$ bit/symbol. The abscissa is the net system

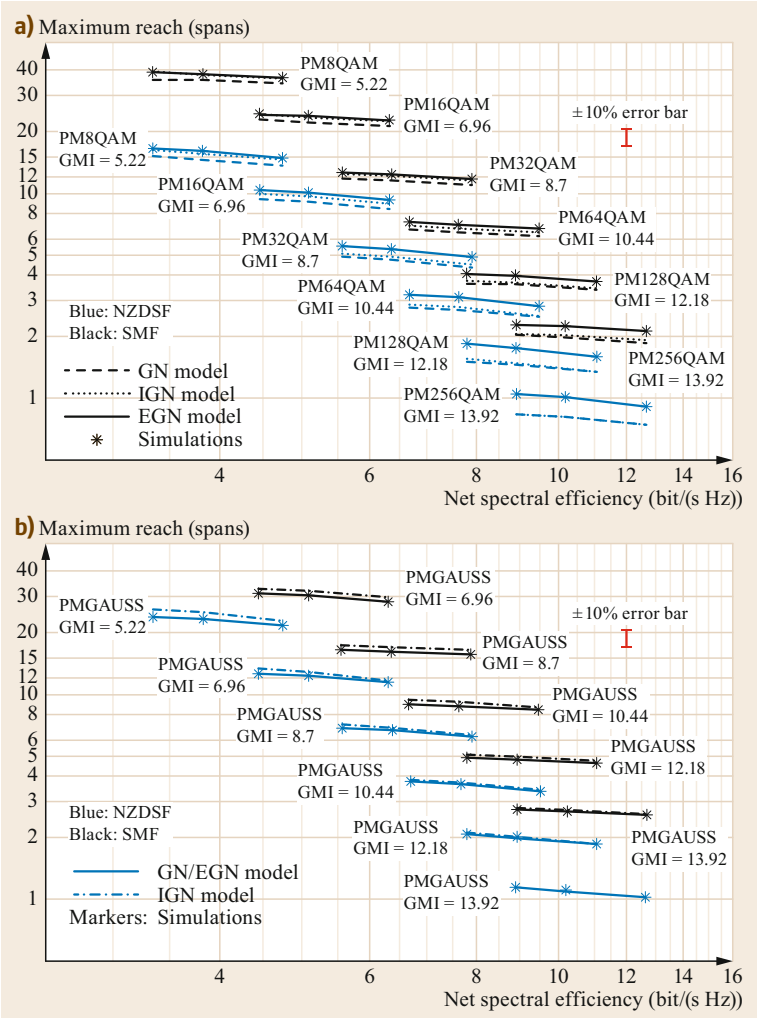


Fig. 9.6a,b System maximum reach (MR) in number of spans, versus net spectral efficiency. *Asterisks* are simulation results, *lines* are analytical predictions using the GN model, incoherent GN model, and EGN model (see legends). Results are shown in *blue* for NZDSF and in *black* for SMF. System data: the span length is 100 km, EDFA noise figure 6 dB, symbol rate $R_s = 64$ GBaud, 15 channels, the tested channel is the center one. For each format the target GMI (for PM-QAM) or MI (for PM-Gaussian) is shown. The abscissa was calculated as $\mathrm{GMI} \times R_s/\Delta f$, MI replacing GMI for PM-Gaussian. Three channel spacings Δf were used for each format: 76.2, 87.5, and 100 GHz. The error bars indicate ±10% error interval. Fibers were: SMF, $D = 16.7$ ps/(nm km), $\alpha = 0.21$ dB/km, $\gamma = 1.3$ 1/(W km); NZDSF, $D = 2.0$ ps/(nm km), $\alpha = 0.22$ dB/km, $\gamma = 1.77$ 1/(W km)

spectral efficiency in bit/(s Hz), calculated as shown in the caption.

Figure 9.6a confirms that the GN model always underestimates MR, in the SMF case by 7% for PM-8QAM, up to 17% for PM-256QAM. This larger error is due to the very short reach of PM-256QAM (1 span), severely violating condition (5). Errors are somewhat larger in the case of the NZDSF (which violates condition (2)), again especially at short reach (violating (2) and (5) together). The incoherent GN model is better overall, for the reasons explained at the end of Sect. 9.4.7.

In summary, the GN model and its incoherent version provide a low-complexity, conservative tool for the estimation of system performance across a very wide range of system scenarios. As we will later show, it also lends itself to the derivation of fully-closed form approximate solutions. These aspects justify its success and widespread use. Later, we will also show that the possible adoption of Gaussian-shaped constellation for optical transmission may actually provide a renewed boost to the use of the GN model.

9.4.8 The EGN Model

As mentioned above, the GN model loses accuracy when criteria (1)–(5) are violated. Also, because of the Gaussian data signal assumption, it does not account for certain nonlinearity features, such as the dependence of NLI generation on the modulation format. These problems were recognized early on [9.51–53], and in [9.53] a procedure was proposed by which the GN model could be modified to avoid them. In [9.65], the so-called *EGN model* was fully derived following up on the procedure introduced in [9.53]. Extended calculation details are reported in [9.66].

Both the GN and the EGN model descend from (9.37). They differ as to the *result of the expectation* $\mathrm{E}\{\mathrm{S}_i^{\mathrm{T}*}\mathrm{S}_{i'}\}$, where S_i is given by (9.35). This is because the RV's ζ_n, which make up S_i, are assumed SI for the GN model, whereas in the EGN model their statistical dependence is taken into account. As a consequence, the sixth-order moments of the ζ_n, which appear within $\mathrm{E}\{\mathrm{S}_i^{\mathrm{T}*}\mathrm{S}_{i'}\}$, yield a much more complex result than (9.40). When such a result is put into (9.37), and the transition from discrete frequency to continuous frequency is made (Sect. 9.4.5), not just one, as in the GN model case, but a total of *12* integral terms are generated, each with different integrand functions and different integration domains.

Interestingly, one of these integral terms *corresponds to the GN model itself*, so that it is possible to write the EGN model as

$$G_{\mathrm{NLI}}^{\mathrm{EGN}}(f) = G_{\mathrm{NLI}}^{\mathrm{GN}}(f) - G_{\mathrm{NLI}}^{\mathrm{corr}}(f)\,, \tag{9.61}$$

where $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$ is the GN model term and $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$ collects the other 11 integral terms. It can be thought of as a *correction* that originates from the statistical dependence of the ζ_n, not taken into account in $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$. Incidentally, such statistical dependence changes depending on the modulation format, so while $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$ is format independent, $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$ is not. Note also that $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$ is intentionally presented with a minus sign in (9.61) to stress the fact that the EGN correction typically *decreases* NLI. In fact, it *always* decreases NLI if PM-QAM signals are assumed, and this explains the feature observed of $G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$ always overestimating NLI.

A complete version of the quite complex EGN model formulas is reported in [9.65, 67]. We will not reproduce them here for lack of space. We will focus instead on an approximate *reduced* EGN model in Sect. 9.4.8, *The Reduced EGN Model*.

EGN Model Accuracy

The EGN model accuracy has been extensively investigated simulatively (see for instance [9.50, 63, 64]). The results have consistently shown that the accuracy of the EGN model is excellent, even when one or more of the conditions (1)–(5), which are critical for the GN model (see the beginning of Sect. 9.4.7), are not met. As we did for the GN model, in Fig. 9.6a we show a collection of tests on QAM systems operating at 64 GBaud, over SMF and NZDSF, with 100 km spans (see Sect. 9.4.7 and the caption of Fig. 9.6a for setup details). The very remarkable feature of the plot is the flawless coincidence of the EGN model MR prediction with the simulation results, across all system configurations, MR values, channel spacings, and fibers.

Note that the star markers in Fig. 9.6a are simulations that use a receiver that is optimum in AWGN and *makes no attempt* at mitigating the component of NLI, which can be classified as *nonlinear phase noise* (NLPN). Such an NLI component turns out to be correlated over time, to an extent that roughly grows as the accumulated dispersion along the link and as the symbol rate squared. If enough time correlation is present, NLPN can be removed by means of relatively simple carrier-phase estimation (CPE) algorithms, and a somewhat better performance than predicted by the EGN model can be achieved. This aspect is discussed extensively in [9.50, 63, 64] and is dealt with in Sect. 9.5.2 of this chapter.

The Reduced EGN Model

The EGN model is very accurate but quite complex. On the other hand, in virtually all cases of practical interest, very good accuracy is still achieved by using the so-called EGN-SCI-X1 [9.67] approximation, which we call here *reduced EGN model*. It consists of neglecting those EGN model terms whose contribution is typically the smallest. The remaining terms are still substantially more complex, as a whole, than the GN model, but about only 1/3 as complex as the complete EGN model. With reference to Fig. 9.6a, the curves of the reduced EGN model would be completely superimposed to those of the EGN model (less than 1% MR difference), in all cases presented there.

To deal with the EGN model, we have to introduce a different notation from that used for the GN model in (9.44). The reason is that the term $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$, cannot be expressed in terms of the WDM signal PSD. Rather, the Fourier transforms of the individual WDM channel *pulses* are called into play. Please refer to Appendix 9.B for a detailed list of symbol definitions.

The overall WDM data signal is written in the time domain as

$$s_{\mathrm{WDM}}(t) = \sum_{n_{\mathrm{ch}}=1}^{N_{\mathrm{ch}}} \sum_{r} (a_{x,n_{\mathrm{ch}}}^{r}\hat{x} + a_{y,n_{\mathrm{ch}}}^{r}\hat{y}) \times s_{n_{\mathrm{ch}}}(t - rT_{n_{\mathrm{ch}}})\mathrm{e}^{\mathrm{j}2\pi f_{n_{\mathrm{ch}}}t}\,. \tag{9.62}$$

The $a_{x,n_{\mathrm{ch}}}^{r}$ and $a_{y,n_{\mathrm{ch}}}^{r}$ are the symbols sent on the $\hat{x}$ and $\hat{y}$ polarizations in the r signaling time slot, on the n_{ch} the WDM channel. We then assume that all quantities are normalized, so that the power carried by each WDM channel is given by

$$P_{n_{\mathrm{ch}}} = \mathrm{E}\{|a_{x,n_{\mathrm{ch}}}^{r}|^2 + |a_{y,n_{\mathrm{ch}}}^{r}|^2\}\,. \tag{9.63}$$

Table 9.1 Exact values of Φ and Ψ parameters

Format	Φ	Ψ
PM-BPSK	1	−4
PM-QPSK	1	−4
PM-8QAM	2/3	−2
PM-16QAM	17/25	−52/25
PM-QPSK	69/100	−211/100
PM-64QAM	13/21	−5548/3087
PM-128	1105/1681	−135044/68921
PM-256	257/425	−12532/7225
PM-∞-QAM	3/5	−12/7
PM-Gaussian	0	0

We also define the following quantities related to the fourth and sixth moments of the channel symbols

$$\Phi = 2 - \frac{\mathrm{E}\{|a|^4\}}{\mathrm{E}^2\{|a|^2\}}\,,$$
$$\Psi = -\frac{\mathrm{E}\{|a|^6\}}{\mathrm{E}^3\{|a|^2\}} + 9\frac{\mathrm{E}\{|a|^4\}}{\mathrm{E}^2\{|a|^2\}} - 12\,, \tag{9.64}$$

where a is any of the $a^r_{x,n_{\mathrm{ch}}}$ or of the $a^r_{y,n_{\mathrm{ch}}}$, which are assumed to all be identically distributed. The exact values of Φ and Ψ for the most commonly-used constellations are shown in Table 9.1. They vary substantially among low-cardinality constellations, whereas they change little among high-cardinality ones. We also report the limit values for a QAM constellation made up of infinitely many signal points, uniformly distributed within a square region whose center is the origin (the PM-∞-QAM entry in Table 9.1). Note that the values for PM-64QAM are already very close to such a limit.

To compute the GN model contribution $G^{\mathrm{GN}}_{\mathrm{NLI}}(f)$ according to the signal notation introduced above, (9.44) can still be used, with the substitution

$$G_{\mathrm{S}}(f) = \sum_{n_{\mathrm{ch}}=1}^{N_{\mathrm{ch}}} P_{n_{\mathrm{ch}}} R_{n_{\mathrm{ch}}} \, |\tilde{s}_{n_{\mathrm{ch}}}(f)|^2 \;, \tag{9.65}$$

where $\tilde{s}_{n_{\mathrm{ch}}}(f) = \mathrm{F}\{s_{n_{\mathrm{ch}}}(t)\mathrm{e}^{\mathrm{j}2\pi f_{n_{\mathrm{ch}}}t}\}$. Regarding the correction term $G^{\mathrm{corr}}_{\mathrm{NLI}}(f)$, its overall expression for the *reduced* EGN model is

$$G^{\mathrm{corr}}_{\mathrm{NLI}}(f) = P^3_{m_{\mathrm{c}}}\left[\Phi_{m_{\mathrm{c}}}\rho^{m_{\mathrm{c}}}_{\mathrm{SCI}}(f) + \Psi_{m_{\mathrm{c}}}\tau^{m_{\mathrm{c}}}_{\mathrm{SCI}}(f)\right] + P_{m_{\mathrm{c}}}\sum_{\substack{n_{\mathrm{ch}}=1\\ n_{\mathrm{ch}}\neq m_{\mathrm{c}}}}^{N_{\mathrm{ch}}} P^2_{n_{\mathrm{ch}}}\Phi_{n_{\mathrm{ch}}}\rho^{n_{\mathrm{ch}}}_{\mathrm{X1}}(f)\,, \tag{9.66}$$

where we have assumed that the channel under test (CUT) is the m_{c}-th channel, *not necessarily* the center channel in the WDM comb.

The terms in (9.66) bearing the subscript SCI (self-channel interference) are EGN correction terms to the NLI produced by the CUT onto itself. Their expression is

$$\begin{aligned}\rho^{m_{\mathrm{c}}}_{\mathrm{SCI}}(f) = \frac{80}{81}R^2_m \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f_1 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f_2 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f'_2 \\ \times |\tilde{s}_{m_{\mathrm{c}}}(f_1)|^2\tilde{s}_{m_{\mathrm{c}}}(f_2)\tilde{s}^*_{m_{\mathrm{c}}}(f'_2)\tilde{s}^*_{m_{\mathrm{c}}}(f_1+f_2-f) \\ \times \tilde{s}_{m_{\mathrm{c}}}(f_1+f'_2-f) \\ \times \mathrm{LK}(f_1,f_2,f_1+f_2-f)\mathrm{LK}^*(f_1,f'_2,f_1+f'_2-f) \\ + \frac{16}{81}R^2_{m_{\mathrm{c}}} \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f_1 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f_2 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f'_2 \\ \times |\tilde{s}_{m_{\mathrm{c}}}(f_1+f_2-f)|^2\tilde{s}_{m_{\mathrm{c}}}(f_1)\tilde{s}_{m_{\mathrm{c}}}(f_2) \\ \times \tilde{s}^*_{m_{\mathrm{c}}}(f_1+f_2-f'_2)\tilde{s}^*_{m_{\mathrm{c}}}(f'_2) \\ \times \mathrm{LK}(f_1,f_2,f_1+f_2-f) \\ \times \mathrm{LK}^*(f_1+f_2-f'_2,f'_2,f_1+f_2-f)\,,\end{aligned} \tag{9.67}$$

$$\begin{aligned}\tau^{m_{\mathrm{c}}}_{\mathrm{SCI}}(f) = \frac{16}{81}R_{m_{\mathrm{c}}} \\ \times \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f_1 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f_2 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f'_1 \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f'_2 \\ \times \tilde{s}_{m_{\mathrm{c}}}(f_1)\tilde{s}_{m_{\mathrm{c}}}(f_2)\tilde{s}^*_{m_{\mathrm{c}}}(f_1+f_2-f) \\ \times \tilde{s}^*_{m_{\mathrm{c}}}(f'_1)\tilde{s}^*_{m_{\mathrm{c}}}(f'_2)\tilde{s}_{m_{\mathrm{c}}}(f'_1+f'_2-f) \\ \times \mathrm{LK}(f_1,f_2,f_1+f_2-f)\mathrm{LK}^*(f'_1,f'_2,f'_1+f'_2-f)\,,\end{aligned} \tag{9.68}$$

where the integration limits were made explicit and correspond to the frequency interval occupied by the CUT, i.e.,

$$f \in \left[f_{m_{\mathrm{c}}} - \frac{B_{m_{\mathrm{c}}}}{2}, f_{m_{\mathrm{c}}} + \frac{B_{m_{\mathrm{c}}}}{2}\right]. \tag{9.69}$$

The terms in (9.66) bearing the subscript *X1* are EGN correction terms to the NLI produced by *XPM* (cross-phase modulation) or, according to a different taxonomy [9.37], due to *XCI* (cross-channel interference). Their expression is

$$\begin{aligned}\rho^{n_{\mathrm{ch}}}_{\mathrm{X1}}(f) = \frac{80}{81}R_{m_{\mathrm{c}}}R_{n_{\mathrm{ch}}} \\ \times \int_{f_{m_{\mathrm{c}}}-B_{m_{\mathrm{c}}}/2}^{f_{m_{\mathrm{c}}}+B_{m_{\mathrm{c}}}/2}\mathrm{d}f_1 \int_{f_{n_{\mathrm{ch}}}-B_{n_{\mathrm{ch}}}/2}^{f_{n_{\mathrm{ch}}}+B_{n_{\mathrm{ch}}}/2}\mathrm{d}f_2 \int_{f_{n_{\mathrm{ch}}}-B_{n_{\mathrm{ch}}}/2}^{f_{n_{\mathrm{ch}}}+B_{n_{\mathrm{ch}}}/2}\mathrm{d}f'_2 \\ \times |\tilde{s}_{m_{\mathrm{c}}}(f_1)|^2\tilde{s}_{n_{\mathrm{ch}}}(f_2)\tilde{s}^*_{n_{\mathrm{ch}}}(f'_2) \\ \times \tilde{s}^*_{n_{\mathrm{ch}}}(f_1+f_2-f)\tilde{s}_{n_{\mathrm{ch}}}(f_1+f'_2-f) \\ \times \mathrm{LK}(f_1,f_2,f_1+f_2-f)\mathrm{LK}^*(f_1,f'_2,f_1+f'_2-f)\,,\end{aligned} \tag{9.70}$$

where the integration limits were made explicit and correspond to either the frequency interval occupied by the CUT as written in (9.69) or by the generic n_{ch}-th WDM channel

$$f \in [f_{n_{\mathrm{ch}}} - \frac{B_{n_{\mathrm{ch}}}}{2}, f_{n_{\mathrm{ch}}} + \frac{B_{n_{\mathrm{ch}}}}{2}] \,. \tag{9.71}$$

9.4.9 Gaussian-Shaped Constellations

If a *PM-Gaussian* constellation is used, then all the factors Φ and Ψ in (9.66) are zero, according to Table 9.1. This implies $G_{\mathrm{NLI}}^{\mathrm{corr}}(f) = 0$ or, equivalently, $G_{\mathrm{NLI}}^{\mathrm{EGN}}(f) = G_{\mathrm{NLI}}^{\mathrm{GN}}(f)$.

Given the current rapidly growing interest in Gaussian-shaped constellations (GSCs), this is an important result. To check it, we ran the sample test set of Fig. 9.6a using an ideal GCS, taking as target MIs the values of GMI of the six QAM systems in Fig. 9.6a. The results are shown in Fig. 9.6b. Simulations agree very well with the GN/EGN curve. Therefore, if systems using GSCs earned widespread adoption, NLI modeling complexity would reduce to that of the GN model, but EGN model accuracy should be expected. This circumstance might help in the design and real-time handling of future physical-layer-aware networks based on GCSs. A specifically devoted paper was recently published on this topic, combining it with future ultra-high symbol rates [9.64]. As a caveat, GSCs are known to generate more NLPN than QAM constellations. For a discussion of the possible impact of this aspect on modeling and for more information on how simulations involving GSCs can be performed, see [9.64].

9.4.10 GN Model Closed-Form Approximate Solutions

In some cases, approximate closed-form solutions (ACFSs) can be found not just for the link factor, as shown in Sect. 9.4.6, but for the overall NLI PSD $G_{\mathrm{NLI}}(f)$, for either the GN or the EGN model. Several ACFSs have been proposed, among them [9.27, 37–41, 67, 68]. Being approximate solutions, each one has specific limitations that must be taken into account.

For the EGN model, however, only one ACFS is currently available, consisting of an asymptotic form (in the number of spans) of the correction term $G_{\mathrm{NLI}}^{\mathrm{corr}}(f)$. It was proposed in [9.68] and then upgraded in [9.67, 69]. Though effective, it only works well for relatively long links.

Instead, several ACFSs have been found for the GN model. A very accurate GN model ACFS is available for transparent and uniform systems using an ideal Nyquist-WDM comb (zero roll-off and channel spacing equal to the symbol rate). It was originally derived in [9.37] and consists of (9.7), (9.13), and (9.23) in that paper, combined.

An extension to non-Nyquist systems, for identical channels with uniform channel spacing, was also proposed in [9.37] ((9.7), (9.15) and unnumbered formula after (9.23)). Its accuracy is quite good for narrow channel spacing but may degrade for large channel spacing.

An ACFS not requiring either transparency, uniformity, or identical, equally-spaced channels, and therefore of great potential usefulness, was reported in [9.41]. Such a formula is quite general, but it approximates the iGN model, rather than GN, i.e., it assumes *incoherent NLI accumulation* (Sect. 9.4.7). On the other hand, the results from the literature [9.50, 63, 64] as well as Fig. 9.6, show the iGN model to be close to the GN model in most practical situations, including that of GSC-based systems (Fig. 9.6b). Two more assumptions are made, which were needed in the derivation: the channels have approximately rectangular PSD (that is, very small roll-off) and span loss is at least 7 dB, with greater than 10 dB being the optimal condition. Also, this ACFS provides an estimate of $G_{\mathrm{NLI}}(f)$ at the *center* of any one of the WDM channels. For the purpose of system performance assessment, it is then necessary to assume that $G_{\mathrm{NLI}}(f)$ is flat over the channel of interest. This *local white noise* assumption was studied in [9.37] and [9.41] and was shown to typically induce a small error. The formula is

$$\begin{aligned} G_{\mathrm{NLI}}(f_{i_{\mathrm{ch}}}) = {} & \frac{16}{27} \sum_{n_{\mathrm{s}}=1}^{N_{\mathrm{s}}} \gamma_{n_{\mathrm{s}}}^2 L_{\mathrm{eff},n_{\mathrm{s}}}^2 \\ & \times \prod_{p=1}^{n_{\mathrm{s}}-1} \Gamma_p^3 \mathrm{e}^{-6\alpha_p L_{\mathrm{s}}^{(p)}} \times \prod_{p=n_{\mathrm{s}}}^{N_{\mathrm{s}}} \Gamma_p \mathrm{e}^{-2\alpha_p L_{\mathrm{s}}^{(p)}} \\ & \times \sum_{n_{\mathrm{ch}}=1}^{N_{\mathrm{ch}}} G_{\mathrm{S}}^2(f_{n_{\mathrm{ch}}}) G_{\mathrm{S}}(f_{i_{\mathrm{ch}}}) \times (2 - \delta_{n_{\mathrm{ch}} i_{\mathrm{ch}}}) \times \Theta_{n_{\mathrm{ch}} i_{\mathrm{ch}} n_{\mathrm{s}}} \,, \end{aligned} \tag{9.72}$$

where $G_{\mathrm{NLI}}(f_{i_{\mathrm{ch}}})$ is the NLI PSD at the center frequency $f_{i_{\mathrm{ch}}}$ of the i_{ch}-th channel of the comb; $G_{\mathrm{S}}(f_{n_{\mathrm{ch}}})$ and $G_{\mathrm{S}}(f_{i_{\mathrm{ch}}})$ are the PSDs of the WDM data signal at the center frequency of the n_{ch}-th and i_{ch}-th channels, respectively; $\delta_{n_{\mathrm{ch}} i_{\mathrm{ch}}}$ is Kronecker's delta, i.e., it is one if $n_{\mathrm{ch}} = i_{\mathrm{ch}}$ and zero otherwise; finally, Θ is

$$\begin{aligned} & \Theta_{n_{\mathrm{ch}} i_{\mathrm{ch}} n_{\mathrm{s}}} \\ & \approx \frac{\operatorname{asinh}\left[\pi^2 (2\alpha_{n_{\mathrm{s}}})^{-1} |\beta_{2,n_{\mathrm{s}}}| f_{n_{\mathrm{ch}}} - f_{i_{\mathrm{ch}}} + \frac{B_{n_{\mathrm{ch}}}}{2} B_{i_{\mathrm{ch}}}\right]}{4\pi (2\alpha_{n_{\mathrm{s}}})^{-1} |\beta_{2,n_{\mathrm{s}}}|} \\ & \quad - \frac{\operatorname{asinh}\left[\pi^2 (2\alpha_{n_{\mathrm{s}}})^{-1} |\beta_{2,n_{\mathrm{s}}}| f_{n_{\mathrm{ch}}} - f_{i_{\mathrm{ch}}} - \frac{B_{n_{\mathrm{ch}}}}{2} B_{i_{\mathrm{ch}}}\right]}{4\pi (2\alpha_{n_{\mathrm{s}}})^{-1} |\beta_{2,n_{\mathrm{s}}}|}, \\ & \qquad n_{\mathrm{ch}} \neq i_{\mathrm{ch}} \end{aligned} \tag{9.73}$$

$$\Theta_{i_{\rm ch}i_{\rm ch}n_{\rm s}} \approx \frac{\operatorname{asinh}\left[\frac{\pi^2}{2}|\beta_{2,n_{\rm s}}|(2\alpha_{n_{\rm s}})^{-1}B_{i_{\rm ch}}^2\right]}{2\pi|\beta_{2,n_{\rm s}}|(2\alpha_{n_{\rm s}})^{-1}},$$
$$n_{\rm ch} = i_{\rm ch}\,, \tag{9.74}$$

where $B_{n_{\rm ch}}$ and $B_{i_{\rm ch}}$ are the bandwidth of the $n_{\rm ch}$-th and $i_{\rm ch}$-th channels, respectively. Note that the formula can be easily upgraded to support lumped frequency-dependent gain or loss, as well as to allow us to account for the drop-off and join-in of channels at any of the span starts.

Regarding accuracy, the typical absolute error between the predictions of (9.72)–(9.74) versus numerical integration of the iGN model is typically 2–3% of MR, making it a valuable tool for real-time system appraisal.

While this chapter was being finalized, two GN/iGN model ACFS were proposed, which upgrade (9.72)–(9.74) to also support dispersion slope through β_3, frequency-dependent loss and interchannel stimulated Raman scattering [9.70, 71]. They represent substantial progress, since they make it possible to analyze ultra-broadband systems, such as (C+L)-band ones. Due to their closed form, they could potentially become effective tools for real-time physical-layer-aware management of fully-loaded reconfigurable optical networks.

Closed-Form NLI Modeling and Raman Amplification

In previous sections, the GN and EGN general expressions (9.44)–(9.46) were provided in such a form that they can support arbitrary and possibly frequency-dependent amplification.

One overall ACFS is currently available for the GN model assuming *ideal distributed* amplification, i.e., with $\alpha(z) = g(z)$ at each point along the link, and ideal Nyquist-WDM transmission. It is (9.24) in [9.37], but of course this represents a completely ideal reference scenario.

An exact closed form was provided for the LK factor in Sect. 9.4.6, under simplifying assumptions (undepleted backward-propagating pump, frequency-flat gain). By means of various simplifications and approximations of the LK term and of the integration procedure, an ACFS for frequency-independent Raman amplification was proposed in [9.72] and extensively tested with good results. In [9.73, 74], the LK factor was approximated in suitable ways such that subsequent simplified numerical integration was possible, even in the most general case of depleted-pump and frequency-dependent Raman amplification.

It should also be pointed out that, as shown extensively in [9.50], provided that Raman amplification is backward-pumped and provides gain which is at least 6 dB lower than fiber loss, then its effect on NLI generation is small. In practice, it is possible to neglect it, incurring only about 2–3% MR estimation error. This result is significant, since in many practical cases Raman amplification complies with the above condition, as part of the span loss is compensated for by Raman amplification and part by an EDFA placed at the end of the span. This is called *hybrid Raman-EDFA amplification* and currently represents the typical solution employed in terrestrial (new or refurbished) links. In such cases, the effect of Raman on NLI generation can essentially be ignored.

9.5 Time-Domain Perturbative Model

In this section, we explore a perturbation-based *time-domain* model [9.75], describing the way in which data symbols transmitted into the optical fiber are perturbed by the nonlinearity of the fiber. This approach is alternative to the frequency-domain FWM approach described in the previous section. While it leads to similar results to the EGN model, it also permits us to predict statistical correlations between different temporal NLI symbols and to discern among different types of NLI. In particular, it allows to single out and characterize its so-called PPRN components (phase and polarization rotation noise).

In the following, we ignore nonlinear perturbations generated by the presence of co-propagating ASE noise and focus on SPM perturbations caused by the symbols transmitted over the channel of interest, as well as on XPM and FWM perturbations caused by the data symbols transmitted over neighboring interfering WDM channels.

In the framework of first-order perturbation analysis, the symbols received after ideal dispersion compensation can be written as

$$\boldsymbol{r}_n = \boldsymbol{a}_n + \Delta\boldsymbol{a}_n\,, \tag{9.75}$$

where the two-element column vectors $\boldsymbol{a}_n$ and $\boldsymbol{r}_n$ represent the transmitted and received polarization multiplexed data symbols of the channel of interest in the n-th time slot. The two-element data vector $\Delta\boldsymbol{a}_n$ corresponds to the first-order nonlinear perturbation. The various signal-induced contributions to $\Delta\boldsymbol{a}_n$ can be categorized as resulting from SPM, XPM, and FWM.

Their time-domain representation is given by [9.53, 75]

$$\Delta \boldsymbol{a}_n^{\mathrm{SPM}} = \sum_{l,k,m} \boldsymbol{a}_{n+l} \boldsymbol{a}_{n+k}^{\dagger} \boldsymbol{a}_{n+m} S_{l,k,m}, \tag{9.76}$$

$$\Delta \boldsymbol{a}_n^{\mathrm{XPM}} = \sum_{l,k,m,j} \left(\boldsymbol{b}_{n+k,j}^{\dagger} \boldsymbol{b}_{n+m,j} \mathbf{I} + \boldsymbol{b}_{n+m,j} \boldsymbol{b}_{n+k,j}^{\dagger} \right) \times \boldsymbol{a}_{n+l} X_{l,k,m}(\Omega_j), \tag{9.77}$$

$$\Delta \boldsymbol{a}_n^{\mathrm{FWM}} = \sum_{\substack{l,k,m \\ j_1,j_2,j_3}} \boldsymbol{b}_{n+l,j_1} \boldsymbol{b}_{n+k,j_2}^{\dagger} \boldsymbol{b}_{n+m,j_1} \times F_{l,k,m}(\Omega_{j_1}, \Omega_{j_2}, \Omega_{j_3}) , \tag{9.78}$$

where $\boldsymbol{b}_{n,j}$ represents the two-element data vector transmitted in the n-th time slot over the j-th interfering WDM channel, having a frequency separation of Ω_j from the channel of interest. The FWM kernels $F_{l,k,m}(\Omega_{j_1}, \Omega_{j_2}, \Omega_{j_3})$ satisfy

$$F_{l,k,m}(\Omega_{j_1}, \Omega_{j_2}, \Omega_{j_3}) = \mathrm{j}\gamma \frac{8}{9} \int_0^L \int f(z) s_0^*(z,t) \times s_{\Omega_{j_1}}(z, t - lT) s_{\Omega_{j_2}}^*(z, t - kT) s_{\Omega_{j_3}}(z, t - mT) \mathrm{d}z \mathrm{d}t , \tag{9.79}$$

where γ is the nonlinear coefficient of the fiber, L and $f(z)$ are the length and power profile of the link, T is the symbol time duration, and where $s_{\Omega}(z,t)$ represents the dispersed waveform of the pulse transmitted over a WDM channel spaced by Ω from the channel of interest, when reaching point z along the fiber. The SPM and XPM kernels are given by $S_{l,k,m} = F_{l,k,m}(0,0,0)$ and $X_{l,k,m}(\Omega) = F_{l,k,m}(0,\Omega,\Omega)$, respectively. We note that in cases where the channel spacing is sufficiently low, there are additional XPM contributions that do not appear in (9.77) and that involve interactions between three data-vectors from the closest interfering channel or interactions between two data vectors from the channel of interest and a single data vector from the closest interfering channels.

9.5.1 XPM Time-Varying ISI Representation

The effect of XPM on the received data symbols of the channel of interest can be described as time-varying ISI [9.76, 77]. This can be viewed by rewriting (9.77) as

$$\Delta \boldsymbol{a}_n^{\mathrm{XPM}} = \sum_h \mathbf{H}_l^{(n)} \boldsymbol{a}_{n+l} , \tag{9.80}$$

where the 2×2 ISI matrices are given by

$$\mathbf{H}_l^{(n)} = \sum_{k,m,j} \left(\boldsymbol{b}_{k+n,j}^{\dagger} \boldsymbol{b}_{m+n,j} \mathbf{I} + \boldsymbol{b}_{m+n,j} \boldsymbol{b}_{k+n,j}^{\dagger} \right) X_{l,k,m}(\Omega_j) . \tag{9.81}$$

Assuming that the data symbols transmitted over the interfering channels are unknown at the transmitter and receiver implies that the ISI matrices $\mathbf{H}_l^{(n)}$ are unknown as well. The summation over k and m in (9.81) implies that the set of data vectors $\boldsymbol{b}_{n,j}$ contributing to $\mathbf{H}_l^{(n)}$ changes with n and, hence, the ISI matrices vary with time, as indicated by the superscript (n). In highly dispersive systems a large number of data vectors $\boldsymbol{b}_{n,j}$ may participate in the summation of (9.81), implying that the ISI matrices $\mathbf{H}_l^{(n)}$ change slowly with n. The temporal correlation of these ISI matrices is further discussed in Sect. 9.5.4.

9.5.2 Nonlinear Phase and Polarization Rotation

The zeroth-order XPM contribution $\mathbf{H}_0^{(n)} \boldsymbol{a}_n$ possesses some interesting and unique properties that have important implications on the statistical properties of NLIN and on its mitigation [9.53, 55, 75]. As shown in [9.55] and in the Appendix of [9.78], its effect can be written as

$$\boldsymbol{a}_n + \mathbf{H}_0^{(n)} \boldsymbol{a}_n = \begin{pmatrix} \mathrm{e}^{\mathrm{j}\theta_n^{(x)}} & \mathrm{j}h_n \\ \mathrm{j}h_n^* & \mathrm{e}^{\mathrm{j}\theta_n^{(y)}} \end{pmatrix} \begin{pmatrix} a_n^{(x)} \\ a_n^{(y)} \end{pmatrix} , \tag{9.82}$$

where $a_n^{(x)}$ and $a_n^{(y)}$ are the two elements of the data vector $\boldsymbol{a}_n$, and where $\theta_n^{(x)}$, $\theta_n^{(y)}$ and h_n are given by

$$\theta_n^{(x)} = \gamma \frac{8}{9} \sum_{k,m,j} \left(2 b_{n+k,j}^{(x)*} b_{n+m,j}^{(x)} + b_{n+k,j}^{(y)*} b_{n+m,j}^{(y)} \right) \times X_{0,k,m}(\Omega_j) , \tag{9.83}$$

$$\theta_n^{(y)} = \gamma \frac{8}{9} \sum_{k,m,j} \left(2 b_{n+k,j}^{(y)*} b_{n+m,j}^{(y)} + b_{n+k,j}^{(x)*} b_{n+m,j}^{(x)} \right) \times X_{0,k,m}(\Omega_j) , \tag{9.84}$$

$$h_n = \gamma \frac{8}{9} \sum_{k,m,j} b_{n+k,j}^{(y)*} b_{n+m,j}^{(x)} X_{0,k,m}(\Omega_j) , \tag{9.85}$$

with $b_{n,j}^{(x)}$ and $b_{n,j}^{(y)}$ representing the two elements of $\boldsymbol{b}_{n,j}$.

The zeroth-order XPM contribution can, therefore, be viewed as inducing independent phase noise in each polarization, as well as polarization crosstalk, which is known also as cross polarization modulation (XpolM). Another important observation that arises from (9.82)–(9.84) is that nonlinear phase noise has a strong dependence on the modulation format [9.53, 75]. This

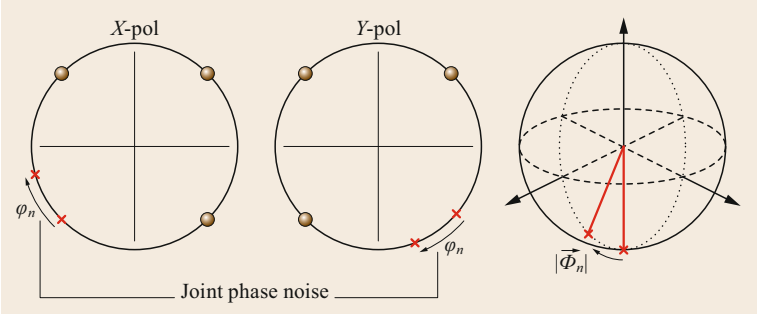


Fig. 9.7 Phase and polarization-rotation noise (PPRN) in QPSK transmission

dependence results from the fact that the terms with $k = m$ in (9.83) and (9.84) are proportional to $|b^{(x)}_{n+m,j}|^2$ and $|b^{(y)}_{n+m,j}|^2$. For constant-modulus formats, these terms are fixed for any n, implying that they induce a constant phase shift for all received data symbols (i.e., only rotating the entire received constellation). On the other hand, when amplitude modulation is introduced to the data symbols, $|b^{(x)}_{n+m,j}|^2$ and $|b^{(y)}_{n+m,j}|^2$ change with n, and $\theta^{(x)}_n$ and $\theta^{(y)}_n$ vary with time. This is why nonlinear phase noise is relatively small in systems using QPSK but significantly larger in systems employing higher-order QAM formats [9.53]. Note however, that even for constant-modulus formats, polarization crosstalk (through ih_n) can be very important.

The effect of the zeroth-order XPM contribution on the dual-polarization signal can be also described as phase and polarization rotation noise. Using the power series expansion of matrix exponentials [9.79, 80], (9.82)–(9.85) can be written as [9.55]

$$\boldsymbol{a}_n + \mathbf{H}_0^{(n)}\boldsymbol{a}_n = \mathrm{e}^{\mathrm{j}\varphi_n}\mathrm{e}^{\mathrm{j}\boldsymbol{\Phi}_n}\boldsymbol{a}_n\,, \tag{9.86}$$

where the term $\exp(\mathrm{j}\varphi_n)$ induces phase-noise and is given by

$$\varphi_n = \gamma\frac{4}{3}\sum_{k,m,j}\boldsymbol{b}^{\dagger}_{n+k,j}\boldsymbol{b}_{n+m,j}X_{0,k,m}(\Omega_j)\,, \tag{9.87}$$

and where the matrix exponential $\exp(\mathrm{j}\boldsymbol{\Phi}_n)$ induces polarization rotation noise and is given by

$$\boldsymbol{\Phi}_n = \gamma\frac{8}{9}\sum_{k,m,j}\left(\boldsymbol{b}_{n+m,j}\boldsymbol{b}^{\dagger}_{n+k,j} - \frac{1}{2}\boldsymbol{b}^{\dagger}_{n+k,j}\boldsymbol{b}_{n+m,j}\mathbf{I}\right) \times X_{0,k,m}(\Omega_j)\,. \tag{9.88}$$

Figure 9.7 visualizes the two contributions of the zeroth-order XPM effect. The phase-noise part $\mathrm{e}^{\mathrm{j}\varphi_n}$ rotates the two polarizations together by a common angle φ_n equal to the average of $\theta^{(x)}_n$ and $\theta^{(y)}_n$, whereas the polarization-state rotation $\exp(\mathrm{j}\boldsymbol{\Phi}_n)$ causes the Stokes vector representing $\boldsymbol{a}_n$ on the Poincaré sphere [9.80] to rotate about $\boldsymbol{\Phi}_n$, the Stokes vector of $\boldsymbol{\Phi}_n$, at an angle equal to its length $|\boldsymbol{\Phi}_n|$. This motivates the term *phase and polarization-rotation noise* (PPRN) to denote the zeroth-order XPM contribution. Furthermore, the PPRN description captures the unitarity of the zeroth-order XPM effect, which does not impair the norm of the transmitted dual-polarization data vector $\boldsymbol{a}^{\dagger}_n\boldsymbol{a}_n$.

9.5.3 Variance of PPRN and Higher-Order XPM Terms

In what follows, we examine the relative importance of the various XPM contributions by examining their individual contribution to the overall XPM variance. The predictions are based on the calculations of [9.81] (also shown in the Appendix of [9.78]), which were shown to be in excellent agreement with split-step simulations. The calculations assume polarization-multiplexed transmission of statistically independent data symbols, isotropically symmetric in their phase space and are given here, for simplicity of notation, for perfect Nyquist pulses with zero roll-off. The dependence of XPM on the roll-off factor is typically small but may become significant for roll-off factors larger than ≈ 0.2. We refer the interested reader to Appendix B of [9.81] for further information on how to modify the formulas below to account for nonzero roll-off factors.

The variance of XPM contributions can be written as $P_0P^2_{\mathrm{int}}\chi_{\mathrm{XPM}}$, where P_0 and P_{int} are the average launch power of the channel of interest and the interfering channel. Following the assumptions before, the various XPM terms are uncorrelated, and the XPM nonlinear coefficient, χ_{XPM}, can be written as [9.81]

$$\chi_{\mathrm{XPM}} = \sum_l \chi^{(l)}_{\mathrm{XPM}} = \sum_l \chi^{(l)}_{\mathrm{XPM},1} + \left(\frac{\langle|b|^4\rangle}{\langle|b|^2\rangle^2} - 2\right)\chi^{(l)}_{\mathrm{XPM},2}\,, \tag{9.89}$$

where b represents a single data symbol transmitted over one of the polarizations of the interfering channel, and where the angled brackets denote statistical averaging. The term $\langle|b|^4\rangle/\langle|b|^2\rangle^2$ accounts for the dependence of XPM on the fourth-order moment of the interfering data symbols.

The contribution of the l-th XPM term $\mathbf{H}_l^{(n)}\boldsymbol{a}_{n+l}$ to the overall XPM variance is, therefore, given by $P_0 P_{\mathrm{INT}}^2 \chi_{\mathrm{XPM}}^{(l)}$, where the coefficients $\chi_{\mathrm{XPM},1}^{(l)}$ and $\chi_{\mathrm{XPM},2}^{(l)}$ are given in [9.81]

$$\chi_{\mathrm{XPM},1}^{(l)} = \frac{32}{27}\frac{\gamma^2}{T^2}\sum_j \int \rho_j(\omega_1,\omega_2,\omega_3)\rho_j^*(\omega_4,\omega_2,\omega_3) \times \mathrm{e}^{\mathrm{j}(\omega_1-\omega_4)lT}\frac{\mathrm{d}^4\omega}{(2\pi)^4},$$

$$\chi_{\mathrm{XPM},2}^{(l)} = \frac{80}{81}\frac{\gamma^2}{T}\sum_j \int \rho_j(\omega_1,\omega_2,\omega_3)\rho_j^*(\omega_4,\omega_5,\omega_6) \times \mathrm{e}^{\mathrm{j}(\omega_1-\omega_4)lT}\frac{\mathrm{d}^5\omega}{(2\pi)^5}, \tag{9.90}$$

with $\omega_6 = \omega_5 - \omega_2 + \omega_3$. The kernels $\rho_j(\omega_1,\omega_2,\omega_3)$ are given by

$$\rho_j(\omega_1,\omega_2,\omega_3) = \tilde{s}_0^*(\omega_1-\omega_2+\omega_3)\tilde{s}_0(\omega_1)\tilde{s}_j^*(\omega_2)\tilde{s}_j(\omega_3) \times \int_0^L f(z)\mathrm{e}^{\mathrm{j}(\omega_1-\omega_2-\Omega_j)(\omega_2-\omega_3)\beta''z}\mathrm{d}z, \tag{9.91}$$

where β'' is the dispersion coefficient of the fiber, and where $\tilde{s}_0(\omega)$ and $\tilde{s}_j(\omega)$ represent the baseband spectral shape of the transmitted pulses from the channel of interest and the j-th interfering channel.

In Fig. 9.8 we examine the significance of the various XPM contributions in standard 16-QAM WDM systems with 100 km spans, 115 32 Gbaud channels, and 37.5 GHz channel spacing. We plot the variance of the various XPM contributions $\mathbf{H}_l^{(n)}\boldsymbol{a}_{n+l}$ normalized by the variance of the zeroth-order XPM contribution $\mathbf{H}_0^{(n)}\boldsymbol{a}_n$. Evidently, the PPRN contribution has the most pronounced effect whereas the significance of the XPM terms $\mathbf{H}_l^{(n)}\boldsymbol{a}_{n+l}$ with $|l|>0$ decreases monotonically with $|l|$.

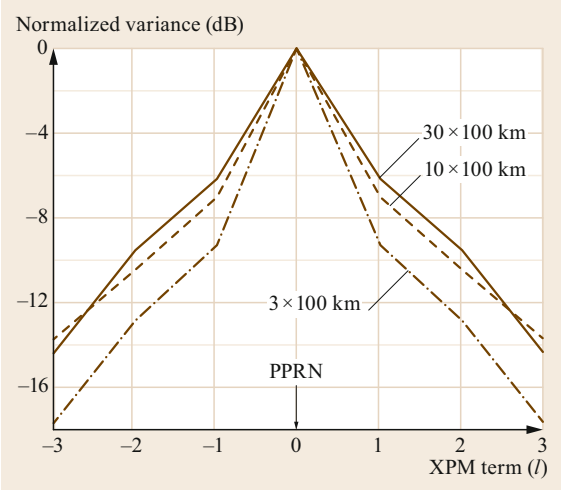


Fig. 9.8 Variance of the various XPM terms normalized by the variance of the zeroth-order XPM term (PPRN) for fully-loaded 115-channel transmission using 16-QAM

9.5.4 Temporal Correlations

An important observation regarding the ISI form of XPM is the fact that the ISI matrices $\mathbf{H}_l^{(n)}$ change slowly with n, as the matrix elements are the result of a summation over a significant number of interfering symbols, see (9.81). Following the assumptions of Sect. 9.5.3, the temporal autocorrelation function of the diagonal elements of $\mathbf{H}_l^{(n)}$ is given in [9.81]

$$R_{\mathrm{diag}}^{(l)}(s) = R_{\mathrm{diag},1}^{(l)}(s) + \left(\frac{\langle|b|^4\rangle}{\langle|b|^2\rangle^2} - 2\right)R_{\mathrm{diag},2}^{(l)}(s), \tag{9.92}$$

where

$$R_{\mathrm{diag},1}^{(l)}(s) = \frac{80}{81}\frac{\gamma^2}{T^2}\sum_j \int \rho_j(\omega_1,\omega_2,\omega_3)\rho_j^*(\omega_4,\omega_2,\omega_3) \times \mathrm{e}^{\mathrm{j}(\omega_1-\omega_4)lT}\mathrm{e}^{\mathrm{j}(\omega_2-\omega_3)sT}\frac{\mathrm{d}^4\omega}{(2\pi)^4},$$

$$R_{\mathrm{diag},2}^{(l)}(s) = \frac{80}{81}\frac{\gamma^2}{T}\sum_j \int \rho_j(\omega_1,\omega_2,\omega_3)\rho_j^*(\omega_4,\omega_5,\omega_6) \times \mathrm{e}^{\mathrm{j}(\omega_1-\omega_4)lT}\mathrm{e}^{\mathrm{j}(\omega_2-\omega_3)sT}\frac{\mathrm{d}^5\omega}{(2\pi)^5}, \tag{9.93}$$

with $\omega_6 = \omega_5 - \omega_2 + \omega_3$. The temporal autocorrelation function of the off-diagonal elements of $\mathbf{H}_l^{(n)}$ is given in [9.81]

$$R_{\mathrm{off-diag}}^{(l)}(s) = \frac{1}{5}R_{\mathrm{diag},1}^{(l)}(s). \tag{9.94}$$

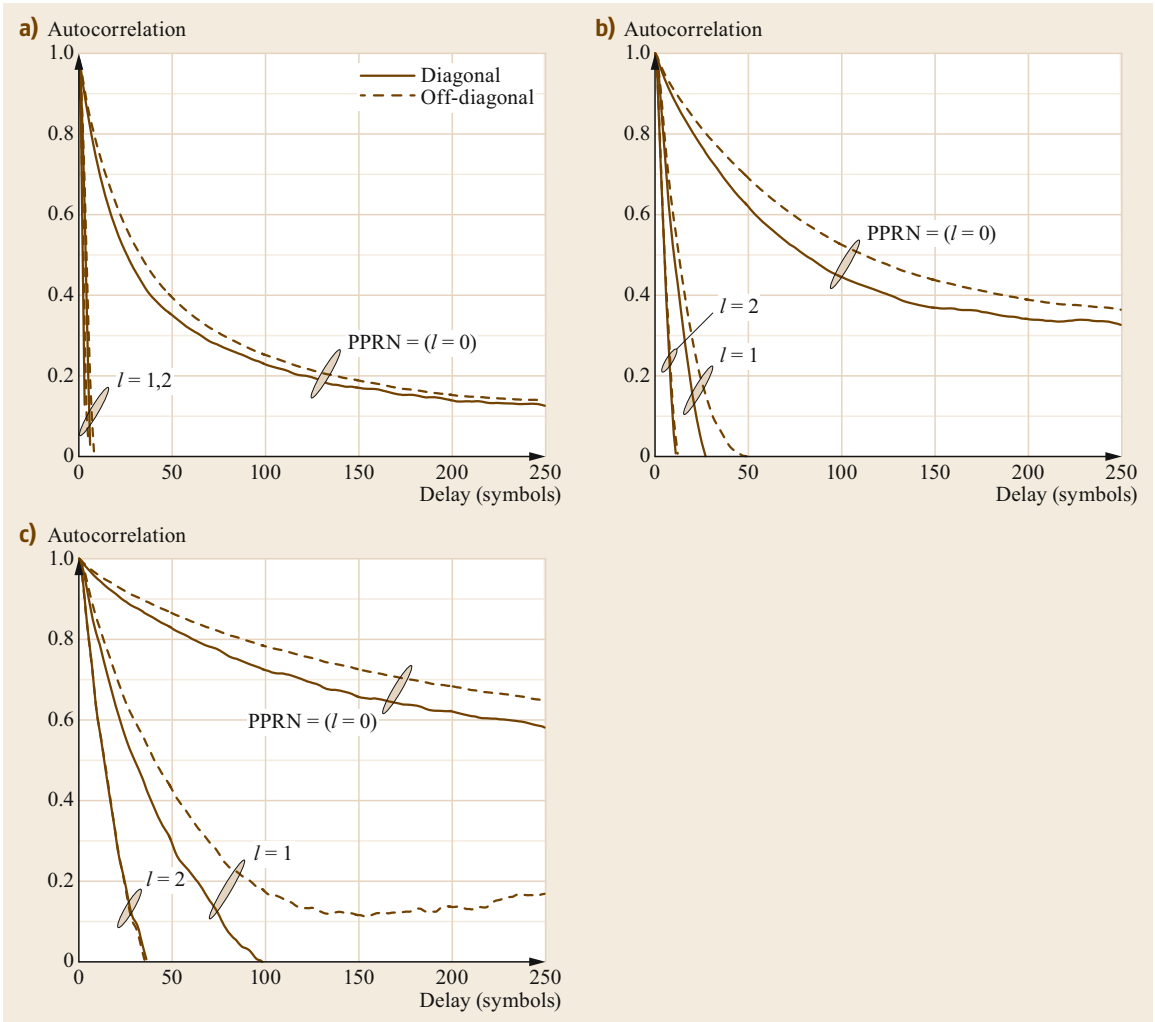


Fig. 9.9a–c Autocorrelation of the diagonal (*solid*) and off-diagonal (*dashed*) elements of $\mathbf{H}_l^{(n)}$ for $l = 0, 1, 2$ in 16-QAM fully loaded systems with 115 channels. The diagonal and off-diagonal elements of $\mathbf{H}_0^{(n)}$ induce nonlinear phase-noise and polarization crosstalk. **(a)** 3×100 km; **(b)** 10×100 km; **(c)** 30×100 km

In Fig. 9.9, we show the autocorrelation function of the diagonal and off-diagonal elements of $\mathbf{H}_l^{(n)}$ for $l = 0, 1, 2$, considering 16-QAM transmission, 100 km spans, and standard fully loaded systems with 115 32 Gbaud channels and 37.5 GHz channel spacing. The correlations of the higher-order XPM terms quickly drop to zero, even in systems operating over a 3000 km link. The correlations of nonlinear PPRN ($l = 0$), however, are relatively long, of the order of tens of symbols. These results are aligned with numerical simulations [9.82–85] and experimental measurements [9.86, 87] verifying the long temporal correlations of the phase-noise contribution of PPRN (diagonal elements with $l = 0$), as well as with the numerical [9.83, 84] and experimental [9.88, 89] verification of the long temporal correlations of the polarization crosstalk contribution of PPRN (off-diagonal elements with $l = 0$).

Regarding the specific system impact of PPRN and NLPN, see Sect. 9.9.2.

9.5.5 Pulse Collision Theory

The time-domain model further provides insights into how the various NLI contributions are formed. By examining the optical field of the transmitted signal as a collection of temporal pulses transmitted from multiple WDM channels, one can characterize and analyze

the different nonlinear interactions between the transmitted pulses as they propagate through the optical fiber.

In particular, in the limits of first-order perturbation analysis, the interactions can be classified as collisions between two pulses, three pulses, and four pulses; these collisions can be either complete, or incomplete. A rigorous and comprehensive analysis of the various types of collisions is provided in [9.55]. Each type of collision is shown to have its unique signature, and the overall nature of NLI is determined by the relative significance of the various collisions in a given WDM transmission. The most important contributions to NLI are shown to follow from two-pulse and four-pulse collisions. Two-pulse collisions generate NLI in the form of PPRN, whereas four-pulse collisions generate complex circular ISI noise. In addition, two-pulse collisions are shown to have a strong dependence on the modulation format of the interfering WDM channels and to be most pronounced when the collision is complete, whereas four-pulse collisions are shown to be modulation-format independent and to be strongest when the collision is incomplete. The theory further shows that in short optical links with lumped amplification, and links with perfect distributed amplification, NLI tends to be dominated by PPRN contributions generated by two-pulse collisions; in long links with lumped amplification, the importance of four-pulse collisions increases, and NLI becomes more complex circular with reduced dependence on modulation format.

9.6 Spatially-Resolved Perturbative Models and Their Applications

The perturbative method discussed in Sects. 9.3.1 and 9.4 shows the perturbation as a double integral in the frequency domain, weighted by the kernel function. Strictly speaking, each frequency of the signal is perturbed by all possible FWM combinations, and the kernel represents their efficiency.

This picture is a global input/output description of the whole optical link, and it is of great interest for the system perspective as shown in several sections of this chapter. Nevertheless, it is of interest to relate such a global scale description to a local scale at a given coordinate. Such a relation highlights several physical connections that may help to understand the perturbative models and to search for novel applications of perturbation theory in nonlinear optical communications.

The local description of the perturbation is the dynamical model expressed by the differential equation (9.7). We find it particularly interesting to solve it by following the numerical split-step idea, much as we did for the NLSE with the SSFM. The corresponding block diagram is shown in Fig. 9.10 [9.23, 25, 90]. For the sake of comparison, we also show the SSFM method in Fig. 9.10.

The RP1 solution is the sum of the unperturbed and the first-order perturbation. The unperturbed term, by definition, has been discretized by the concatenation of linear blocks only, here indicated by the letter L. Each linear block thus summarizes linear effects such as attenuation, GVD, etc., into an infinitesimal step h.

The perturbation is, indeed, the result of many contributions, each represented by a branch in the block diagram. In such paths we find nonlinear blocks, which implement the input/output relation $N(A) = -\mathrm{j}\gamma|A|^2Ah$. We note that, since the received perturbation is linear in γ, the electric field, according to RP1, while traveling along the link can cross only one nonlinear block from input to output. However, such a nonlinearity can appear anywhere, hence the reason for the final sum creating the perturbative term as a sort of multipath interference.

The idea can be iterated, creating higher-order RP schemes [9.91]. For instance, RP2 can be emulated by considering all possible combinations of two nonlinear blocks within the discretized distance grid. The SSFM algorithm can be viewed as an RP∞ method implemented in a clever way, where the nonlinear block is substituted by the SPM operator $N(A) = \mathrm{e}^{-\mathrm{j}\gamma|A|^2h}$.

In Fig. 9.10, we refer to RP1 as parallel RP1 because of its parallel nature, which is not the case of SSFM whose serial structure is unavoidable. Besides the physical intuition, such a parallel structure might find some advantages in numerical implementation when a high degree of parallelization is available [9.92].

We can exploit the RP1 diagram to derive a spatially-resolved GN model [9.25, 27]. Our target is the variance of the output perturbation γA_1, and more generally its autocorrelation function. Since the output perturbation is the sum of many contributions (Fig. 9.10) to get our target we need the cross-correlation among any two generic paths forming the perturbation [9.27, 93]. Let us focus on two such paths, the ones where the nonlinearity appears after s and z km, respectively. See Fig. 9.10 for reference and for the main definitions of variables.

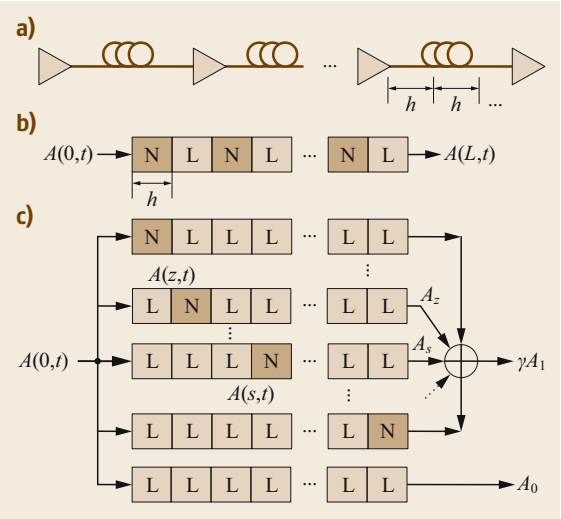


Fig. 9.10 (a) Physical link, (b) its corresponding standard SSFM model, and (c) its parallel RP1 model

The common input to such paths is a linear digital modulated signal $A(z=0,t)$, with t time. In case of interest, such a signal is actually a cyclo-stationary stochastic process, i.e., its statistical properties vary cyclically in time with a period equal to the symbol time. However, we believe that the signal can be safely treated as stationary. This claim finds its ground in the observation that we are planning to work with dispersion uncompensated links where the strong interference induced by dispersion likely removes higher order cyclo-stationary effects, and because we mainly focus on sinc-like pulses such that the signal is almost stationary already at the fiber input.

In this framework, if we know the cross-correlation function of the outgoing fields from the nonlinear blocks, here called $R_{zs}^{(\mathrm{out})}(\tau)$, with τ the time lag, the cross-correlation of the received fields on such paths can be found by using basic linear system theory [9.94]. In detail, the cross-correlation between A_z and A_s at the output coordinate L is related to $R_{zs}^{(\mathrm{out})}(\tau)$ by

$$\mathbb{E}[A_z(t+\tau)A_s^*(t)] = h_{zL}(\tau) \otimes h_{sL}^*(-\tau) \otimes R_{zs}^{(\mathrm{out})}(\tau) \,,$$

with $\otimes$ denoting convolution and $h_{zL}(\tau)$ the impulse response of the filter accounting for all linear effects from coordinate z to output coordinate L. Such an expression can be efficiently evaluated in the frequency domain by working with power spectral densities.

With similar arguments, we can relate the autocorrelation function of the transmitted field $A(0,t)$ to the ingoing fields of the nonlinear blocks at coordinates z and s, respectively,

$$\begin{aligned} R_{zs}^{(\mathrm{in})} &\triangleq \mathbb{E}[A(z,t+\tau)A^*(s,t)] \\ &= h_{0z}(\tau) \otimes h_{0s}^*(-\tau) \otimes R_{00}(\tau) \,, \end{aligned} \tag{9.95}$$

with $R_{00}(\tau) = \mathbb{E}[A(0,t+\tau)A^*(0,t)]$. If we are able to relate $R_{zs}^{(\mathrm{out})}(\tau)$ to $R_{zs}^{(\mathrm{in})}(\tau)$, we therefore have a relation between the system input/output autocorrelation functions and, thus, our target can definitely be solved by summing, i.e., integrating, all such contributions.

Such a relation is possible in the special, yet relevant, case of $A(z,t)$ with Gaussian statistics at any coordinate [9.19]. If we get rid of the constant phase-shift term as in the eRP (Sect. 9.3.1), the relation takes a simple and elegant form [9.27]

$$R_{zs}^{(\mathrm{out})}(\tau) = 2\left|R_{zs}^{(\mathrm{in})}(\tau)\right|^2 R_{zs}^{(\mathrm{in})}(\tau) \,. \tag{9.96}$$

Such assumptions are exactly those of the GN model, whose spatially-resolved reference formula is, thus,

$$\tilde{S}_{\mathrm{GN}}(f) = \int_0^L \int_0^L \tilde{h}_{zL}(f)\tilde{h}_{sL}^*(f) \times \mathcal{F}\{R_{zs}^{(\mathrm{out})}(\tau)\}\mathrm{d}z\mathrm{d}s \,, \tag{9.97}$$

where $\tilde{S}_{\mathrm{GN}}(f)$ is the PSD of the received perturbation, i.e., the Fourier transform of the autocorrelation function. The variance of the nonlinear interference given by the GN model is simply the integral of this PSD.

Such a solution has several similarities with the basic RP idea adopted for the electric field. Strictly speaking, we are propagating the signal correlation function from input to output by performing first a linear operation according to (9.95), then the same first-order nonlinear perturbation identical to the one experienced by the electric field except for a factor 2, then again linear effects up to the output. The main difference is that while the field perturbation is the sum of many perturbative infinitesimal contributions, here we have to account for all possible pairs of cross-correlations, thus increasing the complexity to a double integration. However, it has been shown that the double integral can be reduced to a single integration by exploiting the exponential behavior of attenuation along the distance [9.27].

What about the general case of non-Gaussian statistics for the propagating signal? The previous idea can still be used by properly modifying (9.96). Now the cross-correlation of the signals outgoing the nonlinear blocks cannot be related just to the same cross-correlation between the inputs, but we have to account

for higher-order statistics because of the nonlinear transformation [9.75]. Since nonlinearity is cubic, the cross-correlation involves products of six random variables. Such random variables are the digital symbols of the modulation format under use. Among all possible combinations, many of them have zero average for classical modulation formats with uniform distribution and rotationally symmetric constellations. If we let a_k be the information symbol at time k, we are left with just three kinds of nonzero contributions to the cross-correlation: one involving $\{|a_k|^2|a_n|^2|a_l|^2\}$ with $k \neq n \neq l$, one with $\{|a_k|^4|a_n|^2\}$ with $k \neq n$, and one for $\{|a_k|^6\}$. The first partition is the biggest one and generates terms like the one in (9.96), since indexes are different and symbols independent. Hence, such a partition is the one accounted by the GN model. The other two partitions generate the higher-order contributions of the EGN model [9.65].

In summary, the cross-correlation $R_{zs}^{(\mathrm{out})}(\tau)$ can be related to the shape of the supporting pulse distorted by linear effects up to coordinate z, $p(z,t) \triangleq h_{0z}(t) \otimes p(0,t)$, by [9.25]

$$R_{zs}^{(\mathrm{out})}(\tau) = \underbrace{\frac{\mu_2^3}{T^3} 2|Q(z,s,\tau)|^2 Q(z,s,\tau)}_{\text{GN model}} + \frac{\mu_2 \kappa_{2;2}}{T^2}\left[4F_4(z,s,\tau) + Q_4(z,s,\tau)\right]Q(z,s,\tau) + \frac{\kappa_{3;3}}{T} Q_6(z,s,\tau)\,, \tag{9.98}$$

with $\kappa_{i;i}$ the i-th cumulant of transmission symbol ([9.25] Appendix B), $\mu_2 \triangleq \mathbb{E}[|a_k|^2]$, and

$$\begin{aligned} Q(z,s,\tau) &\triangleq p(z,\tau) \otimes p^*(s,-\tau)\,, \\ Q_4(z,s,\tau) &\triangleq p^2(z,\tau) \otimes (p^*(s,-\tau))^2\,, \\ F_4(z,s,\tau) &\triangleq |p(z,\tau)|^2 \otimes |p(s,-\tau)|^2\,, \\ Q_6(z,s,\tau) &\triangleq [|p(z,\tau)|^2 p(z,\tau)] \\ &\quad \otimes [|p(s,-\tau)|^2 p^*(s,-\tau)]\,. \end{aligned} \tag{9.99}$$

In particular, square M-QAM modulation with i.i.d. symbols has

$$\kappa_{2;2} = -\frac{3}{5}\frac{M+1}{M-1}\mu_2^2, \qquad \kappa_{3;3} = \frac{36}{21}\frac{M^3-1}{(M-1)^3}\mu_2^2\,.$$

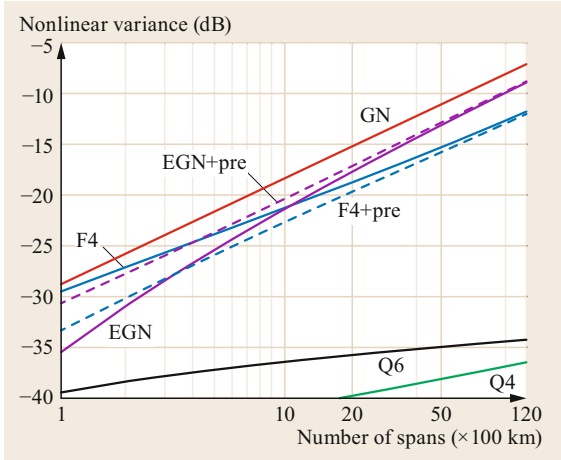


Fig. 9.11 Contributions to nonlinear interference variance (normalized to the cube of signal power P) versus the number of spans N_s. 133 PDM-QPSK channels ($R = 32$ Gbaud, $\Delta f = 37.5$ GHz) transmitted into a $N_s \times 100$ km SMF dispersion uncompensated link; F_4 and Q_4 contributions are plotted in absolute value, since they are actually negative. The impact of a precompensation of 8500 ps/nm on some terms is shown by a *dashed line*. Note that GN model overestimates the EGN prediction (after [9.25])

The expression for $R_{zs}^{(\mathrm{out})}(\tau)$ can finally be inserted in (9.97), and the PSD of the nonlinear interference can be evaluated by numerical integration. Please note that with Gaussian distributed symbols $\kappa_{i;i} = 0$, $i > 1$, so that (9.98) coincides with (9.96), since $R_{zs}^{(\mathrm{in})}(\tau) = (\mu_2/T)Q(z,s,\tau)$, for a digital signal with supporting pulse $p(0,t)$.

The EGN model can be read as an additive correction to the GN model. Although the EGN is more accurate, it is worth noting that, except for very peculiar modulation formats [9.52], the GN model gives a conservative overestimation. The previous derivation was for single channel in single polarization. Extensions to PDM and WDM are available [9.25].

The relative importance of each term forming the EGN model is shown in Fig. 9.11 for a specific example detailed in the figure caption. We observe the mentioned GN overestimation compared to the EGN model. The most important EGN correction is the one provided by the fourth-order correction F_4 [9.75]. Such a term is partially mitigated at short lengths by predistorting the signal with a precompensation fiber.

9.7 Multiplicative Models and Their Applications

We might wonder, at this point, whether the models introduced in the previous sections definitively solve the problem of modeling the optical fiber channel. For instance, the representation of NLI as AWGN provides a good accuracy in terms of performance evaluation (the EGN model, in particular) with a reasonably low complexity (the GN model, in particular). However, it also entails that NLI cannot be mitigated, which is not entirely correct. For instance, deterministic intrachannel nonlinear effects can be exactly compensated for by digital backpropagation (DBP) (Sect. 9.11.1). Moreover, as shown in Sect. 9.5, some NLI terms due to XPM are not independent of the signal and should be more properly interpreted as time-varying ISI, rather than noise. In general, this means that the accuracy of AWGN-like models in terms of performance evaluation does not necessarily correspond to an accurate and detailed characterization of the statistical properties of NLI (e.g., non-Gaussian statistics, temporal correlation, and so on). These properties are not particularly relevant to determine the performance of systems employing conventional symbol-by-symbol detection, but might be the key to devise improved detection and nonlinearity mitigation strategies—e.g., able to mitigate also interchannel nonlinearity and signal–noise interaction, or to perform DBP with a lower complexity. Research of alternative models for the optical-fiber channel is, hence, still in progress. Strictly connected with this topic, is the computation of channel capacity (discussed in Sect. 9.12), which entails an accurate knowledge of the channel and the use of the best possible modulation and detection strategy.

A possible approach is provided by the LP model, briefly introduced in Sect. 9.3.1, which falls within the broader class of *multiplicative models*. As opposed to *additive models*, which describe nonlinear effects by means of a noise-like additive term, multiplicative models describe nonlinear effects through a fading-like multiplicative term—a change of the channel characteristics that causes a distortion of the propagating signal. The latter approach is usually more complex than the former, but also closer to the underlying physics—the Kerr effect being a change of the refractive index induced by the propagating optical signal that causes a phase rotation of the signal itself.

The LP model was originally introduced in [9.26] (and generalized to all orders in [9.91]) to provide an analytical approximation of the signal at the output of a nonlinear dispersive fiber given the input signal. It was later combined with the RP model to investigate signal–noise interaction, providing a joint description of both parametric gain and nonlinear phase-noise effects [9.95]. It has been eventually applied to study interchannel nonlinearity in WDM systems, originating the frequency-resolved LP (FRLP) model [9.76, 96]. This last model and its applications will be briefly described in the sequel. The model has been developed for the single-polarization case, but the concept can be extended to dual polarization signals.

To investigate the impact of interchannel nonlinearity in WDM systems, it is convenient to divide the overall WDM signal into the channel-of-interest (COI) component, still referred to as $A(z,t)$, and the interfering channels (IC) component $A_w(z,t)$—resulting from the combination of all the other WDM channels. Expanding the nonlinear term of the NLSE (9.1) into its SPM, XPM, and FWM components (according to the role played by the COI and IC components) [9.1], and retaining only the XPM one, the equation can be rewritten as

$$\frac{\partial A}{\partial z} = \mathrm{j}\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - 2\mathrm{j}\gamma|A_w|^2 A\,. \tag{9.100}$$

Equation (9.100) does not account for SPM, which can be compensated for by DBP (Sect. 9.11.1), and for FWM and signal–noise interaction, whose impact in WDM systems is usually negligible compared to XPM (this might not be the case when considering WDM channels with a low symbol rate). The evolution of A_w is, in turn, subject to nonlinear interference from A, such that we should also write an analogous equation for A_w, coupled to (9.100). However, in terms of the impact of A_w on A, the nonlinear evolution of A_w is a second-order effect. Thus, we will neglect it and assume that A_w evolves linearly and independently of A. In this case, (9.100) is a linear Schrödinger equation with a space and time-variant potential, whose solution can be formally expressed as

$$A(L,t) = \int_{-\infty}^{\infty} H_w(f,t)\tilde{A}(0,f)\mathrm{e}^{\mathrm{j}2\pi ft}\mathrm{d}f\,, \tag{9.101}$$

where $H_w(f,t)$ is a *time-variant* transfer function [9.97]. In mathematics, the solution for this kind of equation is typically given in terms of *Green's function* (or *propagator* in quantum mechanics), to which the time-variant transfer function is simply related by a Fourier transform.

The transfer function can be obtained from the FRLP model and expressed as [9.76, 96]

$$H_w(f,t) = \mathrm{e}^{-\mathrm{j}\theta(f,t)}\,, \tag{9.102}$$

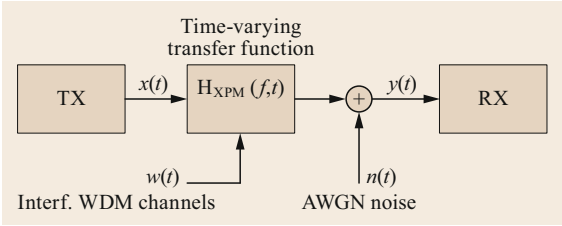


Fig. 9.12 Continuous-time FRLP model

where the XPM term $\theta(f,t)$ depends on the input IC signal $A_w(0,t)$ and on the link characteristics. Some general expressions for $\theta(f,t)$ and its statistics are provided in [9.76, 96].

Eventually, including the effect of amplifier noise, the overall system (from the COI standpoint) can be schematically modeled as in Fig. 9.12, in which $x(t) = A(0,t)$ is the transmitted signal, $y(t)$ the received signal, $H_w(f,t)$ the time-varying transfer function of the channel, which depends on the overall signal transmitted by the other channel users (the interfering channels) $w(t) = A_w(0,t)$, and $n(t)$ an AWGN term that accounts for the accumulated optical amplifier noise.

A key observation here is that all the neglected effects (e.g., FWM and signal–noise interaction) can eventually be reintroduced by properly increasing the power spectral density of the AWGN term, for instance, based on the calculations from the GN or EGN models or extracting the parameters from a numerical simulation. However, when modeling an effect as AWGN, we basically give up the possibility of mitigating it and make a worst-case assumption in terms of system performance. On the other hand, a more accurate modeling of the dominant XPM term may allow for its partial mitigation.

From the COI viewpoint, the model in Fig. 9.12 describes the nonlinear time-invariant optical fiber channel as a linear time-variant one. This apparent paradox is explained by the fact that channel nonlinearity is accounted for by the dependence of $H_w(f,t)$ on $w(t)$ [9.96]. However, assuming that $w(t)$ is unknown to both the transmitter and the receiver, such nonlinearity remains hidden, and the effect of $H_w(f,t)$ is simply perceived as a linear distortion. Moreover, since $w(t)$ depends on time, the channel transfer function $H_w(f,t)$ also depends on time.

The model in Fig. 9.12 is substantially that of a doubly-dispersive fading channel, often used in wireless communications, whose key features are the coherence time and bandwidth over which the channel remains strongly correlated [9.98]. This analogy may help to better understand the channel characteristics and behavior, as well as to devise improved transmission and detection strategies.

The coherence properties of the channel were studied in [9.99] and are illustrated by the contour plots in Fig. 9.13, which shows the correlation between the values $\theta(0,t)$ and $\theta(\Delta f, t+\tau)$ of the XPM term at two different times and frequencies inside the COI bandwidth, as a function of the delay τ and frequency separation Δf. The channel is assumed to be stationary in time but not in frequency, as the impact of the XPM term depends on the frequency distance from the interfering channels and varies inside the COI bandwidth. The correlation is analyzed by holding one frequency fixed in the middle of the COI bandwidth (conventionally set to $f = 0$), and letting the other vary inside the COI bandwidth. The scenarios considered refer to a Nyquist–WDM system with 50 GHz channel spacing and bandwidth, Gaussian symbols (a worst-case assumption in terms of XPM impact), and a 1000 km link of standard single-mode fiber with either ideal distributed amplification (IDA) (Fig. 9.13a) or 10×100 km lumped amplification (LA) (Fig. 9.13b). Only the XPM term generated by the couple of closest interfering channels (i.e., those located at $f = \pm 50$ GHz) is considered. The coherence is quite substantial in the IDA link, but significantly reduced in the LA link.

The FRLP model has been used to derive some closed-form expressions for the AIR with different modulation formats and a mismatched decoder optimized for the AWGN channel [9.96]. In this case, similar results could be obtained also with some additive models such as the RP or EGN models. However, the FRLP model can also be used to devise more efficient detection strategies that exploit the coherence properties of the channel and to derive some improved capacity bounds [9.82]. These issues will be discussed in greater detail in Sect. 9.12.

Equation (9.100) is also the starting point to derive other models and descriptions of interchannel nonlinearity. In [9.100, 101], Feynman path integrals and diagrammatic techniques are used to derive Green's function (propagator) of (9.100), from which the input–output channel statistics are obtained by assuming that the stochastic potential $|A_w|^2$ is a Gaussian process with short-ranged correlations in space and time—a more drastic approximation than the assumption of linear propagation of A_w done in deriving (9.102). The resulting model has been used to compute some capacity lower bounds for the optical fiber channel [9.100, 101].

Another approach to solve (9.100) is the *Magnus expansion*, a general method to derive approximate exponential representations of the solution of linear differential equations with varying coefficients [9.102]. The Magnus expansion provides a power series expansion for the corresponding exponent and, for this reason, is

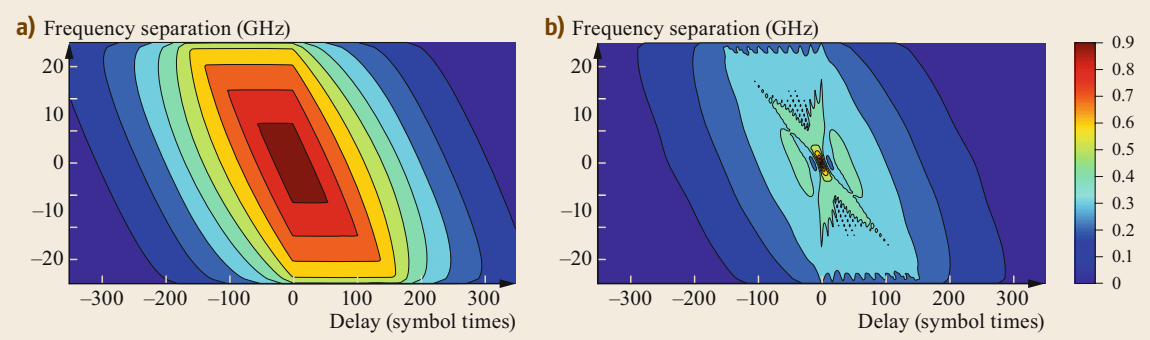


Fig. 9.13a,b XPM coherence: **(a)** IDA link; **(b)** 10×100 km LA link

sometimes referred to as time-dependent exponential perturbation theory. In optical-fiber communications, the Magnus expansion has already been employed to model linear effects such as polarization-mode dispersion and polarization-dependent loss [9.103], while its application to fiber nonlinearity has only been slightly touched [9.104]. When applied to (9.100), the Magnus expansion is quite accurate even at first order—in many cases more than the corresponding first-order RP, eRP, and LP expansions [9.104]—and has the desirable additional property of preserving at any order some important features of the original equation, such as its unitarity. These characteristics make it an interesting subject for future research.

9.8 Model-Specific Features

In this section, we try to provide a list of features of the various fiber nonlinearity models introduced so far, which make them attractive for certain applications. Carrying out a more comprehensive and comparative analysis of the models would be a difficult task, and we refrain from doing it. Some of the aspects mentioned here are dealt with in later sections, to which the reader should refer:

GN model: reasonable compromise between accuracy and complexity, very compact formula.

iGN model: for most systems, very good compromise between accuracy and complexity.

GN/iGN models: lend themselves to closed form solutions, some of which are extremely simple, yet rather accurate, such as for ideal Nyquist WDM with completely identical spans, helping establish fundamental limits; others are quite flexible (any comb, any link) but still fully closed form, very good for management of physical-layer aware optical networks.

EGN model: complete and accurate model in the frequency domain; computationally very complex; a reduced version exists, which is quite accurate and $1/3$ as complex, but complexity is still high. It does not allow us to accurately factor out PPRN, though it does allow us to approximately factor out NLPN by assuming constant-envelope transmission (i.e., assuming PM-QPSK-like parameters).

Time-domain model: very similar to the EGN model in terms of accuracy and complexity when predicting interference variance; also predicts nonlinear phase and polarization rotation noises (PPRN); predicts temporal correlations; describes nonlinear interference as an intersymbol interference (ISI), which can be used to examine and evaluate the performance of various ISI cancellation techniques; predicts the contribution of the various ISI terms.

Spatially-resolved model: alternative description of the EGN model; predicts temporal correlations of the optical nonlinear interference; focuses on nonlinear spatial interactions along the link, such as nonlinear signal–noise interaction; complexity grows with the system length.

FRLP model: represents NLI as a multiplicative distortion rather than as an additive noise; describes the channel through a time-varying transfer function; allows us to identify the NLPN component and calculate its correlation in time and frequency. These characteristics help devise improved detection and nonlinearity mitigation strategies. Closed-form solutions are available for simple system configurations (e.g., ideal Nyquist WDM).

9.9 Impact of Specific Effects on NL Modeling

Amplified spontaneous emission (ASE) noise interacts with the nonlinear Kerr effect along propagation. The result is generally referred to as nonlinear signal–noise interaction (NSNI). NSNI can be classified as an FWM process. However, it finds its own space in the literature because of the distinctive nature of some of the photons joining the FWM interaction, which come from a wide-band Gaussian distributed signal such as ASE.

9.9.1 Nonlinear Signal ASE Interaction

Initial studies on NSNI date back to investigations about the nonlinear interaction between a strong continuous-wave (CW) signal and wide-band noise [9.105, 106]. In such a case, only degenerate FWM yields significant NSNI, which motivated the use of small-signal analysis to analyze the effect. Results highlighted that an optical fiber can induce modulation instability (MI) depending on the dispersion sign. In particular, the white spectrum of noise may experience a gain (hence, the name modulation instability), which is maximum at the frequency $\sqrt{2\gamma P/|\beta_2|}$, with P the power of the strong CW [9.105].

Modulation instability has been exploited to build optical parametric amplifiers (OPA) [9.107]. OPA shows distinguished new features compared to existing optical amplifiers, such as adjustable gain spectra and center frequency, the possibility to work in phase sensitive mode with ideally zero noise figure, the possibility to perform phase conjugation and wavelength conversion [9.108]. However, OPA shows nontrivial problems such as gain-shape fluctuations and polarization dependence that have limited their applications in practice [9.108].

Besides the positive implications of NSNI in building optical amplifications, NSNI originating along the transmission fiber is generally a problem for the system performance. Initial studies showed that NSNI can induce strong phase noise, usually referred to as the Gordon and Mollenauer effect [9.109], whose impact has been shown to severely affect phase-modulated signals such as differential phase-shift keying (DPSK) transmissions [9.110]. A closed formula of the NSNI-induced phase noise variance and its main scaling properties was proposed in [9.109].

The phase noise induced by NSNI is a random variable whose exact probability density function was derived by *K.-P. Ho* in [9.111] in the absence of dispersive effects and before detection filters. The impact of dispersion is much more complex, such that the exact PDF is unknown but closed formulas exist for the NSNI variance due to intrachannel pulse XPM [9.112]. Small-signal analyses of NSNI in dispersive links were carried out in [9.113, 114] and in [9.115, 116], where they were referred to as parametric gain (PG).

ASE noise can be included in the GN and EGN models. Since ASE is inserted in the link along propagation, the spatially-resolved model of Fig. 9.10 is well suited to search for modifications to GN-based formulas [9.43].

Since ASE is Gaussian distributed, its dominant contribution appears as an enhancement of the GN part of the EGN formula (9.98). However, ASE impacts even the fourth-order nonlinear interference contribution proportional to $\kappa_{2;2}$.

One of the main implications of NSNI to GN model variance is that the NSNI nonlinear interference variance does not scale with the cube of the signal power, but with the square. This is not surprising, since the most dominant contribution to NSNI comes from FWM processes involving the beating of signal × signal × ASE [9.43, 117–119]. As a rule of thumb, numerical results showed that NSNI starts to impact the nonlinear interference variance at SNR smaller than 10 dB [9.42, 43].

Some formulas have been proposed in the literature to modify the GN model formula and account for NSNI. *Poggiolini* et al. proposed using the incoherent-GN formula by substituting the channel power at the input of fiber k with $P + P_{\mathrm{ASE}}k$, P being the transmitted power [9.42]. *Serena* ([9.43], eq. (10)) proposed correcting the cross-correlation between the nonlinear interference brought by two different spans by an ASE-dependent term. *Lavery* et al. [9.120] proposed weighting the NSNI accumulation coherence factor in a different way to the signal such that each ASE term contributes from its injection point in the link. *Ghazisaeidi* extended the theory of [9.75] by providing an efficient algorithm for NSNI variance computation [9.121].

NSNI is the fundamental limit to the performance in the presence of nonlinear equalization [9.43, 119–121], a feature that motivated the search of effective models of NSNI with the aim of finding the capacity of optical communication links. If we perturb the entire link, including digital backpropagation, we get the spatially resolved model of Fig. 9.14 [9.43]. We note that, without ASE, for any input/output branch, there is a branch with opposite parameters, such that the overall nonlinear interference is zero. However, with ASE, this symmetry is broken, such that a residual NSNI cannot be compensated for by a nonlinear equalizer. *Lavery*

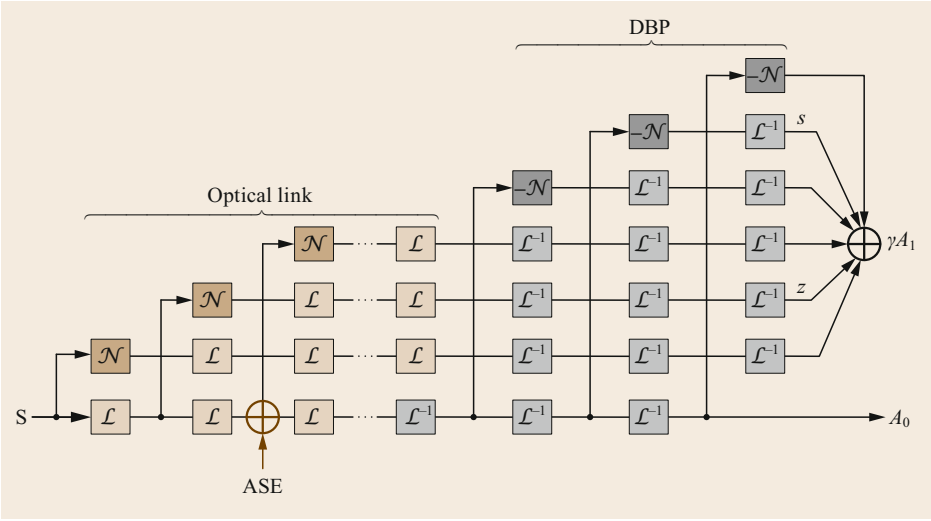


Fig. 9.14 First-order representation of an optical link including digital backpropagation (DBP) (after [9.43]). The paths labelled with s and z do not cancel out in the presence of the ASE source indicated

et al. [9.120] showed that the optimum is to equally split the nonlinear equalizer between transmitter and receiver sides.

9.9.2 System Impact of Nonlinear Phase and Polarization Noise

Section 9.5.2 provides the tools for accurate modeling of nonlinear phase and polarization noise (PPRN). These tools can then be used to assess the actual system impact of PPRN. As a relevant example, Fig. 9.15 shows the SNR gain that can be obtained by ideal removal of all interchannel PPRN. The plot is significant, since it addresses a realistic C-band fully populated 115-channel system, over SMF, with 100 km span length, at 32 GBaud. The SNR gain shown assumes that launch power is optimized to take the maximum advantage out of interchannel PPRN removal. Hence, the term *peak-SNR gain* is used ([9.78] for more details). Note that peak-SNR gain can also be interpreted, conversely, as the *impact* that PPRN has on the system. In other words, QPSK crossing the 0.5 dB line at five spans in Fig. 9.15 can be read as either a 0.5 dB potential gain if PPRN is removed or a 0.5 dB SNR degradation due specifically to PPRN.

Standard CPE (carrier-phase estimation) algorithms in the receiver DSP can mitigate the effect of PPRN, and in particular of its phase-noise component (NLPN). However, these algorithms typically leverage the time-correlation that such disturbances exhibit, quantified in Fig. 9.9. Unfortunately, where mitigation would be most needed, such as in short links according to Fig. 9.15, PPRN has short correlation time (Fig. 9.9), so that mitigation by standard CPE algorithms is only partially effective. PPRN can be almost completely removed in long links, where it exhibits long correlation times. There, though, its impact is relatively small, and the obtainable mitigation gain is limited.

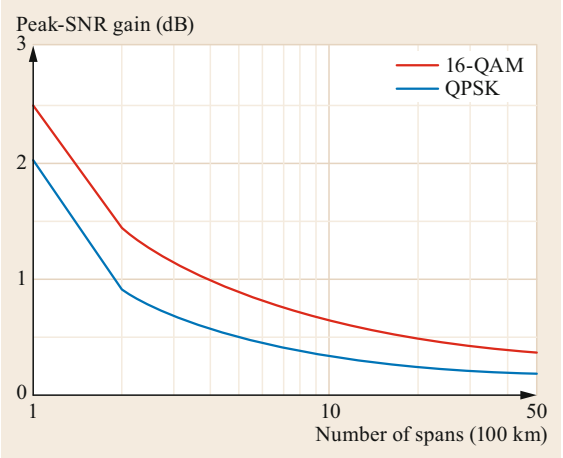


Fig. 9.15 Peak SNR gain resulting from perfect interchannel PPRN removal in fully loaded systems with 115 WDM channels at 32 GBaud. The detailed parameters used here are given by Table 2 in [9.78]. The Peak SNR gain is different between the maximum SNR, with and without interchannel PPRN. For more information on peak SNR gain, see [9.78], Sect. 4

This topic, however, is complex, and Fig. 9.15 only looks at one scenario. Many system features, such as fiber dispersion, distributed amplification, modulation format, etc., affect PPRN. For a comprehensive analysis and extensive references, see [9.78]. The symbol rate also affects PPRN. Recently, the scaling of the impact of NLPN versus *symbol rate* was studied, for instance, in [9.63, 64], using an approximate technique based on

the EGN model. The indication is that, when the symbol rate is increased, the variance of NLPN decreases and its correlation time increases. As a result, overall, its impact appears to decrease at higher symbol rates.

9.9.3 Joint Nonlinear and Polarization Effects, PMD/PDL

Although single-mode fibers support only one propagating mode, the resulting field is the superposition of two polarizations [9.7]. Each polarization can be independently modulated, and the resulting signal is generally referred to as polarization division multiplexing (PDM), also known as dual polarization (DP) or polarization multiplexing (PM). PDM is a simple technique to double the spectral efficiency and represents the starting point of more advanced techniques of spatial division multiplexing [9.122].

However, the two polarizations usually interact along propagation, thus generating polarization crosstalk at the receiver. The reason for such cross-coupling is related to perturbations to the ideal fiber structure, such as changes in the refractive index, non-circularity of the fiber, thermal stresses, variations in the core radius, irregularity at the core-cladding boundary, etc. Perturbations may be even deliberately introduced during fiber fabrication as a way to improve performance, as we will discuss later. All such stresses make the fiber anisotropic, i.e., they introduce birefringence.

As a result of birefringence, the polarizations may travel at different speeds, thus manifesting polarization-mode dispersion (PMD) [9.80]. A main implication of PMD is the end-to-end group delay spread between polarizations, called differential group delay (DGD). The problem is of particular concern in the presence of polarization coupling along transmission, because to undo the coupling at the receiver we must account for the memory introduced by DGD.

The accumulation of PMD along the distance depends on the correlation length over which mode coupling occurs, typically of about 100 m for SMF [9.7]. This way, a long-haul link operates in the strong coupling regime, whose random properties along distance eventually make the DGD a random variable, with Maxwellian statistics in the limit of correlation length approaching zero [9.123, 124]. Most importantly, the average DGD and the DGD standard deviation accumulate proportionally to the square root of the distance [9.9, 125]. This circumstance eases equalization of PMD. On the contrary, GVD has a much smaller randomness, such that it accumulates with a rate proportional to the distance.

For such reasons, short fiber correlation length is intentionally induced during fabrication by properly spinning the fiber [9.126]. Data sheets usually report the PMD coefficient η_{PMD}. Typical values are $0.32\,\mathrm{ps}/\sqrt{\mathrm{km}}$ for pre-1991 fibers, $0.13\,\mathrm{ps}/\sqrt{\mathrm{km}}$ for fibers installed in the period from 1992 to 1998, and $0.05\,\mathrm{ps}/\sqrt{\mathrm{km}}$ for post-1999 fibers [9.127]. For a fiber of length L, The average DGD $\overline{\tau}$ is related to the PMD coefficient by $\overline{\tau} = \eta_{\mathrm{PMD}}\sqrt{L}$, while the root mean square value of DGD is $\tau_{\mathrm{rms}} = \sqrt{(3\pi)/8}\,\overline{\tau}$ [9.124].

PMD is generally a problem in high symbol rate systems and must be properly equalized at the receiver side. Adaptive equalization is mandatory, since PMD varies in time on scales of the order of milliseconds. The problem is best solved with coherent detection where classical algorithms first developed for wireless communications have been successfully used. Some examples are the blind constant-modulus algorithm (CMA), the least mean squares (LMS) algorithm, and the data-aided least squares method [9.128].

Both birefringence and PMD play a role in the nonlinear regime. As mentioned in the introduction, the fast variations of birefringence with distance are usually averaged out as in the Manakov equation [9.9, 45]. In this scenario, the two polarization tributaries still interact nonlinearly along propagation through XPM, but they also interact through cross-polarization modulation (XPolM) [9.8, 9, 129–131]. XPolM is a generalization of the XPM phase rotation in the complex plane into a rotation in the three-dimensional space described by the Poincaré sphere. Now, it is the total power that is preserved; however, power can be exchanged between the polarization tributaries, thus creating nonlinear crosstalk. Contrary to PMD temporal variations, XPolM is rapidly varying in time, with temporal scales of the order of the walk-off between interacting channels. As a consequence, XPolM scattering manifests as a Brownian motion over the Poincaré sphere [9.130, 131], and its equalization is extremely challenging. XPolM has been shown to be one of the dominant nonlinear impairments in dispersion-managed PDM systems [9.132, 133].

PMD joins the polarization and nonlinear Kerr effect interactions [9.134]. The analysis is simplified in the strong coupling regime, where the reference model is the Manakov-PMD equation [9.45]. Numerical simulations and experiments [9.135, 136] showed that PMD helps improve the average performance of an optical link, provided that linear PMD is fully equalized at the receiver. The reason is that PMD is a dispersive effect, hence it decorrelates channels farther away than the PMD correlation bandwidth [9.137], thus mitigating the nonlinear interference accumulation along distance.

The numerical investigation of PMD in the nonlinear regime is particularly burdensome. To correctly

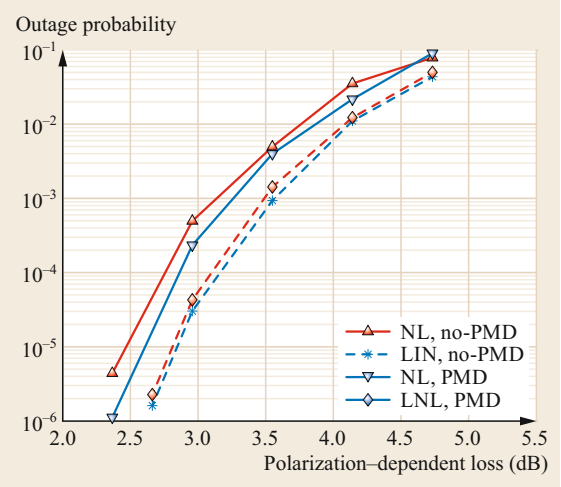


Fig. 9.16 Outage probability versus PDL. An outage event occurs when the Q-factor is smaller than 6.25 dB. *Dashed lines*: equivalent linear (LIN) model with the same average Q-factor as the in nonlinear (NL) propagation. PMD = $0.13\,\mathrm{ps}/\sqrt{\mathrm{km}}$ (after [9.138])

emulate the slow birefringence temporal variations it is mandatory to independently test many fiber realizations to find rare events that set the outage probability induced by PMD, i.e., the probability that the Q-factor is below a given threshold. Attempts to extend the GN model to include PMD are reported in [9.137].

Birefringence may also induce polarization-dependent loss (PDL), i.e., different energy losses for the two polarizations. PDL is mainly present in optical devices such as reconfigurable add-drop multiplexers (ROADM) or EDFA. PDL is expressed in international standards by [9.7]

$$\rho_{\mathrm{dB}} \triangleq 10\log_{10}\left(\frac{T_{\max}}{T_{\min}}\right),$$

where $T_{\max}$ and $T_{\min}$ are the maximum/minimum transmission powers after a PDL element. Typical values of ρ_{dB} are fractions of dB (e.g., 0.4 dB for a ROADM). The overall PDL cumulated along a link is a random variable, whose statistics can be evaluated by following similar methods as for DGD [9.139, 140].

PDL induces a penalty that is enhanced in a nonlinear regime by interaction with the nonlinear Kerr effect, thus increasing outage probability [9.141–143].

Figure 9.16 shows an example of outage probability estimated by numerical simulation in a 15-channel 32-Gbaud QPSK system, traveling in a 35×100 km, $D = 4\,\mathrm{ps}/(\mathrm{nm\,km})$, link at a power of 0.5 dBm and OSNR = 16.1 dB/0.1 nm [9.138]. PMD was either 0 or $0.13\,\mathrm{ps}/\sqrt{\mathrm{km}}$. Dashed lines show the result by an equivalent linear model, where the EDFA noise figure has been changed to get the same average Q-factor as in a nonlinear regime. The difference between the solid and the dashed lines is an indication that PMD/PDL interacted with nonlinearity along transmission.

9.10 Applications

The main outcome of the GN model is the received SNR. In the special, yet relevant, case of the additive white Gaussian noise (AWGN) channel with matched filter detection, the SNR can be converted into bit error rate (BER) by simple formulas, as shown in Appendix 9.A.

9.10.1 Which Performance Metric?

Usually, BER follows a complementary error function (erfc)-like behavior versus SNR, which does not suit a system designer that usually works on dB scales. For this reason, the Q-factor has been introduced. The Q-factor is a one-to-one relation with the BER in a reference system tracking the erfc transformation (Appendix 9.A). This way, the Q-factor coincides with the SNR in the relevant case of QPSK, and with 2·SNR for BPSK. For several other modulation formats, the relation is, luckily, almost linear with slope 1 on a dB/dB scale. Conversion graphs are available in Appendix 9.A.

Whatever the choice, BER or Q-factor, they are normally estimated before FEC decoding. The rationale is to associate a threshold to the FEC code and claim the post-FEC BER to be 0 if the pre-FEC BER is smaller than the threshold. For instance, the second generation hard-decision FEC (HD-FEC) standardized by ITU-T G.975.1 had a Q-factor threshold of 8.5 dB with an FEC overhead of 6.7%.

Nowadays, the trend is toward soft-decision FEC (SD-FEC) that can provide gains of about 1.5 dB with respect to hard-decision [9.144], or even more with iterative decoding. SD-FEC so far operate at overheads of ∼ 20% by exploiting soft decoding with advanced techniques such as low-density parity check codes (LDPC) or turbo product codes (TPC). However, it is worth noting that the higher complexity of SD-FEC may be an issue for some systems, thus HD-FEC with overhead

$\lesssim 15\%$ may be preferred for those applications where low power consumption is mandatory [9.144].

The idea of a FEC threshold works fine with binary HD-FEC and bit-interleaving, and thus BER works great in this scenario. However, it has been shown that it may fail with SD-FEC [9.145].

An alternative performance metric, more suited to SD-FEC, is the mutual information (MI). MI is a concept introduced in information theory to provide a measure of the *amount of information* obtained about one random variable through the knowledge of another random variable [9.146]. The concept of MI is strictly related to the concept of entropy, which provides a measure of unpredictability on average. For the continuous random variable Y with probability density function $p_Y(y)$, the entropy (more correctly, the differential entropy) is defined as

$$h(Y) \triangleq -\mathbb{E}[\log[p_Y(y)]] ,$$

with $\mathbb{E}[.]$ expectation. MI between two random variables X and Y is, thus, defined as

$$\mathrm{MI}(X, Y) \triangleq h(Y) - h(Y|X) , \qquad (9.103)$$

where $h(Y|X)$ is the conditional entropy, i.e., the entropy of Y conditioned to X averaged over all possible values of X. Hence, the computation of (9.103) requires knowledge of the conditional probability $p_{Y|X}(y|x)$.

MI is a relationship between two random variables. For channels with memory, we need to generalize the idea to stochastic processes, hence we should substitute MI with the information rate (IR) [9.146, 147]. Let $\mathbf{X}_N = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y}_N = (Y_1, Y_2, \ldots, Y_n)$ temporal sequences of the stochastic process $X(t)$ and $Y(t)$, respectively, from discrete time $t = 1$ to time $t = n$. The IR is defined as

$$I(\mathbf{X}_N, \mathbf{Y}_N) \triangleq \lim_{n\to\infty} \frac{1}{n}[h(\mathbf{Y}_N) - h(\mathbf{Y}_N|\mathbf{X}_N)] .$$

We are motivated to associate Y_k to the input of the soft decoder and X_k to the transmitted symbol. The GN model tells us that the AWGN channel is a good approximation of the true channel, and, thus, MI can be evaluated from the memoryless model $Y_k = X_k + n_k$, with n_k a complex Gaussian random variable. However, the true optical channel is different, not just in terms of different statistics but also because the optical channel exhibits memory, hence such a straightforward approach cannot exploit all the useful information [9.148, 149]. Nevertheless, the theory of mismatched decoding [9.150] shows that such MI is a lower bound to the maximum IR [9.147] constrained to the modulation format under use. An example of MI for QPSK modulation in AWGN [9.151] is reported in Fig. 9.17.

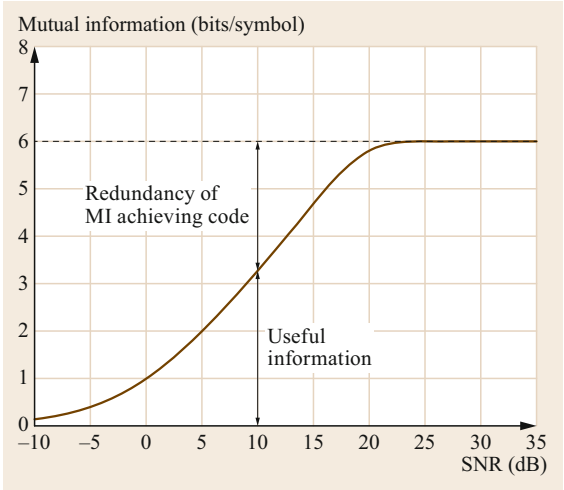


Fig. 9.17 MI of 64-QAM in AWGN. The error probability after FEC can be made arbitrarily small by a code having a redundancy of at least the indicated bits/symbol

In summary, if the soft-decoder has just the information provided by the GN model, i.e., it believes that noise is additive and Gaussian and knows its variance, we say that the corresponding MI described above is achievable, i.e., there exists a code for which error-free transmission is possible at MI bits/symbol. The redundancy of such best-code is reported, for example, in Fig. 9.17 for the reference case of SNR = 10 dB. Practical codes, able to exploit the information provided by the auxiliary channel, will need higher redundancies even if nowadays with LDPC and TPC the loss is very small.

We say that the AWGN channel approximation is an instance of an auxiliary channel of the true channel [9.147]. Other auxiliary channels are possible: the closer the auxiliary channel to the true channel, the tighter the lower bound to the maximum constrained IR. The capacity is the supremum among all possible modulation formats and symbols distributions.

MI is the relevant metric for symbol-wise receivers. However, optical communication systems often use bit-interleaved coded modulation (BICM) [9.152] whose noniterative implementation [9.153] simplifies decoding. For such schemes, the generalized mutual information (GMI) is better suited [9.145, 153]. The main difference with MI is that GMI depends on the bit-symbol mapping. If $X = (B_1 B_2 \ldots B_L)$ is the string of bits, $B_k \in (0, 1)$, corresponding to symbol X, GMI is defined as [9.153, 154]

$$\mathrm{GMI} \triangleq \sum_{k=1}^{L} \mathrm{MI}(B_k, Y) .$$

It is always GMI ≤ MI. For instance, for BPSK, it is GMI = MI because here bits ≡ symbols. QPSK is the composition of two BPSK in quadrature, hence with Gray coding and AWGN, each bit can be detected independently from the other by using the real/imaginary axis as a discriminating threshold. Hence, even for QPSK GMI = MI. In particular, $\mathrm{MI}_{\mathrm{QPSK}}(\mathrm{SNR}) = 2 \times \mathrm{MI}_{\mathrm{BPSK}}(\mathrm{SNR}/2)$.

For higher-order PSK and QAM modulation formats GMI < MI. The reason is that in these cases, the mentioned bit-independence in the detection rule is broken, and thus we can extract more information by using a symbol-wise detector rather than a bit-wise detector after bit interleaving.

For the AWGN channel the difference between GMI and MI is limited to fractions of bits (Appendix 9.A).

The debate about the most appropriate performance metric is still open. *Schmalen* et al. showed that with nonbinary FEC, MI is the right candidate [9.155]. *Cho* et al. showed that with probabilistic shaping, normalized GMI yields a better correlation with the post-FEC [9.156], while *Yoshida* et al. [9.157] proposed asymmetric information (ASI) as a better predictor in the nonlinear regime.

9.10.2 Maximum Reach and Optimum Launch Power

According to the perturbative description of the NLSE, the received SNR can be expressed by [9.3, 4]

$$\mathrm{SNR} = \frac{P}{\sigma^2_{\mathrm{ASE}} + \sigma^2_{\mathrm{NL}}}\,, \tag{9.104}$$

where P is the signal power, σ^2_{ASE} is the cumulative ASE power generated along the link, while σ^2_{NL} is the nonlinear interference variance. For a transparent N-span periodic link with optical amplifiers of noise figure F and gain G, it is $\sigma^2_{\mathrm{ASE}} = h\nu FGBN$, with h Planck's constant, ν the carrier frequency, and B the receiver bandwidth. According to GN model theory, $\sigma^2_{\mathrm{NL}} = \eta P^3$, where η is the unit-power nonlinear interference coefficient discussed in the previous sections.

Several interesting implications for system design can be inferred from (9.104). First, since ASE dominates the SNR at small powers, while nonlinear interference dominates at high powers, an optimal power exists (Fig. 9.1). Such a power is generally referred to as *nonlinear threshold* P_{NLT} and can be easily found by setting to zero $(\mathrm{dSNR})/(\mathrm{d}P)$, with the result

$$P_{\mathrm{NLT}} = \left(\frac{\sigma^2_{\mathrm{ASE}}}{2\eta}\right)^{1/3}.$$

By inverting such a relation, we find that $\sigma^2_{\mathrm{ASE}} = 2\eta P^3_{\mathrm{NLT}}$, i.e., at the optimal launched power, also known as the nonlinear threshold (NLT), ASE variance is twice the nonlinear interference variance. This important result tells us that at best power, the role of linear ASE noise is more important than the role of nonlinear distortions, such that it is better to put efforts into optimizing linear propagation rather than nonlinear propagation.

By substituting P_{NLT} into (9.104), we get the corresponding maximum SNR

$$\mathrm{SNR}_{\mathrm{NLT}} \triangleq \max(\mathrm{SNR}) = \frac{P_{\mathrm{NLT}}}{\frac{3}{2}\sigma^2_{\mathrm{ASE}}} = \frac{1}{\left[27\eta\left(\frac{h\nu FGBN}{2}\right)^2\right]^{\frac{1}{3}}}\,.$$

Besides these key ingredients to set up a connection, it is interesting to have a look at the SNR penalty with respect to linear impairments. By factoring out in (9.104) the SNR impaired by linear effects only, $\mathrm{SNR}_{\mathrm{lin}} = P/(\sigma^2_{\mathrm{ASE}})$, we can define such a penalty SP by

$$\mathrm{SP} \triangleq \frac{\mathrm{SNR}_{\mathrm{lin}}}{\mathrm{SNR}} = \frac{\sigma^2_{\mathrm{ASE}} + \sigma^2_{\mathrm{NL}}}{\sigma^2_{\mathrm{ASE}}}\,.$$

The interesting implication is that at best power P_{NLT}, the SNR penalty is always 3/2, whatever the value of σ^2_{ASE} and η, i.e., whatever the link for which the GN model assumptions work! Such a 3/2 factor is more conveniently expressed in dB by $10\log_{10} 3/2 \cong 1.76\,\mathrm{dB}$. Therefore, if our system works at power P_{NLT}, and we do not want to change such a power, at best we can improve the SNR by 1.76 dB with an ideal receiver able to exactly compensate for nonlinear impairments. Only by increasing the power can we improve the SNR by more than 1.76 dB by using a better receiver.

For the system designer used to work in dB, it is interesting to have a look at what happens by 3 dB changes in ASE and nonlinear interference power. We have

$$\sigma^{2'}_{\mathrm{ASE}} = \sigma^2_{\mathrm{ASE}} + 3\,\mathrm{dB} \Rightarrow \begin{cases} P'_{\mathrm{NLT}} = P_{\mathrm{NLT}} + 1\,\mathrm{dB} \\ \mathrm{SNR}'_{\mathrm{NLT}} = \mathrm{SNR}_{\mathrm{NLT}} - 2\,\mathrm{dB} \end{cases}$$

$$\eta' = \eta + 3\,\mathrm{dB} \Rightarrow \begin{cases} P'_{\mathrm{NLT}} = P_{\mathrm{NLT}} - 1\,\mathrm{dB} \\ \mathrm{SNR}'_{\mathrm{NLT}} = \mathrm{SNR}_{\mathrm{NLT}} - 1\,\mathrm{dB}\,. \end{cases}$$

Interestingly, a 3 dB change in ASE or nonlinear interference induces an absolute change of 1 dB in the nonlinear threshold.

The GN model can be used to infer the reach of the system at a given SNR, i.e., the maximum transmission distance ensuring the signal-to-noise ratio SNR at the receiver. Such an SNR can be, for instance, the SNR threshold of the FEC code under use. An example of signal power contour levels and their related reach is shown in Fig. 9.2. To estimate the reach we need to know how the ASE power and the nonlinear interference η scale with the number of spans N. The ASE power has a simple linear scaling $\sigma^2_{\mathrm{ASE}} = h\nu FGBN$. Experiments, numerical simulations, and theoretical models, indeed, suggest that the nonlinear interference follows the scaling law [9.41, 158, 159]

$$\eta = \eta_1 N^{1+\varepsilon} ,$$

where η_1 is the normalized nonlinear interference variance after one span, while ε is the coherence accumulation factor accounting for the superlinear accumulation of nonlinear effects along the distance. The presence of ε is due to correlations among the different paths creating the first-order perturbation (Fig. 9.10). In highly dispersive links, the strong dispersion makes the signal entering a given span almost uncorrelated with the signal entering a different span, hence we expect also the nonlinear interference to be almost uncorrelated span-by-span, such that the cumulative effect accounted by η scales linearly with distance, i.e., $\varepsilon \approx 0$. On the contrary, the accumulation factor is $0 < \varepsilon < 1$, with the extreme case of $\varepsilon = 1$ in fully-compensated dispersion-managed links. In dispersion-uncompensated links, it is typically $\varepsilon < 0.2$ [9.41, 159]; ϵ would be almost zero if only cross-channel nonlinearities were present, because the channel-walk-off decorrelates cross-nonlinear interference from span to span. An $\epsilon > 0$ is, thus, due to a significant single-channel nonlinearity. Therefore, single-channel nonlinearity grows faster with distance than cross-channel nonlinearity. This implies, for instance, that single-channel nonlinear equalizers become more effective at large propagation distances.

In this framework, the SNR is, thus, a function of three variables, σ^2_{ASE}, η_1 and ε. To get the reach, we thus need at least three measurements of power yielding the reference SNR at different distances. Based on the previous discussion, at the reach $N_0 \triangleq \max(N)$, the ASE variance is twice the nonlinear interference variance, hence the following identities hold

$$\mathrm{SNR} = \frac{P_0}{\frac{3}{2} h\nu FGBN_0} = \frac{P_0}{3\eta_1 N_0^{1+\varepsilon} P_0^3} ,$$

where P_0 is the signal power at the reach. The previous expression gives two equations in two variables (P_0, N_0), from which we get our target, the reach [9.159],

$$\mathrm{reach} \triangleq N_0 = \frac{1}{\left[(3\mathrm{SNR})^3 \eta_1 \left(\frac{h\nu FGB}{2}\right)^2\right]^{\frac{1}{3+\varepsilon}}} \quad (\mathrm{spans}) .$$

The reach is a function of the GN model key parameters η_1, ε, whose estimation is affected by errors. It is interesting to have a feeling of how reliable the formula of the reach is with variations of such parameters. Since it is preferable to work in dB scales, the following derivatives hold [9.159]

$$\frac{\partial N_0^{\mathrm{dB}}}{\partial \mathrm{SNR}^{\mathrm{dB}}} = -\frac{3}{3+\varepsilon}, \quad \frac{\partial N_0^{\mathrm{dB}}}{\partial F^{\mathrm{dB}}} = -\frac{2}{3+\varepsilon}, \quad \frac{\partial N_0^{\mathrm{dB}}}{\partial \eta_1^{\mathrm{dB}}} = -\frac{1}{3+\varepsilon} . \tag{9.105}$$

For instance, at $\varepsilon = 0$, an estimation error on η_1 of 1 dB translates into a reach error of 0.33 dB [9.41, 160].

Another key message from (9.105) is the following. To increase the reach we can play with i) the FEC by using a better code with a smaller SNR threshold, with ii) the optical link, for instance, by using better amplifiers with a smaller noise figure F, or with iii) the nonlinearity, for instance, by using a nonlinear equalizer to reduce η_1. Equation (9.105) tells us that such strategies are in order of effectiveness.

9.10.3 Modeling in Networks

The GN model has proved to be an invaluable analytical tool, able to explain the most important scaling laws and features of the new regime of highly-dispersed transmission entailed by dispersion-uncompensated links with coherent detection and electronic digital-signal processing. However, the GN model also finds one of its most practical uses in route selection in modern optical networks with quality-of-transmission (QoT) guarantees.

The dominant optical network paradigm is that of wavelength-routed optical networks (WRON), where a transparent optical channel (a light path) is established from source to destination on a fixed wavelength across multiple fiber links and optical switching nodes (Fig. 9.18). The most recent evolution of WRONs is that of elastic optical networks (EON) ([9.161, 162] and references therein), where bandwidth can be flexibly allocated to light paths with a granularity down to 6.25 GHz, due to coherent detection and smart wavelength-selective switches. Flexibility may also concern the modulation format and/or the forward error-correction (FEC) code of each light path at

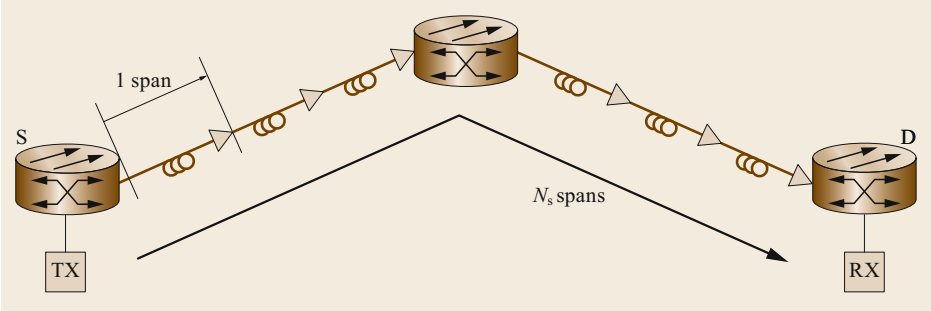


Fig. 9.18 Example of a selected light path from source (S) to destination (D) transmitted on a fixed wavelength across six spans, with two hops and three spans/hops

the transmitter in order to assure reaching the destination without intermediate regenerations. Electro-optical (EO) regenerations at intermediate nodes may, however, be necessary to both allow electrical multiplexing/demultiplexing onto the light path at intermediate nodes (*grooming*) and possibly to change the wavelength of the subsequent section of the light path in order to reduce wavelength blocking (WB). Reduction of WB through wavelength conversion has the beneficial effect of allowing an increase of network utilization (i.e., the fraction of wavelengths used in the network) and, thus, of network throughput, but at the moment it is still considered a too-costly option because of the cost of EO regenerations and is used with great care.

Initial efforts in dimensioning EONs when taking into account nonlinear propagation effects assumed a worst-case full-load scenario where all wavelengths in every fiber link are populated, and nonlinear effects are at their maximum [9.163]. Such a full-load assumption has the great advantage of simplifying the routing, modulation, and spectrum assignment (RMSA), since it decouples the propagation problem from the traffic-dependent linear and nonlinear interference. Hence, every modulation format has an optimal launch power spectral density and a maximum error-free transmission distance, known as the *reach*, and a simple comparison of the reach with the requested source-destination distance is enough to assess whether or not a connection is feasible without intermediate regeneration. The other major advantage of the full-load assumption is that of making existing light paths insensitive to the establishment of new connections, a fact that makes RMSA decisions fast even in a dynamic traffic scenario where connections my be established and released on short time scales.

The first RMSA algorithms were based on look-up tables containing the reach for every possible modulation format. The reach tables were evaluated offline by long statistical computations based on the split-step Fourier propagation method. Semi-analytical methods based on a single span propagation and tuning with variable residual input span dispersion were also proposed to speed up reach computations [9.93].

However, the fully analytical formulas afforded by the GN model quickly found their way into the RMSA design because they drastically speed up the calculation of the required received optical signal-to-noise ratio (OSNR) of a reference light path for a given modulation format, and thus the assessment of whether it is above its FEC threshold, and the connection is thus feasible.

In particular, the incoherent GN model treats the nonlinear interference terms generated at each span as if they were independent additive Gaussian noise, exactly as the optical noise from the amplifiers.

The first proposed use of the incoherent GN model in an EON scenario appeared in [9.47, 164], where the locally-optimal globally-optimal Nyquist (LOGON) strategy was introduced. It was there first observed that, when using the optimal power spectral density at full load, the reach at any load is never larger than $3/2$ of the full-load reach. This still is the major theoretical justification for using the apparently resource-wasteful full-load assumption for nonlinear interference evaluation. The reach computation using the GN model for optical networking was explored in [9.165]. The LOGON strategy was analyzed in terms of blocking probability versus carried traffic in a theoretical setting [9.166] and recently for various network topologies [9.167], while in [9.168] the expected gain in the network throughput due to transceiver adaptation was studied in EONs using the full-load LOGON strategy.

Alternative, equivalent formulations of the GN model that make the single and cross-channel contributions explicit for networking applications appeared in [9.39].

The enhanced GN (EGN) model, which takes into account the effect of the modulation format on nonlinear interference, can alternatively be used in place of the GN model [9.169] for applications where route establishment can take place on time scales of several

minutes, due to the development of fast computing techniques [9.170].

Recently, many papers have appeared that use either the GN or the EGN model within RMSA algorithms and try to remove the full-load assumption in order to make a better use of network resources.

One line of such developments [9.171–176] explores the use of the GN analytical OSNR within a mixed nonlinear programming framework in order to obtain the absolute optimal RMSA in a static traffic scenario (one where a traffic matrix is offered to a green-field network, and light paths are set up one by one to satisfy as many connections as possible) from complete knowledge of each network fiber's active light paths.

Since the problem is provably NP-hard [9.177, 178], such techniques can work in reasonable time only for impractically small networks. Heuristics have, thus, to be found in order to make them work in practical-sized networks. In this context, linearized GN formulas have been proposed in order to effectively use the GN model within a mixed integer linear programming framework [9.178, 179].

Another line of development explores instead a statistical approach to the RMSA problem, i.e., a model-based approach for light path selection, which is extremely fast to compute but has a bounded probability of being incorrect, and thus may sometimes require path recomputation. For instance, in [9.180], a QoT-based routing method was proposed, where a light path is set up based solely on wavelength load measurements along its selected physical path, and the probability that the newly set-up light path is unfeasible is bounded below a desired threshold.

Another instance of the statistical RMSA approach is one that makes use of machine learning techniques for route selection. Here, the GN/EGN model can be used to synthesize positive and negative examples of feasible light paths, and such examples are used to train the weights of a parametric model, for instance a neural network [9.181] or a random forest [9.169].

In summary, the use of the GN/EGN model for network planning and optimization is gaining momentum and is likely to become the method of choice for RMSA in future EoNs.

9.11 Nonlinearity Mitigation

This section discusses the main approaches for nonlinearity mitigation based on DSP. The interested reader might also refer to [9.182] for an extensive review on this subject.

We point out that very recently artificial neural networks (ANNs) have also been advocated as a possible DSP approach for the mitigation of nonlinear effects. We refrain from dealing with it here, but the interested reader can see, for instance, a recent prominent result [9.183].

We also point out that nonlinearity mitigation can be carried out by means of all-optical techniques as well. One example is, for instance, optical phase conjugation [9.184]. All-optical mitigation techniques are a vast and interesting investigation field, but they are outside of the scope of this chapter.

9.11.1 Digital Backpropagation

One of the most promising and studied strategies to combat fiber nonlinearity is digital backpropagation (DBP). DBP is a channel inversion technique that aims at removing fiber propagation effects by digitally emulating the propagation of the received signal through a fictitious fiber link—equal to the actual fiber link but reversed in space and with opposite-sign parameters $-\alpha_P$, $-\beta_2$, and $-\gamma$ for each span of fiber [9.185–187]—as schematically depicted in Fig. 9.19.

A simple analysis reveals that DBP can, in principle, exactly invert the propagation equation (9.1), hence removing all deterministic propagation effects due to the interplay between dispersion and nonlinearity. DBP can be implemented at the transmitter, at the receiver, or both. Several numerical studies and experimental demonstrations of DBP are available, reporting different gains in terms of SNR and reach, depending on the scenario considered [9.57, 188–190]. There is, anyway, a general consensus that gains up to 1 dB might be realistically achieved, and that DBP is a practical candidate for extending the reach of next-generation transponders.

In practice, DBP operates on a digitized version of the signal and is usually implemented by the SSFM algorithm (Sect. 9.2) for its good characteristics in terms of performance and complexity. The sequence of received samples is divided into several overlapping blocks of length N, with N_m overlapping samples. Each block is propagated through the whole link according to the algorithm in Fig. 9.3, with a total of N_s steps. For a dual-polarization signal, the number of real additions and multiplications required by an SSFM-based implementation of DBP per each processed sample is reported in Table 9.2 [9.191]. The parameter N_c accounts for alternative implementations of DBP (see

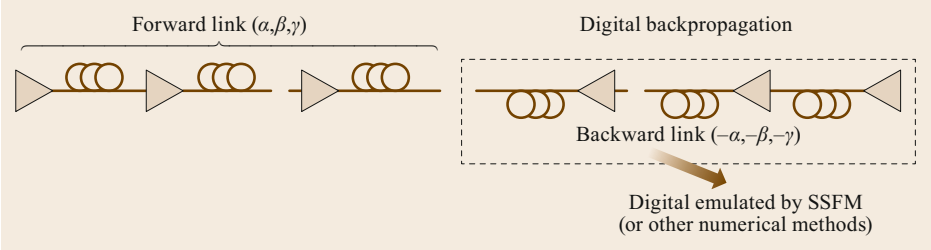


Fig. 9.19 The DBP strategy

below) and must be set to zero for a standard SSFM implementation; N_m is of the order of channel memory induced by GVD (Sect. 9.2), and N is optimized to minimize the overall computational complexity. For comparison, Table 9.2 shows the same figures also for GVD compensation based on a frequency-domain feedforward equalizer.

Each DBP step typically requires about 10–20% more operations than GVD compensation, such that the overall complexity of conventional DBP (based on standard SSFM and employing one step per span) easily exceeds a factor ×10 the complexity of GVD compensation in terrestrial links [9.190].

Attaining better improvements with lower complexity is one of the big challenges to bring DBP from the lab to the field. A first solution is filtered DBP [9.117, 192, 193], in which the nonlinear step uses a low-pass filtered version of the signal power. The nonlinear step (9.5) is, thus, replaced by

$$A_{k,i} = A'_{k,i} \mathrm{e}^{-\alpha_\mathrm{p} h/2} \times \mathrm{e}^{-\mathrm{j}\gamma h_\mathrm{eff}[c_0|A'_{k,i}|^2 + \sum_{\ell=1}^{N_\mathrm{c}} c_i(|A'_{k-\ell,i}|^2 + |A'_{k+\ell,i}|^2)]}, \quad i = 1, \ldots, N, \tag{9.106}$$

where $c_0, \ldots, c_{N_\mathrm{c}}$ are $N_\mathrm{c}+1$ real coefficients of a tapped-delay-line low-pass filter. For $N_\mathrm{c} = 0$, the algorithm reduces to the standard SSFM. The idea is to overcome the inability of the standard SSFM to account for the continuously varying phase mismatch induced by GVD along each nonlinear propagation step—a problem that is particularly significant at large frequencies. Filtered DBP attempts to mitigate such a problem by low-pass filtering signal power during the nonlinear step in a way to limit the intensity fluctuations induced by the artificial phase matching condition of GVD [9.192]. Few taps are generally required in the additional filtering operation, such that the additional complexity per step (shown in Table 9.2) is usually more than compensated for by the step elongation, with about one order of magnitude reduction of the overall complexity. A similar strategy is the one based on the enhanced SSFM algorithm, in which the nonlinear step is still replaced by (9.106), with the difference that the expression is derived from a perturbation analysis based on the FRLP model (Sect. 9.7), and the $N_\mathrm{c}+1$ coefficients are optimized by a minimum mean square error criterion [9.194]. A single-step DBP achieving the same performance as a standard DBP with a 16 times higher complexity has been experimentally demonstrated [9.191].

Alternative DBP design methods showed that finely positioning the nonlinear step in a symmetric-SSFM implementation of DBP can enhance the algorithm performance [9.195]. *Gonçalves* et al. [9.196] showed that DBP can be aided by a memory polynomial model, a technique used in wireless communications to relax the requirements of the Volterra equation. In dispersion-managed links, the strong correlation of the nonlinear interference from span to span can be efficiently exploited to build a folded DBP, where many spans are folded into a single span [9.197].

Low-complexity intrachannel nonlinearity mitigation techniques based on the perturbation analysis in [9.22] have also been proposed. In the special case of QPSK symbols, *Tao* et al. [9.198] showed how to exploit the constant modulus property of the constellation to build a multiplier-free intrachannel Volterra-based equalizer by substituting multiplications with summations. *Tao* et al. [9.199] proposed reducing the number of multiplications in the perturbation evaluation by quantization of the kernel function. By using 50% pre-

Table 9.2 Number of real operations required by GVD compensation and SSFM-based DBP (per step)—$N_\mathrm{c} = 0$ for standard SSFM and $N_\mathrm{c} > 0$ for filtered or enhanced SSFM with N_c coefficients

	GVD compensation	DBP (per step)
Additions	$\frac{N}{N-N_\mathrm{m}}(8\log_2 N + 8)$	$\frac{N}{N-N_\mathrm{m}}(8\log_2 N + 21 + N_\mathrm{c})$
Multiplications	$\frac{N}{N-N_\mathrm{m}}(8\log_2 N + 4)$	$\frac{N}{N-N_\mathrm{m}}(8\log_2 N + 11 + 2N_\mathrm{c})$

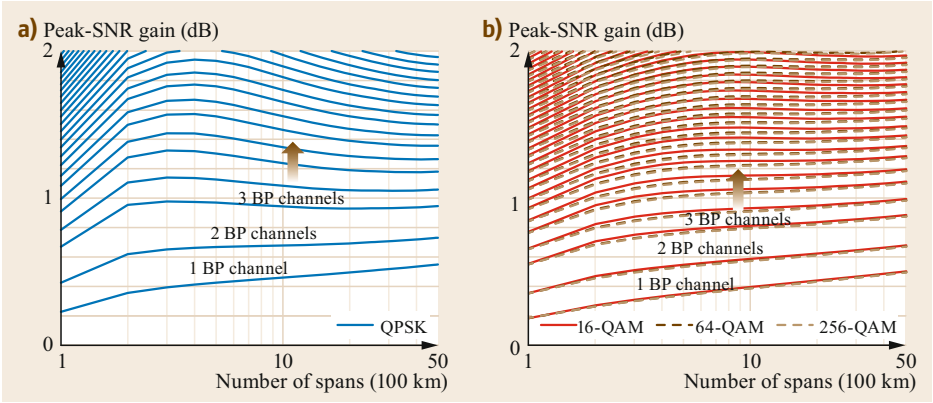


Fig. 9.20a,b Peak-SNR gains achievable by single and multichannel DBP over a fully-loaded (115 channels) dispersion-unmanaged WDM system over SMF, with a channel symbol rate of 32 GBaud. The system parameters are set as in [9.78]

compensation, the number of quantized coefficients has been shown to be reduced down to few units, still keeping significant gains [9.200]. *Gao* et al. [9.201] showed that with 50% precompensation, the symmetry of the dispersion map with respect to the midpoint of transmission link can be used to reduce the number of multiplications in a perturbation-based nonlinear precompensation. The reason is that, this way the kernel coefficients are i) real valued, thus halving the number of complex multiplications and ii) evaluated over half link length. A reduction factor of 6.8 has been experimentally demonstrated. Although the kernel description of the perturbation is a global input/output description of the link, it has been shown that implementing the perturbative nonlinear equalization by a multistage compensation cascade may relax complexity at a given accuracy [9.202]. Indeed, when dividing the link into N stages, the equalization procedure is repeated N times, once per stage. Nevertheless, since the total number of nonlinear interactions per stage decreases *quadratically* with N due to the shorter memory of each stage, the overall complexity is reduced. A comprehensive analysis of the symmetries in Volterra-based nonlinear equalization that help practical implementation was investigated in [9.203]. An analytical expression for the computation of the coefficients of perturbative nonlinear precompensation for Nyquist pulses was derived and experimentally verified in [9.204]. The gains achievable by DBP or perturbative nonlinear compensation were investigated and compared in a C+L band ultra-longhaul transmission experiment in [9.205].

Besides computational issues, and despite the theoretical possibility of an exact channel inversion, both fundamental and practical limitations prevent DBP from completely removing nonlinear impairments. A first limitation arises in WDM systems, which are impaired by both intra and interchannel nonlinearity. In typical systems, each channel is separately detected and processed, such that only a *single-channel DBP* can be actually implemented, effectively compensating only for intrachannel nonlinearity. In principle, it is possible to implement a *multichannel DBP* by jointly detecting and processing two or more WDM channels [9.117, 189]. As the number of backpropagated channel increases, a higher portion of interchannel nonlinearity can be effectively mitigated. However, the required computational resources grow more than linearly with the number of backpropagated channels.

The peak-SNR gains that can be theoretically achieved by ideal single and multichannel DBP over the typical dispersion-unmanaged WDM systems are shown in Fig. 9.20 as a function of the number of spans [9.78], for systems operating at 32 GBaud.

The various solid curves correspond to the number of channels that are jointly backpropagated. The potential gain of single and three-channel DBP is limited to about 0.5 and 1 dB, respectively. Beyond three-channel DBP, each additional backpropagated channel provides an incremental gain of no more than ≈ 0.1 dB, possibly not enough to justify the corresponding increase of the computational resources required. Slightly higher gains are achievable when considering systems with interchannel phase noise and polarization rotation compensation [9.78]. Moreover, in optically routed networks, signals at different wavelengths may follow different paths in the network, such that an effective multichannel DBP is no longer possible.

The possibility to obtain a perfect channel inversion may be limited also by the presence of PMD [9.137]. In principle, an exact knowledge of the PMD evolution both in time and along the fiber link would still allow for a perfect channel inversion without any performance degradation. In practice, such a knowledge is unavail-

able both at the transmitter and receiver, and DBP is usually performed by ignoring the presence of PMD (which is separately compensated for by an adaptive linear equalizer after DBP). In typical WDM systems, the impact of PMD is negligible for single-channel DBP, but becomes relevant when increasing the number of backpropagated channels, practically limiting the SNR gains achievable [9.189]. Some possible countermeasures, based on the insertion of PMD sections among the DBP steps, were investigated in [9.206–208], considering different strategies to configure the PMD sections.

Eventually, the nonlinear compensation achievable by DBP is fundamentally limited by signal–noise interaction, for which a channel inversion strategy turns out to be ineffective. This is easily understood by noting that the ASE noise injected by each in-line amplifier interacts with the propagating signal along a different portion of the optical link, such that DBP reduces the signal–noise interaction due to the first amplifiers, but enhances that due to the last amplifiers. The impact of signal–noise interaction on DBP is generally negligible compared to that of interchannel nonlinearity and of PMD. It is, therefore, of little practical relevance in real systems. However, it becomes essential when investigating the ultimate limitations of optical fiber systems—for instance, when considering single-channel scenarios or, equivalently, in the case of an ideal multichannel DBP covering the whole WDM spectrum. This kind of analysis is complicated by the unavailability of an exact model of the optical fiber channel accounting for signal–noise interaction, though some approximated models, based on perturbation theory, are, indeed, available, e.g., [9.43, 95, 114–116, 121, 209], which would be required to optimize the detection strategy, as discussed more in detail in Sect. 9.12. In fact, DBP is not sufficient to implement an optimum detector, and alternative strategies have been proposed to replace or complement DBP [9.210–212]. One possible approach is that of extending the DBP concept to include the uncertainty due to amplifier ASE noise when backpropagating the received signal. This procedure, known as stochastic DBP (SDBP), consists in backpropagating distributions rather than signals, eventually obtaining an estimate of the posterior distribution of the transmitted symbols given the received signal, which is required to implement an optimum detector [9.211]. The specific SDBP algorithm proposed in [9.211] is based on a particle representation of the distributions. It is too complex for a practical implementation but offers a benchmark for simpler implementations.

Most DBP demonstrations assume that the required system parameters (e.g., launch powers and fiber parameters for each span of a fiber) are known at the receiver. Often, this is not the case. Moreover, some parameters may change over time due, for instance, to network reconfigurations. It is, therefore, essential to devise adaptive strategies for the configuration and dynamic control of DBP parameters [9.213, 214]. While there are no particularly stringent requirements in terms of convergence and adaptation speed for standard DBP algorithms, this issue becomes critical when considering PMD-aware DBP algorithms.

9.11.2 Symbol-Rate Optimization

One of the features of nonlinear fiber propagation is that the amount of generated NLI power depends on the symbol rate of the WDM channels. Specifically, it was shown simulatively [9.215, 216] and then experimentally [9.217] that, by distributing the same total data throughput across the same optical bandwidth, over a *larger* number of WDM channels at a *smaller* symbol rate, the NLI power impairing reception would *decrease*. Initial theoretical studies concurred [9.132, 218]. These findings were then investigated by means of the EGN model [9.219]. Further investigation was carried out in [9.78], using both time-domain and the pulse-collision models, which also clarified the role and impact of NLPN in this phenomenon. Interestingly, SRO was predicted by the EGN model results to actually slightly increase its effectiveness when the overall system WDM bandwidth is widened [9.219], contrary, for instance, to DBP. Due to all these findings, the idea of reducing the symbol rate per carrier in order to improve system performance has been considered as a potential NLI mitigation technique, under the acronym SRO (symbol-rate optimization).

SRO theory predicts that the optimum symbol rate is, in uniform links of N_s spans, each of length L_s [9.219]

$$R_{\mathrm{opt}} = \sqrt{\frac{2}{\pi |\beta_2| L_s N_s}} \,. \tag{9.107}$$

Values typically range between 1 and 10 GBaud. These rates are quite low as compared to industry standards. In addition, the industrial trend, for techno-economic reasons, has historically been that of constantly ramping up symbol rates, with a transition between 32 to 64 GBaud currently taking place. This circumstance clearly clashes with the SRO prescription of operating at low symbol rates.

One way to make an increase in *total* symbol rate per transponder compatible with a decrease of symbol rate *per carrier* is to resort to *subcarrier multiplex-*

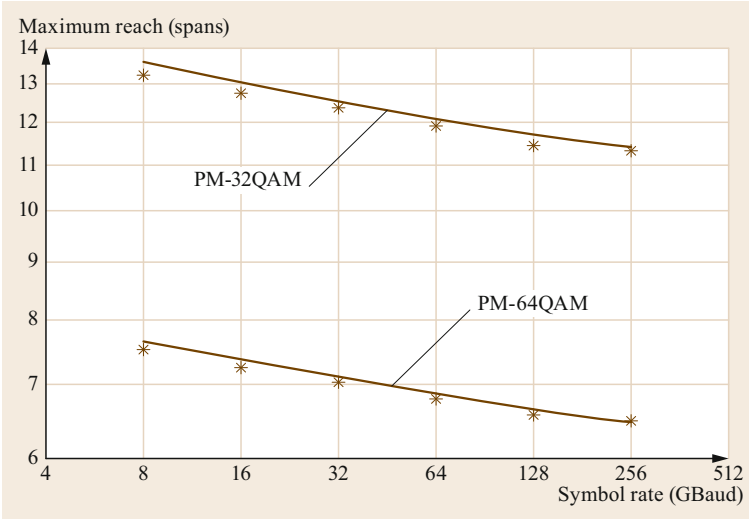


Fig. 9.21 Maximum reach versus per-channel symbol rate. *Solid lines*: calculations using the EGN model. *Markers*: simulations. Span length 100 km, SMF, EDFA noise figure 6 dB, roll-off 0.2. Target performance: GMI 8.7 and 10.44 bits/symb, for PM-32QAM and PM-64QAM, respectively. The total optical bandwidth is 1.83 THz for all rates. For symbol rates from 8 to 256 GBaud, the number of channels ranges from 192 to 6. Carrier spacing: $\Delta f = 76.2(R_s/64)$, where R_s is the symbol rate in GHz

ing. Essentially, high symbol-rate channels are created as a collection of DAC-generated electrical subcarriers, which operate at the symbol rate that is optimum from the viewpoint of NLI mitigation. For instance, a 64 GBaud-channel could consist of 16 4 GBaud subcarriers.

No closed-form formula currently exists for the prediction of the potential MR gain. Estimating it requires the numerical evaluation of the EGN model. Incidentally, SRO is one of those contexts where the GN model is inadequate, as suggested by the strong violation of criterion (1), which prescribes a high symbol rate for GN model accuracy (Sect. 9.4.7). Direct EGN (or time-domain model) calculations or Monte-Carlo simulations are required.

Figure 9.21 provides a pictorial appreciation of the predicted MR gain for two high-cardinality systems, either PM-64QAM or PM-32QAM. The relevant system data is reported in the figure caption. Both the solid line (EGN model) and the markers (simulations) concur in predicting a 16–18% MR increase between the figure extremes, i.e., from 256 GBaud down to 8 GBaud. The latter is close to the theoretical optimum, which according to (9.107) is about 5 GBaud for PM-32QAM and about 6.5 GBaud for PM-64QAM. Note that in order to clearly show an optimum, Fig. 9.21 would have to be pushed down to about 2 GBaud. This was not possible because the CPU time of the simulations, as well as the their memory occupation, is inversely proportional to the symbol rate. For the almost 2 THz total optical bandwidth considered in Fig. 9.21, 8 GBaud was the practical limit. It was done instead in [9.219], Figs. 9.1 and 9.4, because the total bandwidth was 500 GHz. There, the optimum is clearly visible.

The MR percentage gain due to SRO is actually larger for longer links and for smaller constellations [9.78, 219], with PM-QPSK being the format benefiting the most, with potential MR gains on the order of 25–30% across the same symbol rate range as in Fig. 9.21. Also, there are important aspects as to the role played by NLPN in the context of SRO. For more details on this [9.78].

In addition, when GSCs are used, the gain due to SRO essentially vanishes, as predicted by the GN model (which is equivalent to the EGN model for GCSs). Part or all of the gain could be gained back for GSCs if NLPN mitigation was possible at low symbol rates. This is currently unclear and the subject of ongoing research [9.78].

Despite all the limitations, SRO could be a possible way to partially mitigate NLI, at least for QAM-based systems, with relatively low complexity. In [9.219], it was shown by EGN-model calculations that for full C-band ultra-long-haul systems, at 32 GBaud, its MR gain could be comparable to DBP.

It should, however, be mentioned that negative findings regarding SRO have also been published. In particular, in [9.220] very limited gain from SRO was found, which also tended to vanish as the number of WDM channels was increased, in contrast with SRO theory. Overall, a conclusive specific experiment proving the effectiveness of SRO over a fully populated C-band system is not yet available. Therefore, SRO should still be considered a research topic needing confirmation. We should also mention that presently there are commercial systems that use multisubcarrier transmission. Whether its use grants actual performance gain in these commercial systems, and why, is currently undisclosed information.

9.11.3 The Nonlinear Fourier Transform

The nonlinear Fourier transform (NFT)—also known in the mathematical community as the inverse scattering transform—is a mathematical tool to solve a certain class of nonlinear partial differential equations—the so-called *integrable* equations. The NFT was introduced in [9.221] for the Kortweg–de Vries equation and later applied to many other equations, including the NLSE [9.222] and the Manakov equation [9.223]. A general treatise on the NFT can be found in [9.224], while a more specific review of the NFT for the NLSE and its application to optical fiber communications can be found in [9.225, 226].

The NFT can be regarded as the generalization of the ordinary Fourier transform to nonlinear systems [9.224]. The basic idea is that of representing the propagating waveform through a proper set of spectral data—the scattering data or *nonlinear spectrum*—whose evolution along the fiber link is simple and linear. As with linear systems and the ordinary Fourier transform, the initial value problem is, therefore, solved by performing three main steps:

1. Computation of the input nonlinear spectrum by a direct NFT. For the NLSE, this operation consists in solving the Zakharov–Shabat spectral problem [9.222]. In general, the Zakharov–Shabat operator has both a discrete and a continuous spectrum. The former consists of a set of eigenvalues in the complex plane and the related norming constants and is associated with *soliton* components. The latter is defined over the real line and is an analogue to the ordinary Fourier transform, to which it converges in the low-power limit.
2. Propagation of the nonlinear spectrum. For the NLSE, this is equivalent to the propagation of the linear spectrum in a linear dispersive fiber.
3. Reconstruction of the output waveform from its nonlinear spectrum by an inverse NFT (INFT). For the NLSE, this is classically done by solving the Gelfand–Levitan–Marchenko integral equation [9.222, 224].

Many integrable equations, including the NLSE, have *soliton* solutions, which play an important role in NFT theory. Solitons maintain their shapes (or return periodically to them) during propagation due to a cancellation of nonlinear and dispersive effects. This property makes them attractive for communications. Indeed, soliton-based communications were extensively explored during the 1980s and 1990s ([9.227] and references therein). However, they never made their way to commercial systems because of their low spectral efficiency (about 0.2 (bits/s)/Hz) and the detrimental impact of the Gordon–Haus effect and intersoliton interaction.

Soliton communication is just the simplest way in which NFT-related concepts can be used to design nonlinearity-tolerant communication systems. A more general approach, originally named *eigenvalue communication* [9.228], is that of avoiding nonlinear interference by encoding information on the spectral data of the Zakharov–Shabat operator. After going almost unnoticed for a couple of decades, this approach was recently resurrected—with different names and flavors, such as *eigenvalue communication* [9.229], *nonlinear frequency-division multiplexing (NFDM)* [9.225], and *nonlinear inverse synthesis* [9.230]—due to the impressive progress in DSP technology, which makes conceivable, if not yet feasible, a realtime implementation of the NFT.

All the NFT-based approaches proposed so far can be generally represented by one of the four schemes depicted in Fig. 9.22 (the first three are also discussed in [9.226]). In the standard NFDM scheme of Fig. 9.22a, information is encoded and decoded in the nonlinear frequency domain, using INFT and NFT to convert the nonlinear spectrum into a time-domain waveform and the other way around, respectively [9.225, 228, 230–232]. On the other hand, in the scheme of Fig. 9.22b, information is encoded and decoded in the time domain, as in conventional systems, while NFT and INFT are used to implement DBP [9.233]. Figure 9.22c represents a hybrid approach in which information is encoded in the time domain and decoded in the nonlinear frequency domain [9.229, 234]. Finally, Fig. 9.22d shows the complementary hybrid approach in which information is encoded in the nonlinear frequency domain and decoded in the time domain [9.235, 236]. While the first two approaches achieve a diagonalization of the channel, such that symbol-by-symbol detection can be employed, the last two do not, in fact, they require some kind of sequence detection. Quite interestingly, time-domain detection of NFDM signals (the fourth approach) seems to achieve a superior performance compared to the conventional NFDM scheme (the first approach), though with some additional complexity, probably due to a better matching between the detection metrics employed and the actual channel statistics [9.236].

NFDM schemes can be further classified according to which part of the nonlinear spectrum is used to encode information: the discrete part [9.225, 228, 229, 235], the continuous part [9.225, 230, 232, 236], or both [9.237]. Furthermore, NFT theory can also be applied to the Manakov equation, such that the proposed

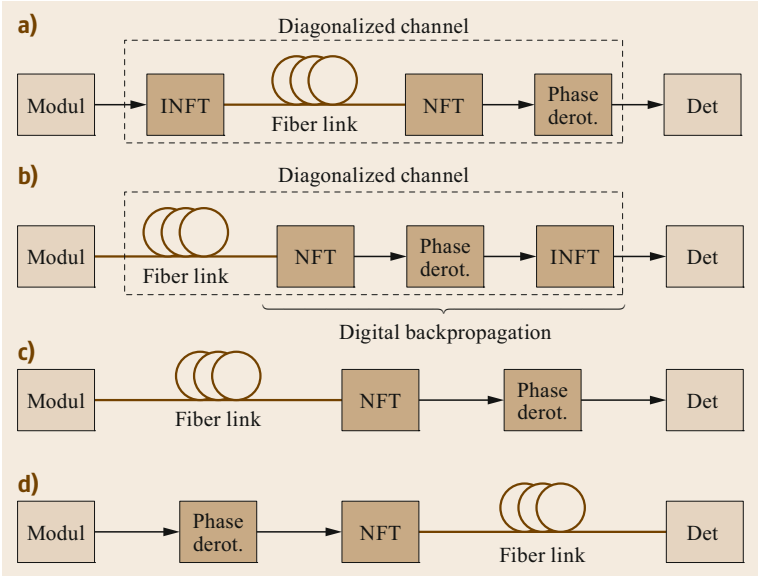


Fig. 9.22a–d Different ways to use NFT in optical communications. **(a)** Modulation and detection in the nonlinear frequency domain; **(b)** modulation and detection in the time domain, with NFT-based DBP; **(c)** modulation in the time domain and detection in the nonlinear frequency domain; **(d)** modulation in the nonlinear frequency domain and detection in the time domain

schemes can be extended to consider polarization-multiplexed signals [9.238, 239].

As for the ordinary Fourier transform, the NFT and the INFT can be computed analytically only in a few simple cases. In general, one has to resort to numerical methods. An overview of the numerical methods available can be found in [9.225, 226]. While the complexity of the most classical methods usually scales at least quadratically with the number of processed samples, super fast algorithms with log-linear complexity (such as the FFT) have been also developed [9.240]. Research of low-complexity methods with general validity is still ongoing and will play an important role in determining the feasibility of NFT-based techniques.

The NFT theory and the techniques mentioned in this section are based on *vanishing boundary conditions*, i.e., they assume that the optical signal decays sufficiently fast as $t \to \pm\infty$. In a real transmission system, this condition can be emulated by operating in burst mode and inserting a sufficiently long guard time between bursts [9.230]. The NFT theory has also been developed under different boundary conditions. A typical case, which has been extensively studied in the literature, is that of *periodic boundary conditions*. This approach may offer some advantages compared to the case of vanishing boundary conditions [9.226], but its application to optical communications has been much less explored [9.241].

Research on NFT-based techniques is still at an early stage, and there is no general consensus about their potentials, perspectives, and suitability for optical communications. Research mostly follows a trial-and-error approach, and the lack of a unified theory makes it hard to understand which strategy (among those depicted in Fig. 9.22) is the most promising, which part of the spectrum should be modulated, how information should be encoded on it, and what kind of boundary conditions should be considered. Several experimental demonstrations of the schemes in Fig. 9.22 can be found. Yet, so far, there is no clear evidence of performance improvement compared to conventional systems. In fact, some critical issues that seriously hamper the use of NFT operations at high power [9.242] and the achievement of high spectral efficiencies [9.232] have been highlighted. Performance needs to be significantly improved and complexity reduced before NFT-based techniques can make their way to real systems. Moreover, some propagation effects, which are not included in the NLSE (9.1) but can be easily described and coped with in the time domain (e.g., attenuation, higher-order dispersion, PMD), break the integrability of the system and are extremely hard to model in the nonlinear frequency domain. Their impact on NFDM systems needs to be better investigated, as it might be potentially more detrimental than the impact of nonlinear effects on conventional systems. Yet, the promise of a tool able to harness fiber nonlinearity keeps research alive.

9.11.4 Other Nonlinear Compensation Techniques

While DBP is certainly one of the most studied nonlinear compensation techniques, other techniques have been also proposed in the literature in recent years, ranging from the quite practical and simple approach of SRO to the more visionary techniques based on the

NFT, both of which are reviewed in this section. Nonlinearity compensation based on the inverse Volterra series transfer function [9.203] has been also studied. Given the equivalence between the RP method and the Volterra series approach, discussed in Sect. 9.3, this technique can be included in the broader category of perturbation-based channel inversion techniques for alternative DBP implementations.

According to the time-domain perturbative model and the FRLP model, which were introduced in Sects. 9.5 and 9.7, respectively, interchannel nonlinearity causes the emergence of phase noise, polarization rotations, and ISI. These effects can be partly mitigated by some classical algorithms typically employed to counteract them, when generated by causes other than fiber nonlinearity. These algorithms can, hence, in all respects be considered as nonlinearity mitigation strategies. Some examples are reported in [9.82, 83, 85, 243].

Another class of nonlinearity mitigation techniques aims at reducing the amount of NLI generated by a proper combination of coding and modulation. Possible examples are the design of nearly constant-envelope modulation formats [9.244], the ad-hoc probabilistic shaping of QAM constellations over one [9.245] or several time slots [9.246], and the use of twin waves [9.247] or conjugate data repetition [9.248] to partly cancel NLI.

9.12 Capacity of the Nonlinear Channel

One of the key results of *Claude Shannon's* pioneering work was the demonstration that a reliable communication over a noisy channel is possible, provided that the information rate is less than a characteristic quantity determined by the statistical properties of the channel, which he named *channel capacity* [9.249].

9.12.1 Capacity and Spectral Efficiency

Other than laying the foundations of what is now called *information theory* and establishing such a general result, Shannon also derived a specific closed-form expression for the capacity of an AWGN channel. As we shall see, this expression is also widely employed in the context of optical fibers, though its validity and interpretation need some care. Shannon's work has been since extended and generalized both to account for a broader class of channels and to obtain closed-form capacity expressions for other specific channels ([9.250] and references therein). In this context, the optical-fiber channel is a major challenge due to its peculiar nonlinear behavior. In fact, as we shall see in this section, research to determine channel capacity is still ongoing, and the problem, so far, remains essentially open [9.251].

We start by considering the generic digital communication system schematically depicted in Fig. 9.23. Firstly, the channel encoder takes a stream of information bits from the source and adds some redundancy to it (e.g., parity check bits) to be used for error correcting purposes. Then, the encoded bits are mapped onto a sequence of information symbols belonging to the input modulation alphabet $\mathcal{X}$. Since the physical channel is usually a *waveform channel* (e.g., an optical fiber link), a DAC converts the sequence of symbols into a waveform (e.g., by linearly modulating a sequence of pulses of given shape) that propagates through the channel. The opposite operations are then performed at the receiver; the output waveform (which is possibly distorted and corrupted by noise) is converted to a sequence of symbols belonging to the output alphabet $\mathcal{Y}$ (e.g., by filtering and sampling the waveform). The detector converts the output symbols into a sequence of bits, which are finally decoded (correcting possible errors) and sent to the destination. The combination of DAC, waveform channel, and ADC constitutes the *discrete-time channel*, which is usually the subject of the information theoretical analysis. In practical cases, the input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ are determined by the quantization levels available in the DAC and ADC and are, hence, finite. Often, when no constraints about the DAC and ADC characteristics are given, the whole field of complex numbers is considered for both alphabets. The discrete-time channel is characterized by a family of conditional distributions $p(\boldsymbol{y}_N|\boldsymbol{x}_N)$, for $N = 1, 2, \ldots$, where $\boldsymbol{x}_N = (x_1, x_2, \ldots, x_N)$, with $x_i \in \mathcal{X}$, and $\boldsymbol{y}_N = (y_1, y_2, \ldots, y_N)$, with $y_i \in \mathcal{Y}$, are length-N realizations of the input and output processes $X = X_1, X_2, \ldots$, and $Y = Y_1, Y_2, \ldots$, respectively.

A typical problem in information theory is that of determining channel capacity—operationally defined as the maximum bit rate at which information can be reliably (i.e., at an arbitrarily low error probability) transmitted through the channel with the best possible combination of coding and modulation—without actually trying such (infinitely many) possible combinations. Given the statistics of the input process, i.e., the input distributions $p(\boldsymbol{x}_N)$ for $N = 1, 2, \ldots$, the maximum bit rate at which information can be reliably transmitted through the channel is given by the infor-

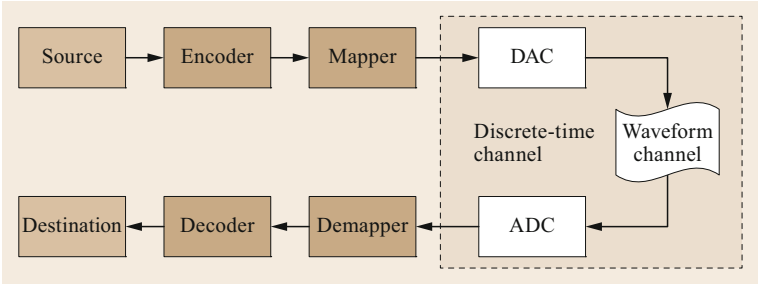


Fig. 9.23 A digital communication system

mation rate [9.252]

$$I(X;Y) = \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{X}_N;\boldsymbol{Y}_N)\,, \tag{9.108}$$

where

$$I(\boldsymbol{X}_N;\boldsymbol{Y}_N) = E\left\{\log_2 \frac{p(\boldsymbol{Y}_N|\boldsymbol{X}_N)}{p(\boldsymbol{Y}_N)}\right\} \tag{9.109}$$

is the mutual information between $\boldsymbol{X}_N$ and $\boldsymbol{Y}_N$, $E\{\cdot\}$ is the expectation operator. The information rate (9.108) depends both on the input statistics—which are determined by the channel user by selecting a combination of code and modulation—and on the channel itself. Eventually, the capacity of the channel is obtained by selecting the input statistics for which the information rate is maximized

$$C = \lim_{N\to\infty} \frac{1}{N} \sup_{\boldsymbol{X}_N} I(\boldsymbol{X}_N;\boldsymbol{Y}_N)\,, \tag{9.110}$$

where the supremum $\sup_{\boldsymbol{X}_N}$ is taken with respect to all input distributions $p(\boldsymbol{x}_N)$ satisfying a specific constraint (usually on the average power) [9.252]. The capacity is, hence, a specific property of the channel, which determines a fundamental limit for transferring information.

Both quantities in (9.108) and (9.110) refer to the discrete-time channel and are measured in bit/symbol (or bit/channel use). The corresponding information rate and capacity of the waveform channel are expressed in bit/s and are obtained by multiplying (9.108) and (9.110) times the baud rate R at which symbols are mapped to and demapped from the input and output waveforms. A further normalization by the *occupied bandwidth* (whenever a suitable definition of it is available) yields the *spectral efficiency* (SE) in (bit/s)/Hz.

In general, only the capacity of some specific channels can be evaluated analytically, the AWGN channel perhaps being the most notable example [9.249]. A discrete-time AWGN channel is characterized by the input–output relation $y_k = x_k + n_k$, where the noise samples n_k are realizations of i.i.d. variables with a zero-mean proper complex Gaussian distribution [9.253]—often referred to as circularly symmetric complex Gaussian (CSCG)—with variance P_n. The analysis in this case is greatly simplified by the absence of memory, which allows us to compute (9.109) by considering only a single pair of input and output symbols ($N = 1$). Given a constraint on the average signal power P_s, the capacity of the AWGN channel is given by the well-known expression $C = \log_2(1 + P_\mathrm{s}/P_\mathrm{n})$ and is achievable by i.i.d. CSCG input symbols [9.249]. As the capacity achieving input distribution and the corresponding output distribution are zero-mean, this capacity can also be expressed in terms of the input variance $\sigma_x^2 = P_\mathrm{s}$, output variance $\sigma_y^2 = P_\mathrm{s} + P_\mathrm{n}$, and covariance $\sigma_{xy} = P_\mathrm{s}$ as

$$C = \log_2\left(1 + \frac{P_\mathrm{s}}{P_\mathrm{n}}\right) = \log_2\left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2\sigma_y^2 - |\sigma_{xy}|^2}\right). \tag{9.111}$$

In this case, it is apparent that, since the noise power P_n is fixed and independent of the signal, the channel capacity grows unbounded with signal power P_s. This means that the channel capacity is limited only if the available resources (power) are limited (e.g., due to economic or technological constraints).

When turning our attention to the optical fiber channel, the picture is more complicated. At low signal power, the nonlinear term of the NLSE (9.1) is practically negligible, and the channel essentially behaves like an AWGN channel with a capacity that grows logarithmically with launch power according to (9.111). (Linear propagation impairments, such as chromatic dispersion and polarization mode dispersion, are characterized by a slowly varying (or constant) unitary transfer matrix. Hence, in the linear regime, their presence has no impact on channel capacity as they can be fully compensated by a linear equalizer.) However, at higher signal power, the nonlinear term of the NLSE is no longer negligible and must be accounted for in the evaluation of channel capacity. This is an extremely difficult task. First of all, an exact and mathemati-

cally tractable expression for $p(\boldsymbol{y}_N|\boldsymbol{x}_N)$ is unavailable, making the analytical evaluation of (9.108) unfeasible. Moreover, its numerical estimation, which must be performed in a $4N$-dimensional space, is practically limited to very small values of N. On the other hand, the convergence of (9.108) and (9.110) to their actual limit can be expected when N is of the order of the channel memory, which can be many hundreds of symbols in typical dispersion-unmanaged systems. For the same reasons, the optimization of $p(\boldsymbol{x}_N)$ in (9.110) cannot be performed analytically either, and, when resorting to numerical approaches, it is limited to small values of N. A further issue arises when considering the discretization of the waveform channel. In fact, due to nonlinearity, the propagating signal is subject to spectral broadening and signal–noise interaction during propagation. Therefore, the structure and bandwidth of an optimum demodulator (providing a sufficient statistic) is also unknown. Moreover, since the signal bandwidth changes during propagation, there is no general consensus on how the spectral efficiency should be defined [9.254].

Typical optical systems employ WDM, in which many transmitter/receiver pairs share the same fiber to transmit and receive signals at different wavelengths. This multi-user scenario requires some extra care when being modeled and analyzed. A typical and reasonable assumption is that each user has access only to its allotted portion of the WDM spectrum, and that all the signals are independently generated and detected. In this situation, the capacity analysis is performed from a single-user perspective, i.e., by focusing on a single WDM channel (user), typically referred to as the COI, and considering all the other channels as a source of possible interference. Moreover, a specific behavioral model for the interfering channels, i.e., a rule that relates the input distribution on the interfering channels to the COI input distribution, should be specified [9.255]. The most common (and fair) assumption is that all the users transmit with same input distribution and power. Different choices are possible, which correspond to different network scenarios and lead to completely different results in terms of capacity [9.255]. Eventually, in the WDM scenario, spectral efficiency is usually defined by considering that the bandwidth occupied by each channel equals the channel spacing.

All the difficulties highlighted in this section make the exact evaluation of the capacity of the optical fiber channel extremely difficult. In cases like this, a typical approach is that of resorting to the computation of upper and lower bounds on the capacity, trying to identify a possibly narrow range of values in which the actual (unknown) capacity lies. This will be discussed in the next sections.

9.12.2 Achievable Information Rate with Mismatched Decoding

When the actual channel law is unknown, such as in the case of the optical fiber channel, it is useful to consider the more general case, depicted in Fig. 9.24, of a discrete-time channel with a detector that makes maximum-a-posteriori-probability (MAP) decisions based on a mismatched channel law $q(\boldsymbol{y}_N|\boldsymbol{x}_N) \neq p(\boldsymbol{y}_N|\boldsymbol{x}_N)$. Following [9.147], we define the *achievable information rate* (AIR) with the mismatched decoding metric $q(\boldsymbol{y}_N|\boldsymbol{x}_N)$ as

$$I_q(X;Y) \triangleq \lim_{N\to\infty} \frac{1}{N} E\left[\log \frac{q(\boldsymbol{y}_N|\boldsymbol{x}_N)}{\int p(\boldsymbol{x}_N) q(\boldsymbol{y}_N|\boldsymbol{x}_N) d\boldsymbol{x}_N}\right]. \tag{9.112}$$

(In [9.147], this quantity is actually referred to as the *auxiliary-channel lower bound* to the information rate.) With respect to the average mutual information rate (9.108), (9.112) is obtained by replacing the actual channel law $p(\boldsymbol{y}_N|\boldsymbol{x}_N)$ with an arbitrary mismatched law $q(\boldsymbol{y}_N|\boldsymbol{x}_N)$, while the expectation is still taken with respect to the actual true distribution $p(\boldsymbol{y}_N|\boldsymbol{x}_N)p(\boldsymbol{x}_N)$ induced by the input distribution and the actual channel law.

The AIR (9.112) has some interesting properties that hold for any true and auxiliary channel and make it suitable for practical use in optical communications: i) it is a lower bound to the information rate and, therefore, to the channel capacity $I_q(X;Y) \leq I(X;Y) \leq C$; ii) its maximization over any possible detection law (obtained for $q(\boldsymbol{y}_N|\boldsymbol{x}_N) = p(\boldsymbol{y}_N|\boldsymbol{x}_N)$) leads to the actual information rate; iii) its further maximization over the input distribution $p(\boldsymbol{x}_N)$ leads to channel capacity; iv) it is achievable over the true channel with source probability $p(\boldsymbol{x}_N)$ and a MAP detector matched to $q(\boldsymbol{y}_N|\boldsymbol{x}_N)$; and v) it can be computed through numerical simulations, without an explicit knowledge of the true channel law $p(\boldsymbol{y}_N|\boldsymbol{x}_N)$, provided that $q(\boldsymbol{y}_N|\boldsymbol{x}_N)$ can be computed.

In practice, the system is designed by selecting a modulation format that determines the input distribution $p(\boldsymbol{x}_N)$, and an approximated (auxiliary) channel model (e.g., one of the ones discussed in Sect. 9.3), which determines the mismatched channel law $q(\boldsymbol{y}_N|\boldsymbol{x}_N)$ and, hence, the detection metrics. The

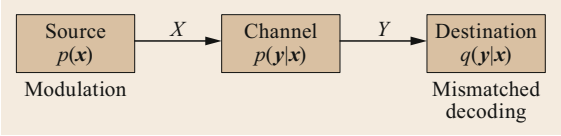


Fig. 9.24 Discrete-time channel with mismatched decoding

AIR (9.112) for this configuration is then computed through numerical simulations. Possibly, $p(\boldsymbol{x}_N)$ and/or $q(\boldsymbol{y}_N|\boldsymbol{x}_N)$ can be numerically optimized by using the AIR as a performance metric to be maximized.

While an exact analytical evaluation of (9.112) is still unfeasible—as the joint distribution $p(\boldsymbol{x}_K, \boldsymbol{y}_K)$ of the true channel, with respect to which the expectation in (9.112) must be computed, is generally unknown—an accurate numerical estimate can be efficiently obtained by relying on the asymptotic equipartition property [9.146] and following the procedure described in [9.147]:

1. Draw a long input sequence $\boldsymbol{x}_N = (x_1, \ldots, x_N)$ of samples from the selected input distribution $p(\boldsymbol{x}_N)$.
2. Compute the corresponding output sequence $\boldsymbol{y}_N = (x_1, \ldots, x_N)$ by using the SSFM to simulate the true channel.
3. Compute the selected conditional distribution $q(\boldsymbol{y}_N|\boldsymbol{x}_N)$.
4. Compute the corresponding output distribution $q(\boldsymbol{y}_N) = \int p(\boldsymbol{x}_N) q(\boldsymbol{y}_N|\boldsymbol{x}_N) \mathrm{d}\boldsymbol{x}_N$.
5. Estimate the AIR as

$$\hat{I}_q(X;Y) = \frac{1}{N} \log \frac{q(\boldsymbol{y}_N|\boldsymbol{x}_N)}{q(\boldsymbol{y}_N)} \,.$$

For finite-state source and auxiliary channel models, the actual computation of $q(\boldsymbol{y}_N|\boldsymbol{x}_N)$ and $q(\boldsymbol{y}_N)$ can be efficiently carried out by the sum-product algorithm on a suitably defined factor graph [9.147]. The procedure can be extended to the more general case of a continuous state space by resorting to particle methods [9.256].

9.12.3 The Nonlinear Shannon Limit (NSL) Lower Bound

As was already mentioned in the previous section, at low signal powers the optical fiber channel behaves like an AWGN channel; in fact, modern optical systems can achieve information rates close to channel capacity (9.111) when working in such a linear regime [9.257]. However, contrarily to the AWGN case, when the signal power increases, the ensuing nonlinearities impair the existing systems to the point that they cease to work. This naturally poses questions about the impact of nonlinearities on the capacity of a fiber-optic channel and the possible existence of an ultimate limit to channel capacity.

A variety of studies and facts hint at the possibility that a finite limit to the fiber-optic channel capacity does exist, so much that a name for it was coined—the *nonlinear Shannon limit* (NSL) [9.257]. To the best of our knowledge, a finite capacity limit for a fiber-optic channel was predicted for the first time in [9.28]. Similar results were later obtained by using different models and either analytical approximations or numerical methods [9.24, 75, 100, 258, 259]. Such a convergence of results has induced the belief that the actual channel capacity is very close to the NSL or, at least, follows the same trend. However, an analysis of the literature reveals that the NSL, as defined in [9.100, 257], but also as computed in many other publications [9.24, 75, 258], is just an instance of the AIR (9.112), computed with specific input distributions (usually i.i.d. symbols with CSCG or uniform-ring distribution) and mismatched channel laws (usually memoryless Gaussian). Therefore, it is only a lower bound to channel capacity, whose tightness is not known a priori.

A special case of (9.112)—referred to as the Gaussian AIR in the sequel—is obtained when considering i.i.d. CSCG input variables and a detector matched to an AWGN channel with the same input–output covariance matrix of the real channel

$$I_{\mathrm{G}}(X;Y) = \log_2 \left(\frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 \sigma_y^2 - |\sigma_{xy}|^2} \right) . \tag{9.113}$$

This result shows that the Gaussian AIR can be computed for *any* channel by simply estimating—either analytically, through an approximated channel model, or numerically, through accurate but time-consuming SSFM simulations—a covariance matrix. Remarkably, (9.113) equals the expression of the AWGN channel capacity (9.111). This means that the same expression can be used (and, in fact, it is commonly used) for any channel, but with different meanings. It gives the true channel capacity for AWGN channels, while it provides only a lower bound (achievable by a conventional detector optimized for the AWGN channel) for generic channels [9.260]. In fact, the information theoretical limits computed in [9.28, 75, 76, 100, 259] were all obtained by using (9.113), though computing the covariance matrix with different approaches. A good agreement between those results can generally be observed.

When considering the GN model, (9.113) reduces to the simple expression

$$I_{\mathrm{G}}(X;Y) = \log_2 \left(1 + \frac{\mathrm{P_s}}{P_{\mathrm{ASE}} + P_{\mathrm{NLI}}} \right) , \tag{9.114}$$

where P_{ASE} is the power of the accumulated ASE noise over the signal bandwidth W, and P_{NLI} is the nonlinear interference power obtained by integrating (9.34). Note that in this case, since the input process is Gaussian by assumption, the EGN model gives exactly the same result, as the correction term vanishes.

An alternative closed-form expression is obtained with the FRLP model in Sect. 9.7. In this case, (9.113) reduces to

$$I_{\mathrm{G}}(X;Y) = \log_2\left(1 + \frac{P_{\mathrm{s}}\mathrm{e}^{-\overline{\sigma_\theta^2}}}{P_{\mathrm{ASE}} + P_{\mathrm{s}}(1-\mathrm{e}^{-\overline{\sigma_\theta^2}})}\right), \tag{9.115}$$

where $\overline{\sigma_\theta^2}$ is the effective variance of the XPM term $\theta(f,t)$ over the signal bandwidth [9.96]. A very good agreement between (9.115) and the numerical computation of (9.113) based on SSFM simulations has been demonstrated [9.96]. A similar expression, with the same functional dependence on signal power P_{s}, can be obtained through Feynman's path integral approach [9.100]. Moreover, when $\overline{\sigma_\theta^2} \ll 1$—i.e., in most cases of practical interest— (9.115) reduces to (9.114), with $P_s\overline{\sigma_\theta^2}$ replacing P_{NLI}. In this case, the scaling of (9.114) and (9.115) with system parameters becomes the same and equals the one obtained in [9.75] with the RP method.

As an example, we compute the Gaussian AIR (9.115) for a single-polarization fully-loaded Nyquist-WDM system (81 channels) with 50 GHz-channel spacing, employing a standard transmission fiber with attenuation, dispersion, and nonlinearity coefficients $\alpha = 0.2\,\mathrm{dB/km}$, $D = 17\,\mathrm{ps/(nm\,km)}$, and $\gamma = 1.27\times10^{-3}\,\mathrm{W^{-1}m^{-1}}$, respectively. We consider both ideal distributed amplification (IDA), with spontaneous emission coefficient $\eta_s = 1$, and the more practical case of lumped amplification (LA) with $\eta_s = 1.6$ (a noise figure of about 5 dB) and different dispersion maps and span lengths. The system includes single-channel ideal DBP to remove intrachannel nonlinearity. The difference between the single and dual-polarization cases is usually small and will be investigated in the next section. Figure 9.25 shows the Gaussian AIR as a function of the signal launch power (per channel) P_{s}, for dispersion-unmanaged IDA links with different lengths. All the curves reach a maximum at the same optimum power ($\approx -10\,\mathrm{dBm}$ per channel). Doubling the distance, the AIR decreases of about 1 bit/symbol.

The more practical case of an LA link is considered in Fig. 9.26, which shows the maximum Gaussian AIR (at optimum launch power) for different link lengths. In particular, Fig. 9.26a considers the case of a dispersion-unmanaged link and shows the dependence of the maximum AIR on the amplifier spacing (span length). Longer spacing means higher attenuation; hence, more ASE noise, higher optimum power, higher nonlinear interference, and, eventually, lower maximum AIR. The limit of zero amplifier spacing corresponds to the case of an IDA link (though with a higher noise figure and, hence, a lower maximum AIR than Fig. 9.25). On the other hand, Fig. 9.26b considers the case of a dispersion-managed link with an amplifier spacing of 100 km and shows the dependence of the maximum AIR on the residual dispersion per span. Inline dispersion compensation turns out to be detrimental in terms of AIR, as it reduces the walk-off between channels causing a coherent accumulation of the nonlinear interference generated in different fiber spans. However, since nonlinear interaction mostly takes place at the beginning of each fiber span, a small residual dispersion per span (about 20–30%) is sufficient to avoid such a coherent accumulation and practically approach the dispersion-uncompensated case.

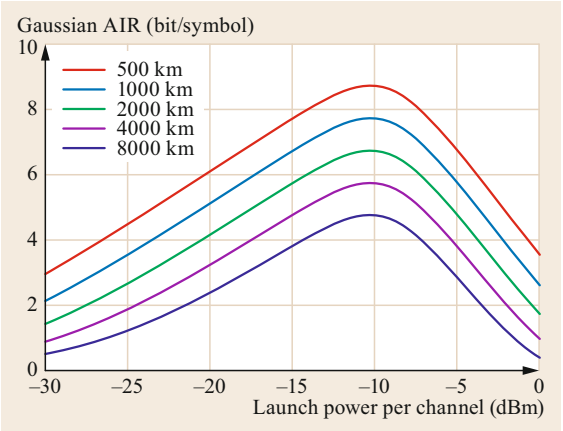


Fig. 9.25 Gaussian AIR for a single-polarization fully-loaded WDM system (81 channels) with 50 GHz channel spacing on a dispersion-unmanaged IDA link

9.12.4 Improved Lower Bounds

The Gaussian AIRs shown in Figs. 9.25 and 9.26 are the capacity lower bounds achievable by a detection strategy that does not account for residual channel memory (after DBP) and for the peculiar characteristics of nonlinear interference. The possible exploitation of the long time coherence of nonlinear interchannel interference (Sect. 9.7) to compute tighter capacity lower bounds was suggested in [9.76] and then demonstrated in [9.261], with further improvements obtained in [9.82, 262].

Figure 9.27 shows the AIR for a dispersion-unmanaged WDM system, considering the same configuration of Fig. 9.25, an IDA link of 1000 km, and different detection strategies. Both the single and dual-polarization cases are shown by solid and dashed lines, respectively. Since these curves are obtained by full SSFM simulations, only five WDM channels are

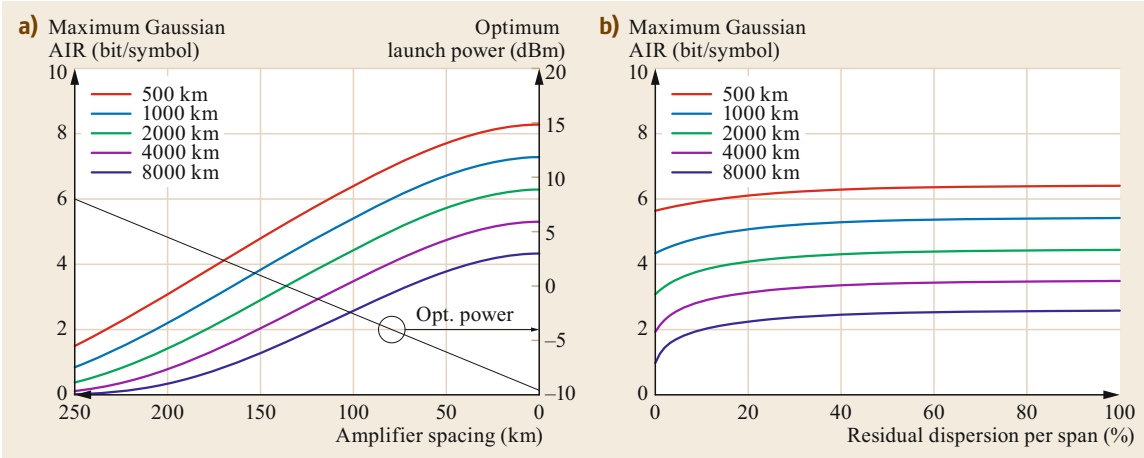


Fig. 9.26a,b Maximum Gaussian AIR for a single-polarization fully-loaded (81 channels) WDM system with 50 GHz channel spacing on an LA link as a function of: **(a)** amplifier spacing (without inline compensation); **(b)** inline residual dispersion (and a fixed amplifier spacing of 100 km)

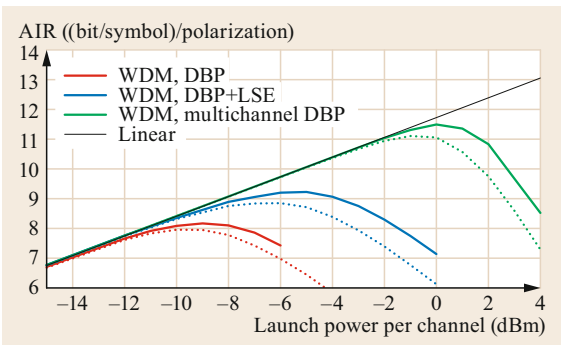


Fig. 9.27 AIR for a single or dual-polarization five-channel WDM system employing different detection strategies

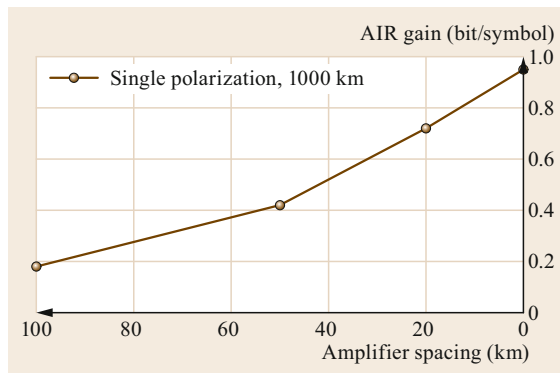


Fig. 9.28 AIR gain as a function of the amplifier spacing for a five channel single-polarization WDM system over a 1000 km link with lumped amplification

considered rather than a fully-loaded system. Before estimating the Gaussian AIR (9.113), different nonlinear mitigation strategies are considered: single-channel DBP as in Fig. 9.25 (the AIR is slightly higher because only five channels are considered), single-channel DBP followed by adaptive least-squares equalization (LSE), and multichannel DBP performed on the whole received WDM spectrum. It turns out that the maximum AIR can be improved by at least 1 bit/symbol by including LSE to mitigate interchannel nonlinearity (the gain increases with the number of WDM channels [9.82]).

As is shown in Sect. 9.7, the XPM term generated by nonlinear interference is highly coherent (both in time and frequency) for IDA links, but much less coherent for links with lumped amplification. Thus, lower AIR gains are expected over the latter. More generally, the coherence decreases as the portion of the link along which nonlinear interaction takes place decreases, such that the AIR gain achievable by LSE (or similar techniques) compared to the Gaussian AIR over links with lumped amplification decreases with amplifier spacing. Figure 9.28 shows the AIR gain (difference between the maximum AIRs obtained with and without LSE) as a function of the amplifier spacing for the same single-polarization system of Fig. 9.27 over a 1000 km link with lumped amplification (but unitary spontaneous emission factor). The limit of zero amplifier spacing corresponds to the case of an IDA link.

9.12.5 Beyond the Nonlinear Shannon Limit

So far, only some capacity *lower bounds* have been presented, all sharing the same typical dependence on optical power—they all reach a maximum at some op-

timum power and then decrease again. We might be tempted to conclude that even the true channel capacity follows the same trend and has a finite maximum. However, as we have already argued, there is no proof of the existence of such a *nonlinear Shannon limit*. In fact, the tightest available *upper bound* is the one proposed in [9.254], which states that the capacity of the optical-fiber channel is upper bounded by $\log_2(1 + \mathrm{SNR})$—the capacity of an AWGN channel with same total accumulated noise. This has important theoretical and practical implications. In fact, it essentially leaves the capacity problem open, as the distance between the available lower and upper bounds diverges at high power, and stimulates the research to build communication systems that operate in the highly nonlinear regime with increased spectral efficiency. Nevertheless, despite significant research efforts, only small improvements with respect to the Gaussian AIR (the conventional NSL) have been demonstrated so far, which does not leave much room for optimism. Yet, some theoretical results suggest that more significant improvements might be possible, and that channel capacity might even increase unbounded with signal power.

One such result is the nondecreasing capacity lower bound [9.263], which states that the capacity of a discrete-time channel cannot decrease with power. This entails that the typical behavior of the capacity lower bounds discussed before—their decrease after some optimum power—is not a characteristic of the true channel capacity and should be more properly ascribed to the use of nonoptimized input distributions. For instance, the capacity bounds obtained with a Gaussian input distribution (e.g., the Gaussian AIR) are very tight at low power but vanish at high power, where, instead, satellite distributions [9.263] or ripple distributions [9.264] can be used to obtain nondecreasing lower bounds. This is a clear indication that, in a strongly nonlinear regime, conventional modulation formats are highly suboptimal, and that significant gains might be expected from the optimization of the input distribution.

Other important results in this sense are the demonstrations of the infinite asymptotic capacity (for $P \to \infty$) of some simplified channels related to the optical-fiber channel. This is the case, for instance, of the zero-dispersion fiber channel [9.265], of the memoryless FWM model [9.255] and of the RP model [9.260]. In all those cases, the Gaussian AIR has a finite maximum, but the (per symbol) capacity grows unbounded with power. This is still very far from a practical scheme that can guarantee such an unlimited capacity over a realistic fiber channel. Moreover, none of the previous results account for spectral broadening induced by fiber nonlinearity, such that a finite spectral efficiency limit can be still expected [9.266].

The asymptotic capacity of the optical fiber channel in the presence of signal–noise interaction is eventually investigated in [9.267], which suggests an alternative approach to address the capacity problem in this case.

9.13 Conclusion

The field of the investigation of the generation, characterization, and assessment of the impact of nonlinear fiber effects has been extremely active over the last 10–15 years. Undoubtedly, remarkably useful results and practical tools have been obtained and put at the community's disposal. Their use has become quite pervasive, both in transmission and in optical networking sectors.

Investigations are still ongoing. What is probably at the forefront of current research is the devising of mitigation techniques to suppress the nonlinear disturbance affecting signals. To this end, the current (and future) results obtained in modeling provide an indispensable tool. The next few years will certainly see further progress in mitigation, also because the DSP power that can be used in transceivers is still ramping up exponentially, so that sophisticated techniques that seemed to be unrealistically complex not long ago, are gradually becoming viable. This might lead to further substantial increase in the performance of optical transmission systems and networks.

9.A Conversion Among SNR, *Q*-Factor, MI and GMI for the AWGN Channel

The BER is related to the *Q*-factor by the following definition

$$Q\text{-factor} \triangleq 20\log_{10}\left[\sqrt{2}\,\mathrm{erfc}^{-1}(2\mathrm{BER})\right] \quad (\mathrm{dB})\,,$$

with erfc^{-1} the inverse of the complementary error function. The main advantage of the *Q*-factor is that it is linearly, or almost linearly, related to SNR with slope 1 in a dB/dB scale for several modulation formats, thus helping rule of thumb design.

The relation between BER and SNR is modulation format dependent. For the AWGN channel with square *M*-QAM, and by neglecting more than one bit error per symbol error, we have

$$\mathrm{BER} \cong \frac{2}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right) \times \mathrm{erfc}\left(\sqrt{\frac{3}{2(M-1)}\mathrm{SNR}}\right), \quad \text{square QAM}\,. \tag{9.116}$$

The formula can be used even for BPSK with $M = 4$ but by multiplying the SNR by 2, since QPSK, aka 4-QAM, has the same minimum symbol distance as BPSK at twice SNR.

In the relevant case of a signal modulated with symbol rate *R* GBaud at 1550 nm and traveling in a *N* span periodic link with EDFAs of gain G_{dB} and a noise figure

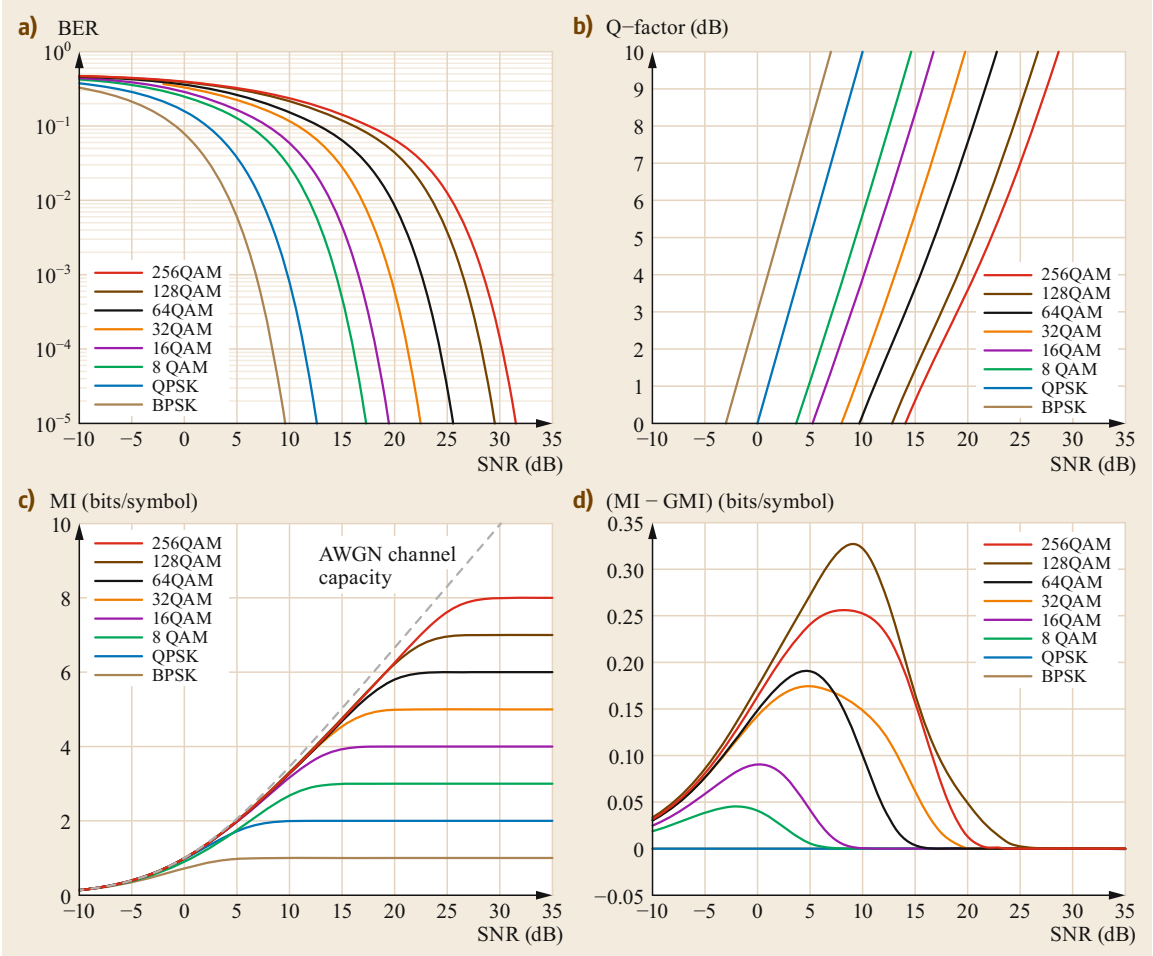


Fig. 9.29a–d Bit error rate (BER), *Q*-factor, mutual information (MI), and generalized mutual information (GMI) for the additive white Gaussian noise channel; 8 QAM is a rectangular QAM, 32 and 128 QAM are cross-QAM

F_{dB}, the SNR in dB is related to signal power P_{dBm} by

$$\mathrm{SNR}_{\mathrm{dB}} = P_{\mathrm{dBm}} + 58 - 10\log_{10} N - G_{\mathrm{dB}} - F_{\mathrm{dB}} - 10\log_{10}\frac{R_{\mathrm{GBaud}}}{12.5}\,.$$

If we neglect the last term accounting for R, we get the optical SNR (OSNR) that could be measured on an optical spectrum analyzer (OSA) over a bandwidth of 12.5 GHz (0.1 nm).

Figure 9.29 shows the relation between SNR, BER, MI, and GMI for the AWGN channel for the most popular modulation formats. We reported the difference between GMI and MI to better highlight the numerical values. The figure for single polarization is: for PDM without polarization crosstalk just multiply the MI and GMI values by 2.

The SNR mismatch between MI and GMI curves is a fraction of dB at practical code overheads. For instance, error-free transmission of single-polarization 64-QAM in AWGN is possible by using FEC with an overhead of 50% at SNR = 9 dB without BICM (i.e., at MI = 3 bits/symbol), or at SNR = 9.44 dB with BICM (i.e., at GMI = 3 bits/symbol). For smaller overhead the difference is smaller.

9.B Detailed List of Symbols, Notation, and Conventions

- z is the longitudinal spatial coordinate, along the link (km).
- α is the fiber *field* loss coefficient (km^{-1}), such that the signal *power* is attenuated as $\exp(-2\alpha z)$.
- α_{P} is the fiber power-loss coefficient (km^{-1}), with $\alpha_{\mathrm{P}} \triangleq 2\alpha$.
- β_2 is the dispersion coefficient ($\mathrm{ps}^2\,\mathrm{km}^{-1}$). The relationship between β_2 and the widely used dispersion parameter D in ps/(nm km) is: $D = -(2\pi c/\lambda^2)\beta_2$, with c the speed of light in km/s and λ the light wavelength in nm.
- γ is the fiber nonlinearity coefficient ($\mathrm{W}^{-1}\,\mathrm{km}^{-1}$).
- L_{s} is the span length (km).
- L_{eff} is the span effective length defined as: $[1-\exp(-2\alpha L_{\mathrm{s}})]/2\alpha$ (km).
- N_{s} is the total number of spans in a link, sometimes written as N_{span} when necessary for clarity.
- $G_{\mathrm{S}}(f)$ is the PSD of the overall WDM transmitted signal (W/Hz).
- $G_{\mathrm{NLI}}(f)$ is the PSD of NLI noise (W/Hz).
- N_{ch} is the total number of channels present in the WDM comb.
- P_n is the launch power of the n-th channel in the WDM comb (W). The power of a single channel is also sometimes written P_{ch} when necessary for clarity.
- R_n is the symbol rate of the n-th channel (TBaud). The symbol rate of a single channel is also written R, or R_{ch} when necessary for clarity.
- $T_n = R_n^{-1}$ is the symbol time of the n-th channel (ps).
- Δf is the channel spacing, used for systems where it is uniform (THz).
- $s_n(t)$ is the pulse used by the n-th channel, in the time domain. Its Fourier transform is $s_n(f)$. The pulse is assumed to be normalized so that the integral of its absolute value squared is T_n. (Note that according to such normalization a channel with an ideal rectangular spectrum and bandwidth R_n would have the flat-top value of its Fourier transform $s_n(f)$ equal to R_n^{-1}.) If any predistortion or dispersion precompensation is applied at the transmitter, this should be taken into account in $s_n(t)$ and $s_n(f)$.
- B_n is the full bandwidth of the n-th channel (THz). If the channel is Nyquist, then $B_n = R_n$.
- f_n is the center frequency of the n-th channel (THz).
- $a^k_{x,n}$, $a^k_{y,n}$ are random variables corresponding to the symbols sent on the n-th channel at the k-th signaling time, on either the polarization $\hat{x}$ or $\hat{y}$; note that $|a^k_{x,n}|^2$, $|a^k_{y,n}|^2$ must have dimensions of power (W). See also (9.63).
- $\mathbb{E}[.]$ is the expected value of a random variable.
- $F_{l,k,m}(\Omega_1, \Omega_2, \Omega_3)$ are kernels describing the NLI contribution resulting from FWN interaction between frequency tones Ω_1, Ω_2, and Ω_3. The SPM and XPM kernels are given by $S_{l,k,m} = F_{l,k,m}(0,0,0)$ and $X_{l,k,m}(\Omega) = F_{l,k,m}(0,\Omega,\Omega)$, respectively.
- $g_\Omega(z,t)$ represents the dispersed waveform of the pulse transmitted over a WDM channel spaced by Ω from the channel of interest, when reaching point z along the fiber; $\tilde{g}_j()$ represents the baseband spectral shape of the transmitted pulses from the j-th interfering WDM channel.
- $\boldsymbol{a}_n$ and $\boldsymbol{r}_n$ represent the transmitted and received polarization multiplexed data symbols of the channel of interest in the n-th time slot.
- $\boldsymbol{b}_{n,j}$ represents the two-element data vector transmitted in the n-th time-slot over the j-th interfering WDM channel.
- $\mathbf{H}_l^{(n)}$ are 2×2 matrices representing the l-th nonlinear ISI term in the n-th time slot.
- $h_{zs}(t)$ is the impulse response of a filter collecting all linear effects accumulated from coordinate z to coordinate s.

$R_{zs}(\tau)$ is the cross-correlation function between the electric field at coordinate z and coordinate s, i.e., $R_{zs}(\tau) = \mathbb{E}[A(z, t+\tau)A^*(s,t)]$.

$\kappa_{n;n}$ is n-th-order cumulant of data symbols a. They are also written as $\kappa_{2;2} = -\mu_2^2\Phi$ and $\kappa_{3;3} = -\mu_2^3\Psi$, with $\mu_2 = \mathbb{E}[|a|^2]$.

F is the noise figure (dB) of an EDFA.

P_{NLT} is the nonlinear threshold, i.e., the optimal launched power maximizing the SNR at the receiver.

η is the normalized nonlinear interference variance $(1/\mathrm{mW}^2)$, i.e., $\sigma^2_{\mathrm{NLI}} = \eta P^3$. For a single-span it is indicated by η_1.

ϵ is the coherence accumulation factor of nonlinear interference along the distance. It is $0 \leq \epsilon \leq 1$, where $\epsilon = 0$ and $\epsilon = 1$ indicate incoherent and fully-coherent accumulation, respectively.

N_0 is the system reach (spans) of an homogeneous optical link, i.e., the maximum distance for which the SNR is equal to a given threshold. The reach in km is $N_0 L_s$.

$\mathcal{L}$ is a linear operator accounting for dispersive effects, i.e., $\mathcal{L} = \mathrm{j}(\beta_2/2)(\partial^2/\partial t^2)$.

$\mathcal{N}$ is a nonlinear operator accounting for the nonlinear Kerr effect and fiber attenuation, i.e., $\mathcal{N} = -\alpha_{\mathrm{p}}/2 - \mathrm{j}\gamma|A|^2$. Fiber attenuation, although a linear effect, is included in the nonlinear operator to make the SSFM more efficient.

$h(X)$ is the differential entropy of the random variable X. It gives a measure of the average rate at which information is produced by X. The differential entropy is also called continuous entropy to emphasize focus on a continuous random variable X, contrary to the entropy that refers to a discrete random variable.

References

9.1 G.P. Agrawal: *Non-Linear Fiber Optics*, 5th edn. (Academic Press, New York 2012)

9.2 C.R. Menyuk, B.S. Marks: Interaction of polarization mode dispersion and nonlinearity in optical fiber transmission systems, J. Lightwave Technol. **24**(7), 2806–2826 (2006)

9.3 G. Bosco, A. Carena, R. Cigliutti, V. Curri, P. Poggiolini, F. Forghieri: Performance prediction for WDM PM-QPSK transmission over uncompensated links. In: *Proc. OFC 2011, Los Angeles* (2011), paper OTh07

9.4 E. Grellier, A. Bononi: Quality parameter for coherent transmissions with Gaussian-distributed nonlinear noise, Opt. Express **19**(13), 12781–12788 (2011)

9.5 C.R. Menyuk: Nonlinear pulse propagation in birefringent optical fibers, IEEE J. Quantum Electron. **23**, 174–176 (1987)

9.6 C.R. Menyuk: Pulse propagation in an elliptically birefringent Kerr medium, IEEE J. Quantum Electron. **25**(12), 2674–2682 (1989)

9.7 J.N. Damask: *Polarization Optics in Telecommunications* (Springer, Berlin, Heidelberg, New York 2005)

9.8 S.G. Evangelides Jr., L.F. Mollenauer, J.P. Gordon, N.S. Bergano: Polarization multiplexing with solitons, J. Lightwave Technol. **10**(1), 28–35 (1992)

9.9 P.K.A. Wai, C.R. Menyuk: Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence, J. Lightwave Technol. **14**(2), 148–157 (1996)

9.10 T.R. Taha, M.J. Ablowitz: Analytical and numerical aspects of certain nonlinear evolution equation, ii, numerical, nonlinear schroedinger equation, J. Comput. Phys. **5**, 203–230 (1984)

9.11 R. Hardin, F.D. Tappert: Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations, SIAM Rev. Chron. **15**, 423 (1973)

9.12 A. Hasegawa, F. Tappert: Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. Anomalous dispersion, Appl. Phys. Lett. **23**, 142–144 (1973)

9.13 J. Shao, X. Liang, S. Kumar: Comparison of split-step Fourier schemes for simulating fiber optic communication systems, IEEE Photonics J. **6**(4), 1–16 (2014)

9.14 A.V. Oppenheim, R.W. Schafer: *Discrete-Time Signal Processing* (Prentice Hall, Upper Saddle River 1999)

9.15 O.V. Sinkin, R. Holzlöhner, J. Zweck, C.R. Menyuk: Optimization of the split-step Fourier method in modeling optical-fiber communications systems, J. Lightwave Technol. **21**(1), 61–68 (2003)

9.16 Q. Zhang, M.I. Hayee: Symmetrized split-step Fourier scheme to control global simulation accuracy in fiber-optic communication systems, J. Lightwave Technol. **26**(2), 302–316 (2008)

9.17 S. Savory: Digital filters for coherent optical receivers, Opt. Express **16**(2), 804–817 (2008)

9.18 D. Zwillinger: *Handbook of Differential Equations*, 3rd edn. (Academic, New York 1997)

9.19 M. Schetzen: *The Volterra and Wiener Theories of Nonlinear Systems* (Wiley, Hoboken 1980)

9.20 K.V. Peddanarappagari, M. Brandt-Pearce: Volterra series transfer function of single-mode fibers, J. Lightwave Technol. **15**(12), 2232–2241 (1997)

9.21 M.J. Ablowitz, G. Biondini: Multiscale pulse dynamics in communication systems with strong dispersion management, Opt. Lett. **23**(21), 1668–1670 (1998)

9.22 A. Mecozzi, C. Balslev Clausen, M. Shtaif: Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission, IEEE Photonics Technol. Lett. **12**(4), 392–394 (2000)

9.23 A. Vannucci, P. Serena, A. Bononi: The RP method: a new tool for the iterative solution of the non-

linear Schrodinger equation, J. Lightwave Technol. **20**(7), 1102–1112 (2002)
9.24 J. Tang: The channel capacity of a multispan DWDM system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver, J. Lightwave Technol. **20**(7), 1095–1101 (2002)
9.25 P. Serena, A. Bononi: A time-domain extended Gaussian noise model, J. Lightwave Technol. **33**(7), 1459–1472 (2015)
9.26 E. Ciaramella, E. Forestieri: Analytical approximation of nonlinear distortions, IEEE Photonics Technol. Lett. **17**(1), 91–93 (2005)
9.27 P. Serena, A. Bononi: An alternative approach to the Gaussian noise model and its system implications, J. Lightwave Technol. **31**(22), 3489–3499 (2013)
9.28 A. Splett, C. Kurzke, K. Petermann: Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), Montreux*, Vol. 2 (1993) pp. 41–44
9.29 H. Louchet, A. Hodzic, K. Petermann: Analytical model for the performance evaluation of DWDM transmission systems, IEEE Photonics Technol. Lett. **15**(9), 1219–1221 (2003)
9.30 E.E. Narimanov, P.P. Mitra: The channel capacity of a fiber optics communication system: perturbation theory, J. Lightwave Technol. **20**(3), 530–537 (2002)
9.31 M. Nazarathy, J. Khurgin, R. Weidenfeld, Y. Meiman, R. Noe, P. Cho, I. Shpantzer, V. Karagodsky: Phased-array cancellation of nonlinear FWM in coherent OFDM dispersive multi-span links, Opt. Express **16**, 15778–15810 (2008)
9.32 X. Chen, W. Shieh: Closed-form expressions for nonlinear transmission performance of densely spaced coherent optical OFDM systems, Opt. Express **18**, 19039–19054 (2010)
9.33 P. Poggiolini, A. Carena, V. Curri, G. Bosco, F. Forghieri: Analytical modeling of non-linear propagation in uncompensated optical transmission links, IEEE Photonics Technol. Lett. **23**(11), 742–744 (2011)
9.34 A. Carena, V. Curri, G. Bosco, P. Poggiolini, F. Forghieri: Modeling of the impact of non-linear propagation effects in uncompensated optical coherent transmission links, J. Lightwave Technol. **30**(10), 1524–1539 (2012)
9.35 P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, F. Forghieri: A detailed analytical derivation of the GN model of non-linear interference in coherent optical transmission systems, arXiv: 1209.0394 (2012)
9.36 P. Johannisson, M. Karlsson: Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system, J. Lightwave Technol. **31**(8), 1273–1282 (2013)
9.37 P. Poggiolini: The GN model of non-linear propagation in uncompensated coherent optical systems, J. Lightwave Technol. **30**(24), 3857–3879 (2012)
9.38 S.J. Savory: Approximations for the nonlinear self-channel interference of channels with rectangular spectra, IEEE Photonics Technol. Lett. **25**(10), 961–964 (2013)
9.39 P. Johannisson, E. Agrell: Modeling of nonlinear signal distortion in fiber-optic networks, J. Lightwave Technol. **32**(23), 3942–3950 (2014)
9.40 A. Bononi, O. Beucher, P. Serena: Single- and cross-channel nonlinear interference in the gaussian noise model with rectangular spectra, Opt. Express **21**(26), 32254–32268 (2013)
9.41 P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, F. Forghieri: The GN model of fiber non-linear propagation and its applications, J. Lightwave Technol. **32**(4), 694–721 (2014)
9.42 P. Poggiolini, A. Carena, Y. Jiang, G. Bosco, V. Curri, F. Forghieri: Impact of low-OSNR operation on the performance of advanced coherent optical transmission systems. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), Cannes* (2014), paper arXiv:1407.2223
9.43 P. Serena: Nonlinear signal–noise interaction in optical links with nonlinear equalization, J. Lightwave Technol. **34**(6), 1476–1483 (2016)
9.44 P.K.A. Wai, C.R. Menyuk, H.H. Chen: Stability of solitons in randomly varying birefringent fibers, Opt. Lett. **16**(16), 1231–1233 (1991)
9.45 D. Marcuse, C.R. Menyuk, P.K.A. Wai: Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence, J. Lightwave Technol. **15**(9), 1735–1746 (1997)
9.46 Matlab support website: https://it.mathworks.com/help/matlab/functionlist.html (2019)
9.47 P. Poggiolini, G. Bosco, A. Carena, R. Cigliutti, V. Curri, F. Forghieri, R. Pastorelli, S. Piciaccia: The LOGON strategy for low-complexity control plane implementation in new-generation flexible networks. In: *Proc. OFC 2013, Los Angeles* (2013), paper OW1H.3
9.48 R. Pastorelli, S. Piciaccia, G. Galimberti, E. Self, M. Brunella, G. Calabretta, F. Forghieri, D. Siracusa, A. Zanardi, E. Salvadori, G. Bosco, A. Carena, V. Curri, P. Poggiolini: Optical control plane based on an analytical model of non-linear transmission effects in a self-optimized network. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), London* (2013), paper We.3.E.4
9.49 R. Pastorelli, G. Bosco, A. Nespola, S. Piciaccia, F. Forghieri: Network planning strategies for next-generation flexible optical networks. In: *Proc. OFC 2014, San Francisco* (2014), paper M2B.1
9.50 P. Poggiolini, Y. Jiang: Recent advances in the modeling of the impact of nonlinear fiber propagation effects on uncompensated coherent transmission systems, tutorial review, J. Lightwave Technol. **35**(3), 458–480 (2017)
9.51 A. Carena, G. Bosco, V. Curri, P. Poggiolini, F. Forghieri: Impact of the transmitted signal initial dispersion transient on the accuracy of the GN-model of non-linear propagation. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), London* (2013), paper Th.1.D.4
9.52 P. Serena, A. Bononi: On the accuracy of the Gaussian nonlinear model for dispersion-unmanaged coherent links. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), London* (2013), paper Th.1.D.3

9.53 R. Dar, M. Feder, A. Mecozzi, M. Shtaif: Properties of nonlinear noise in long, dispersion-uncompensated fiber links, Opt. Express **21**(22), 25685–25699 (2013)

9.54 R. Dar, M. Feder, A. Mecozzi, M. Shtaif: Accumulation of nonlinear interference noise in fiber-optic systems, Opt. Express **22**(12), 14199–14211 (2014)

9.55 R. Dar, M. Feder, A. Mecozzi, M. Shtaif: Pulse collision picture of inter-channel nonlinear interference noise in fiber-optic communications, J. Lightwave Technol. **34**, 593–607 (2016)

9.56 E. Torrengo, R. Cigliutti, G. Bosco, A. Carena, V. Curri, P. Poggiolini, A. Nespola, D. Zeolla, F. Forghieri: Experimental validation of an analytical model for nonlinear propagation in uncompensated optical links, Opt. Express **19**(26), B790–B798 (2011)

9.57 J.-X. Cai, H. Zhang, H.G. Batshon, M. Mazurczyk, O.V. Sinkin, D.G. Foursa, A.N. Pilipetskii, G. Mohs, N.S. Bergano: 200 Gb/s and dual wavelength 400 Gb/s transmission over transpacific distance at 6.0 b/s/Hz spectral efficiency, J. Lightwave Technol. **32**(4), 832–839 (2014)

9.58 J.-X. Cai, O.V. Sinkin, H. Zhang, H.G. Batshon, M. Mazurczyk, D.G. Foursa, A. Pilipetskii, G. Mohs: Nonlinearity compensation benefit in high capacity ultra-long haul transmission systems. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), London* (2013), paper We.4.D.2

9.59 A.J. Stark, Y.-T. Hsueh, T.F. Detwiler, M.M. Filer, S. Tibuleac, S.E. Ralph: System performance prediction with the Gaussian noise model in 100G PDM-QPSK coherent optical networks, J. Lightwave Technol. **31**(21), 3352–3360 (2013)

9.60 A. Nespola, S. Straullu, A. Carena, G. Bosco, R. Cigliutti, V. Curri, P. Poggiolini, M. Hirano, Y. Yamamoto, T. Sasaki, J. Bauwelinck, K. Verheyen, F. Forghieri: GN-model validation over seven fiber types in uncompensated PM-16QAM Nyquist-WDM links, IEEE Photonics Technol. Lett. **26**(2), 206–209 (2014)

9.61 J.-X. Cai, H.G. Batshon, H. Zhang, M. Mazurczyk, O.V. Sinkin, D.G. Foursa, A.N. Pilipetskii: Transmission performance of coded modulation formats in a wide range of spectral efficiencies. In: *Proc. OFC, San Francisco* (2014), paper M2C.3

9.62 J. Pan, P. Isautier, M. Filer, S. Tibuleac, S.E. Ralph: Gaussian noise model aided in-band crosstalk analysis in ROADM-enabled DWDM networks. In: *Proc. OFC, San Francisco* (2014), paper Th1I.1

9.63 P. Poggiolini, G. Bosco, A. Carena, F. Guiomar, M. Ranjbar Zefreh, F. Forghieri, S. Piciaccia: Nonlinearity modeling at ultra-high symbol rates. In: *Proc. OFC 2018, paper W1G.3, San Diego* (2018)

9.64 P. Poggiolini, G. Bosco, A. Carena, D. Pilori, A. Nespola, M. Ranjbar Zefreh, M. Bertino, F. Forghieri: Non-linearity modeling for Gaussian-constellation systems at ultra-high symbol rates. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), paper Tu4G.3, Rome* (2018)

9.65 A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri: EGN model of non-linear fiber propagation, Opt. Express **22**(13), 16335–16362 (2014)

9.66 A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, F. Forghieri: On the accuracy of the GN-model and on analytical correction terms to improve it, arXiv:1401.6946 (2014)

9.67 P. Poggiolini, Y. Jiang, A. Carena, F. Forghieri: Analytical modeling of the impact of fiber nonlinear propagation on coherent systems and networks. In: *Enabling Technologies for High Spectral-Efficiency Coherent Optical Communication Networks*, ed. by X. Zhou, C. Xie (Wiley, Hoboken 2016)

9.68 P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, F. Forghieri: A simple and effective closed-form GN model correction formula accounting for signal non-Gaussian distribution, J. Lightwave Technol. **33**(2), 459–473 (2015)

9.69 P. Poggiolini, Y. Jiang, A. Carena, F. Forghieri: A simple and accurate closed-form EGN model formula, arXiv:1503.04132 (2015)

9.70 D. Semrau, R.I. Killey, P. Bayvel: A closed-form approximation of the Gaussian noise model in the presence of inter-channel stimulated Raman scattering, arXiv:1808.07940 [eess.SP] (2018)

9.71 P. Poggiolini: A generalized GN-model closed-form formula, arXiv:1810.06545 [eess.SP] (2018)

9.72 D. Semrau, G. Saavedra, D. Lavery, R.I. Killey, P. Bayvel: A closed-form expression to evaluate nonlinear interference in Raman-amplified links, J. Lightwave Technol. **35**(19), 4316–4328 (2017)

9.73 M. Cantono, J.-L. Auge, V. Curri: Modelling the impact of SRS on NLI generation in commercial equipment: An experimental investigation. In: *Proc. OFC 2018, paper M1D.2, San Diego* (2018)

9.74 M. Cantono, D. Pilori, A. Ferrari, C. Catanese, J. Thouras, J.L. Auge, V. Curri: On the interplay of nonlinear interference generation with stimulated Raman scattering for QoT estimation, J. Lightwave Technol. **35**(15), 3131–3141 (2018)

9.75 A. Mecozzi, R.-J. Essiambre: Nonlinear Shannon limit in pseudolinear coherent systems, J. Lightwave Technol. **30**(12), 2011–2024 (2012)

9.76 M. Secondini, E. Forestieri: Analytical fiber-optic channel model in the presence of cross-phase modulations, IEEE Photonics Technol. Lett. **24**(22), 2016–2019 (2012)

9.77 R. Dar, M. Feder, A. Mecozzi, M. Shtaif: Time varying ISI model for nonlinear interference noise. In: *Opt. Fiber Commun. Conference (OFC), paper W2A.62* (2014)

9.78 R. Dar, P.J. Winzer: Nonlinear interference mitigation: Methods and potential gain, J. Lightwave Technol. **35**, 903–930 (2017)

9.79 R.C. Jones: A new calculus for the treatment of optical systems. VII. Properties of the N-matrices, J. Opt. Soc. Am. **38**, 671–685 (1948)

9.80 J.P. Gordon, H. Kogelnik: PMD fundamentals: Polarization mode dispersion in optical fibers, Proc. Nat. Acad. Sci. Am. **97**(9), 4541–4550 (2000)

9.81 O. Golani, R. Dar, M. Feder, A. Mecozzi, M. Shtaif: Modeling the bit-error-rate performance of nonlinear fiber-optic systems, J. Lightwave Technol. **34**, 3482–3489 (2016)

9.82 M. Secondini, E. Forestieri: On XPM mitigation in WDM fiber-optic systems, IEEE Photonics Technol. Lett. **26**(22), 2252–2255 (2014)

9.83 R. Dar, M. Feder, A. Mecozzi, M. Shtaif: Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation, J. Lightwave Technol. **33**, 1044–1053 (2015)

9.84 R. Dar, O. Geller, M. Feder, A. Mecozzi, M. Shtaif: Mitigation of inter-channel nonlinear interference in WDM systems. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), paper P.5.6, Cannes* (2014)

9.85 M.P. Yankov, T. Fehenberger, L. Barletta, N. Hanik: Low-complexity tracking of laser and nonlinear phase noise in WDM optical fiber systems, J. Lightwave Technol. **33**, 4975–4984 (2015)

9.86 C. Pan, H. Bülow, W. Idler, L. Schmalen, F.R. Kschischang: Optical nonlinear phase noise compensation for 9×32-GBaud PolDM-16 QAM transmission using a code-aided expectation-maximization algorithm, J. Lightwave Technol. **33**, 3679–3686 (2015)

9.87 C. Schmidt-Langhorst, R. Elschner, F. Frey, R. Emmerich, C. Schubert: Experimental analysis of nonlinear interference noise in heterogeneous flexgrid WDM transmission. In: *Europ. Conf. Opt. Commun. (ECOC)* (2015), paper Tu.1.4.3

9.88 L. Li, Z. Tao, L. Liu, W. Yan, S. Oda, T. Hoshida, J.C. Rasmussen: Nonlinear polarization-crosstalk canceller for dual-polarization digital coherent receivers. In: *Opt. Fiber Communic. Conf. (OFC)* (2010), paper OWE3

9.89 A. Ghazisaeidi, M. Salsi, J. Renaudier, O. Bertran-Pardo, H. Mardoyan, G. Charlet: Performance analysis of decision-aided nonlinear cross-polarization mitigation algorithm. In: *Europ. Conf. Opt. Commun. (ECOC)* (2012), paper We-3

9.90 P.M. Lushnikov: Fully parallel algorithm for simulating dispersion-managed wavelength-division-multiplexed optical fiber systems, Opt. Lett. **27**(11), 939 (2002)

9.91 E. Forestieri, M. Secondini: Solving the nonlinear Schrödinger equation. In: *Optical Communication Theory and Techniques*, ed. by E. Forestieri (Springer, New York 2005)

9.92 A.O. Korotkevich, P.M. Lushnikov: Proof-of-concept implementation of the massively parallel algorithm for simulation of dispersion-managed WDM optical fiber systems, Opt. Lett. **36**(10), 1851–1853 (2011)

9.93 E. Seve, P. Ramantanis, J.-C. Antona, E. Grellier, O. Rival, F. Vacondio, S. Bigo: Semi-analytical model for the performance estimation of 100Gb/s PDM-QPSK optical transmission systems without Inline dispersion compensation and mixed fiber types. In: *Proc. Europ. Conf. Opt. Commun. (ECOC) 2013* (2013), paper Th.1.D.2

9.94 A. Papoulis: *Probability, Random Variables, and Stochastic Processes*, 3rd edn. (McGraw-Hill, New York 1991)

9.95 M. Secondini, E. Forestieri, C.R. Menyuk: A combined regular-logarithmic perturbation method for signal-noise interaction in amplified optical systems, J. Lightwave Technol. **27**(16), 3358–3369 (2009)

9.96 M. Secondini, E. Forestieri, G. Prati: Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps, J. Lightwave Technol. **31**(23), 3839–3852 (2013)

9.97 P. Bello: Characterization of randomly time-variant linear channels, IEEE Trans. Commun. Syst. **11**, 360–393 (1963)

9.98 K. Liu, T. Kadous, A.M. Sayeed: Orthogonal time-frequency signaling over doubly dispersive channels, IEEE Trans. Inform. Theory **50**(11), 2583–2603 (2004)

9.99 M. Secondini, E. Agrell, E. Forestieri, D. Marsella: Fiber nonlinearity mitigation in WDM systems: Strategies and achievable rates. In: *Proc. Europ. Conf. Exhib. Opt. Commun., Sweden* (2017)

9.100 P.P. Mitra, J.B. Stark: Nonlinear limits to the information capacity of optical fiber communications, Nature **411**(6841), 1027–1030 (2001)

9.101 A.G. Green, P.B. Littlewood, P.P. Mitra, L.G.L. Wegener: Schroedinger equation with a spatially and temporally random potential: effects of cross-phase modulation in optical communication, Phys. Rev. E **66**(4), 046627 (2002)

9.102 S. Blanes, F. Casas, J. Oteo, J. Ros: The Magnus expansion and some of its applications, Phys. Rep. **470**(5–6), 151–238 (2009)

9.103 M. Reimer, D. Yevick, D. Dumas: The accuracy of the Magnus expansion for polarization-mode dispersion and polarization-dependent loss, J. Lightwave Technol. **26**(19), 3337–3344 (2008)

9.104 M. Secondini, E. Forestieri: A comparative analysis of different perturbation models for the nonlinear fiber channel. In: *Asia Commun. Photonics Conf. 2015, OSA Tech. Dig. (online), paper AM3D.1* (2015), https://www.osapublishing.org/abstract.cfm?uri=ACPC-2015-AM3D.1

9.105 G.P. Agrawal: *Fiber-Optic Communications Systems*, 3rd edn. (Wiley, Hoboken 2002)

9.106 V.E. Zakharov, L.A. Ostrovsky: Modulation instability: The beginning, Physica D **238**(5), 540–548 (2009)

9.107 J. Hansryd, P.A. Andrekson, M. Westlund, J. Li, P.O. Hedekvist: Fiber-based optical parametric amplifiers and their applications, IEEE J. Sel. Top. Quantum Electron. **8**(3), 506–520 (2002)

9.108 M.E. Marhic, P.A. Andrekson, P. Petropoulos, S. Radic, C. Peucheret, M. Jazayerifar: Fiber optical parametric amplifiers in optical communication systems, Laser Photon. Rev. **9**(1), 50–74 (2015)

9.109 J.P. Gordon, L.F. Mollenauer: Phase noise in photonic communications systems using linear amplifiers, Opt. Lett. **15**(23), 1351–1353 (1990)

9.110 H. Kim, A.H. Gnauck: Experimental investigation of the performance limitation of DPSK systems due to nonlinear phase noise, IEEE Photonics Technol. Lett. **15**(2), 320–322 (2003)

9.111 K.-P. Ho: Error probability of DPSK signals with cross-phase modulation induced nonlinear phase noise, IEEE J. Sel. Topics Quantum Electron. **10**(2), 421–427 (2004)

9.112 K.-P. Ho, H.-C. Wang: Effect of dispersion on nonlinear phase noise, Opt. Lett. **31**(14), 2109–2111 (2006)

9.113 R. Hui, K.R. Demarest, C.T. Allen: Cross-phase modulation in multispan WDM optical fiber systems, J. Lightwave Technol. **17**(6), 1018–1026 (1999)

9.114 A. Carena, V. Curri, R. Gaudino, P. Poggiolini, S. Benedetto: New analytical results on fiber parametric gain and its effects on ASE noise, IEEE Photonics Technol. Lett. **9**(4), 535–537 (1997)

9.115 B. Xu, M. Brandt-Pearce: Optical fiber parametric-gain-induced noise coloring and amplification by modulated signals, J. Opt. Soc. Am. B **21**(3), 499–513 (2004)

9.116 P. Serena, A. Orlandini, A. Bononi: Parametric-gain approach to the analysis of single-channel DPSK/DQPSK systems with nonlinear phase noise, J. Lightwave Technol. **24**(5), 2026–2037 (2006)

9.117 D. Rafique, J. Zhao, D. Ellis: Digital back-propagation for spectrally efficient WDM 112 Gbit/s PM m-ary QAM transmission, Opt. Express **19**(6), 5219–5224 (2011)

9.118 L. Beygi, N. Irukulapati, E. Agrell, P. Johannisson, M. Karlsson, H. Wymeersch, P. Serena, A. Bononi: On nonlinearly-induced noise in single-channel optical links with digital backpropagation, Opt. Express **21**, 26376–26386 (2013)

9.119 A.D. Ellis, M.E. McCarthy, M.A.Z. Al-Khateeb, S. Sygletos: Capacity limits of systems employing multiple optical phase conjugators, Opt. Express **23**(16), 20381 (2015)

9.120 D. Lavery, D. Ives, G. Liga, A. Alvarado, S.J. Savory, P. Bayvel: The benefit of split nonlinearity compensation for single-channel optical fiber communications, IEEE Photonics Technol. Lett. **28**(17), 1803–1806 (2016)

9.121 A. Ghazisaeidi: A theory of nonlinear interactions between signal and amplified spontaneous emission noise in coherent wavelength division multiplexed systems, J. Lightwave Technol. **35**(23), 5150–5175 (2017)

9.122 I.P. Kaminow, T. Li, A.E. Willner: *Optical Fiber Telecommunications Volume VIB: Systems and Networks* (Academic, New York 2013)

9.123 G.J. Foschini, C.D. Poole: Statistical theory of polarization dispersion in single mode fibers, J. Lightwave Technol. **9**(11), 1439–1456 (1991)

9.124 M. Karlsson: Probability density functions of the differential group delay in optical fiber communication systems, J. Lightwave Technol. **19**(3), 324–331 (2001)

9.125 C.D. Poole: Statistical treatment of polarization dispersion in single-mode fiber, Opt. Lett. **13**(8), 687 (1988)

9.126 D. Nolan, X. Chen, M.-J. Li: Fibers with low polarization-mode dispersion, J. Lightwave Technol. **22**(4), 1066–1077 (2004)

9.127 D. Breuer, H. Tessmann, A. Gladisch, H.M. Foisel, G. Neumann, H. Reiner, H. Cremer: Measurements of PMD in the installed fiber plant of Deutsche Telekom. In: *Proc. LEOS 2003* (2003)

9.128 S. Haykin: *Adaptive Filter Theory*, 3rd edn. (Prentice, Upper Saddle River 1996)

9.129 B.C. Collings, L. Boivin: Nonlinear polarization evolution induced by cross-phase modulation and its impact on transmission systems, IEEE Photonics Technol. Lett. **12**, 1582–1584 (2000)

9.130 M. Karlsson, H. Sunnerud: Effects of nonlinearities on PMD-induced system impairments, J. Lightwave Technol. **24**(11), 4127–4137 (2006)

9.131 M. Winter, C.-A. Bunge, D. Setti, K. Petermann: A statistical treatment of cross-polarization modulation in DWDM systems, J. Lightwave Technol. **27**(17), 3739–3751 (2009)

9.132 N. Rossi, P. Serena, A. Bononi: Symbol-rate dependence of dominant nonlinearity and reach in coherent WDM links, J. Lightwave Technol. **33**(14), 3132–3143 (2015)

9.133 C. Xie: WDM coherent PDM-QPSK systems with and without inline optical dispersion compensation, Opt. Express **17**(6), 4815–4823 (2009)

9.134 M. Boroditsky, M. Bourd, M. Tur: Effect of nonlinearities on PMD, J. Lightwave Technol. **24**, 4100–4107 (2006)

9.135 O. Bertran-Pardo, J. Renaudier, G. Charlet, P. Tran, H. Mardoyan, M. Bertolini, M. Salsi, S. Bigo: Demonstration of the benefits brought by PMD in polarization-multiplexed systems. In: *Proc. Europ. Conf. Opt. Commun. (ECOC) 2010, Turin, paper Th.10.E.4* (2010)

9.136 P. Serena, N. Rossi, O. Bertran-Pardo, J. Renaudier, A. Vannucci, A. Bononi: Intra- versus inter-channel PMD in linearly compensated coherent PDM-PSK nonlinear transmissions, J. Lightwave Technol. **29**(11), 1691–1700 (2011)

9.137 G. Gao, X. Chen, W. Shieh: Influence of PMD on nonlinearity compensation using digital back propagation, Opt. Express **20**, 14406–14418 (2012)

9.138 N. Rossi, P. Serena, A. Bononi: Stratified-sampling estimation of PDL-induced outage probability in nonlinear coherent systems, J. Lightwave Technol. **32**(24), 4905–4911 (2014)

9.139 A. Mecozzi, M. Shtaif: Signal-to-noise-ratio degradation caused by polarization-dependent loss and the effect of dynamic gain equalization, J. Lightwave Technol. **22**(8), 1856–1871 (2004)

9.140 M. Karlsson, M. Petersson: Quaternion approach to PMD and PDL phenomena in optical fiber systems, J. Lightwave Technol. **22**(4), 1137–1146 (2004)

9.141 Z. Tao, L. Dou, T. Hoshida, J.C. Rasmussen: A fast method to simulate the PDL impact on dual-polarization coherent systems, IEEE Photonics Technol. Lett. **21**(24), 1882–1884 (2009)

9.142 N. Rossi, P. Serena, A. Bononi: Polarization-dependent loss impact on coherent optical systems in presence of fiber nonlinearity, IEEE Photonics Technol. Lett. **26**(4), 334–337 (2014)

9.143 O. Vassilieva, S. Oda, T. Hoshida, J.C. Rasmussen, M. Sekiya: Experimental investigation of the statistics of the interplay between nonlinear and PDL effects in polarization multiplexed systems. In: *Proc. OFC 2013*, Vol. OM3B.6 (2013) pp. 6–8

9.144 D.A. Morero, M.A. Castrillon, A. Aguirre, M.R. Hueda, O.E. Agazzi: Design tradeoffs and challenges in practical coherent optical transceiver implementations, J. Lightwave Technol. **34**(1), 121–136 (2016)

9.145 A. Alvarado, E. Agrell, D. Lavery, R. Maher, P. Bayvel: Replacing the soft-decision FEC limit paradigm in the design of optical communication systems, J. Lightwave Technol. **34**(2), 707–721 (2016)

9.146 T.M. Cover, J.A. Thomas: *Elements of Information Theory*, 2nd edn. (Wiley, Hoboken 2005)

9.147 D.M. Arnold, H.-A. Loeliger, P.O. Vontobel, A. Kavvcic, W. Zeng: Simulation-based computation of information rates for channels with memory, IEEE Trans. Inf. Theory **52**(8), 3498–3508 (2006)

9.148 I.B. Djordjevic, B. Vasic, M. Ivkovic, I. Gabitov: Achievable information rates for high-speed long-haul optical transmission, J. Lightwave Technol. **23**(11), 3755–3763 (2005)

9.149 G. Colavolpe, T. Foggi, A. Modenini, A. Piemontese: Faster-than-Nyquist and beyond: How to improve spectral efficiency by accepting interference, Opt. Express **19**, 26600–26609 (2011)

9.150 N. Merhav, G. Kaplan, A. Lapidoth, S. Shamai: On information rates for mismatched decoders, IEEE Trans. Inform. Theory **40**(6), 1953–1967 (1994)

9.151 R.E. Blahut: *Principles and Practice of Information Theory* (Addison-Wesley, Boston 1988)

9.152 A. Guillén i Fàbregas, A. Martinez, G. Caire: Bit-interleaved coded modulation, Found. Trends Commun. Inf. Theory **5**(1), 1–153 (2008)

9.153 G. Caire, G. Taricco, E. Biglieri: Bit-interleaved coded modulation, IEEE Trans. Inf. Theory **44**(3), 927–946 (1998)

9.154 M. Karlsson, E. Agrell: Multidimensional modulation and coding in optical transport, J. Lightwave Technol. **35**(4), 876–884 (2017)

9.155 L. Schmalen, A. Alvarado, R. Rios-Müller: Performance prediction of nonbinary forward error correction in optical transmission experiments, J. Lightwave Technol. **35**(4), 1015–1027 (2017)

9.156 J. Cho, L. Schmalen, P.J. Winzer: Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM. In: *Proc. Europ. Conf. Opt. Commun. (ECOC) 2017, Göteborg*, Vol. M.2.D.2. (2017)

9.157 T. Yoshida, M. Karlsson, E. Agrell: Performance metrics for systems with soft-decision FEC and probabilistic shaping, IEEE Photonics Technol. Lett. **29**(23), 2111–2114 (2017)

9.158 F. Vacondio, O. Rival, C. Simonneau, E. Grellier, A. Bononi, L. Lorcy, J.-C. Antona, S. Bigo: On nonlinear distortions of highly dispersive optical coherent systems, Opt. Express **20**(2), 1022–1032 (2012)

9.159 A. Bononi, P. Serena, N. Rossi, E. Grellier, F. Vacondio: Modeling nonlinearity in coherent transmissions with dominant intrachannel-four-wave-mixing, Opt. Express **20**, 7777–7791 (2012)

9.160 V. Curri, A. Carena, A. Arduino: Design strategies and merit of system parameters for uniform uncompensated links supporting Nyquist-WDM transmission, J. Lightwave Technol. **33**(18), 3921–3932 (2015)

9.161 I. Tomkos, S. Azodolmolky, J. Solé-Pareta, D. Careglio, E. Palkopoulou: A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges, Proc. IEEE **102**(9), 1317–1337 (2014)

9.162 G. Zhang, M. De Leenheer, A. Morea, B. Mukherjee: A survey on OFDM-based elastic core optical networking, IEEE Commun. Surv. Tutor. **15**, 65–87 (2013)

9.163 S. Azodolmolky, M. Angelou, I. Tomkos, A. Morea, Y. Pointurier, J. Solé-Pareta: A comparative study of impairments aware optical networks planning tools. In: *Broadband Communications, Networks, and Systems. BROADNETS 2010*, Lect. Notes Inst. Comp. Sci. Soc. Inf. Telecomm. Eng., Vol. 66 (2012) pp. 491–500

9.164 R. Pastorelli, G. Bosco, S. Piciaccia, F. Forghieri: Network planning strategies for next-generation flexible optical networks, J. Opt. Commun. Netw. **7**(3), A511–A525 (2015)

9.165 E. Palkopoulou, G. Bosco, A. Carena, D. Klonidis, P. Poggiolini, I. Tomkos: Nyquist-WDM-based flexible optical networks: exploring physical layer design parameters, J. Lightwave Technol. **31**(14), 2332–2339 (2013)

9.166 S.J. Savory: Congestion aware routing in nonlinear elastic optical networks, IEEE Photonics Technol. Lett. **26**(10), 1057–1060 (2015)

9.167 V. Curri, M. Cantono, R. Gaudino: Elastic all-optical networks: a new paradigm enabled by the physical layer. How to optimize network performances?, J. Lightwave Technol. **35**, 1211–1221 (2017)

9.168 D.J. Ives, A. Alvarado, S.J. Savory: Throughput gains from adaptive transceivers in nonlinear elastic optical networks, J. Lightwave Technol. **35**(6), 1280–1289 (2017)

9.169 C. Rottondi, L. Barletta, A. Giusti, M. Tornatore: A machine learning method for quality of transmission estimation of unestablished lightpaths, J. Opt. Commun. Netw. **10**(2), A286 (2018)

9.170 R. Dar, M. Feder, A. Mecozzi, M. Shtaif: Nonlinear interference noise wizard, http://nlinwizard.eng.tau.ac.il

9.171 L. Yan, E. Agrell, H. Wymeersch, P. Johannisson, R. Di Taranto, M. Brandt-Pearce: Link-level resource allocation for flexible-grid nonlinear fiber-optic communication systems, IEEE Photonics Technol. Lett. **27**(12), 1250–1253 (2015)

9.172 L. Yan, E. Agrell, H. Wymeersch, M. Brandt-Pearce: Resource allocation for flexible-grid optical networks with nonlinear channel model, J. Opt. Commun. Netw. **7**(11), B101–B108 (2015)

9.173 J. Zhao, H. Wymeersch, E. Agrell: Nonlinear impairment-aware static resource allocation in elastic optical networks, J. Lightwave Technol. **32**(22), 4554–4564 (2015)

9.174 M.N. Dharmaweera, J. Zhao, L. Yan, M. Karlsson, E. Agrell: Traffic-grooming- and multipath-routing-enabled impairment-aware elastic optical networks, J. Opt. Commun. Netw. **8**(2), 58–70 (2016)

9.175 D.J. Ives, P. Bayvel, S.J. Savory: Adapting transmitter power and modulation format to improve optical network performance utilizing the Gaussian

noise model of nonlinear impairments, J. Lightwave Technol. **32**(21), 3485 (2014)
9.176 D.J. Ives, P. Bayvel, S.J. Savory: Routing, modulation, spectrum and launch power assignment to maximize the traffic throughput of a nonlinear optical mesh network, Photon. Netw. Commun. **29**(3), 244–256 (2015)
9.177 K. Christodoulopoulos, I. Tomkos, E.A. Varvarigos: Elastic bandwidth allocation in flexible OFDM-based optical networks, J. Lightwave Technol. **29**, 1354–1366 (2011)
9.178 L. Yan, E. Agrell, M.N. Dharmaweera, H. Wymeersch: Joint assignment of power, routing, and spectrum in static flexible-grid networks, J. Lightwave Technol. **35**(10), 1766–1774 (2017)
9.179 Y. Xu: Resource Allocation in Elastic Optical Networks with Physical-Layer Impairments, Dissertation, University of Virginia (2017)
9.180 A. Bononi, P. Serena, A. Morea, G. Picchi: Regeneration savings in flexible optical networks with a new load-aware reach maximization, Opt. Switch. Netw. **19**, 123–134 (2016)
9.181 T. Panayiotou, S.P. Chatzis, G. Ellinas: Performance analysis of a data-driven quality-of-transmission decision approach on a dynamic multicast-capable metro optical network, J. Opt. Commun. Netw. **9**(1), 98–108 (2017)
9.182 J.C. Cartledge, F.P. Guiomar, F.R. Kschischang, G. Liga, M.P. Yankov: Digital signal processing for fiber nonlinearities, Opt. Express **25**(3), 1916–1936 (2017)
9.183 V. Kamalov, L. Jovanovski, V. Vusirikala, S. Zhang, F. Yaman, K. Nakamura, T. Inoue, E. Mateo, Y. Inada: Evolution from 8QAM live traffic to PS 64-QAM with neural-network based nonlinearity compensation on 11000 km open subsea cable. In: *Proc. OFC 2018, post-deadline paper Th4D.5, San Diego* (2018)
9.184 A.D. Ellis, M.A. Zaki Al Khateeb, M.E. McCarthy: Impact of optical phase conjugation on the nonlinear Shannon limit, J. Lightwave Technol. **35**(4), 792–798 (2017)
9.185 R.-J. Essiambre, P.J. Winzer: Fibre nonlinearities in electronically pre-distorted transmission. In: *Proc. Europ. Conf. Opt. Commun., paper Tu.3.2.2* (2005)
9.186 X. Li, X. Chen, G. Goldfarb, E. Mateo, I. Kim, F. Yaman, G. Li: Electronic post-compensation of WDM transmission impairments using coherent detection and digital signal processing, Opt. Express **16**, 880–888 (2008)
9.187 E. Ip, J.M. Kahn: Compensation of dispersion and nonlinear impairments using digital backpropagation, J. Lightwave Technol. **26**(20), 3416–3425 (2008)
9.188 S.J. Savory, G. Gavioli, E. Torrengo, P. Poggiolini: Impact of interchannel nonlinearities on a split-step intrachannel nonlinear equalizer, IEEE Photonics Technol. Lett. **22**(10), 673–675 (2010)
9.189 G. Liga, T. Xu, A. Alvarado, R.I. Killey, P. Bayvel: On the performance of multichannel digital backpropagation in high-capacity longhaul optical transmission, Opt. Express **22**, 30053–30062 (2014)
9.190 D. Rafique: Fiber nonlinearity compensation: Commercial applications and complexity analysis, J. Lightwave Technol. **34**(2), 544–553 (2016)
9.191 M. Secondini, S. Rommel, G. Meloni, F. Fresi, E. Forestieri, L. Potì: Single-step digital backpropagation for nonlinearity mitigation, Photon. Netw. Commun. **31**(3), 493–502 (2016)
9.192 L.B. Du, A.J. Lowery: Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems, Opt. Express **18**(16), 17075–17088 (2010)
9.193 L. Li, Z. Tao, L. Dou, W. Yan, S. Oda, T. Tanimura, T. Hoshida, J.C. Rasmussen: Implementation efficient nonlinear equalizer based on correlated digital backpropagation. In: *Opt. Fiber Commun. Conf., Los Angeles, paper OWW3* (2011)
9.194 M. Secondini, D. Marsella, E. Forestieri: Enhanced split-step fourier method for digital backpropagation. In: *Proc. Europ. Conf. Opt. Commun., paper We.3.3.5* (2014)
9.195 A. Napoli, Z. Maalej, V.A.J.M. Sleiffer, M. Kuschnerov, D. Rafique, E. Timmers, B. Spinnler, T. Rahman, L.D. Coelho, N. Hanik: Reduced complexity digital back-propagation methods for optical communication systems, J. Lightwave Technol. **32**(7), 1351–1362 (2014)
9.196 J. Gonçalves, C.S. Martins, F.P. Guiomar, T. Cunha, J. Pedro, A.N. Pinto, P. Lavrador: Nonlinear compensation with DBP aided by a memory polynomial, Opt. Express **24**(26), 30309–30316 (2016)
9.197 L. Zhu, G. Li: Folded digital backward propagation for dispersion-managed fiber-optic transmission, Opt. Express **19**(7), 5953 (2011)
9.198 Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, J.C. Rasmussen: Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate, J. Lightwave Technol. **29**(17), 2570–2576 (2011)
9.199 Z. Tao, L. Dou, Y. Zhao, W. Yan, T. Oyama, S. Oda, T. Tanimura, T. Hoshida, J.C. Rasmussen: Nonlinear compensation beyond conventional perturbation based algorithms. In: *Proc. ACP/IPOC 13*, Vol. ATh4E.1 (2013) pp. 4–6
9.200 Q. Zhuge, M. Reimer, A. Borowiec, M. O'Sullivan, D.V. Plant: Aggressive quantization on perturbation coefficients for nonlinear pre-distortion. In: *Proc. OFC 2014*, Vol. h4D.7 (2014)
9.201 Y. Gao, J.C. Cartledge, A.S. Karar, S.S.-H. Yam, M. O'Sullivan, C. Laperle, A. Borowiec, K. Roberts: Reducing the complexity of perturbation based nonlinearity pre-compensation using symmetric EDC and pulse shaping, Opt. Express **22**(2), 1209 (2014)
9.202 X. Liang, S. Kumar: Multi-stage perturbation theory for compensating intra-channel nonlinear impairments in fiber-optic links, Opt. Express **22**(24), 29733–29745 (2014)
9.203 F.P. Guiomar, A. Pinto: Simplified Volterra series nonlinear equalizer for polarization-multiplexed coherent optical systems, J. Lightwave Technol. **31**(23), 3879–3891 (2013)

9.204 A. Ghazisaeidi, R.-J. Essiambre: Calculation of coefficients of perturbative nonlinear pre-compensation for Nyquist pulses. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), Cannes* (2014)

9.205 A. Ghazisaeidi, I. Fernandez de Jauregui Ruiz, L. Schmalen, P. Tran, C. Simonneau, E. Awwad, B. Uscumlic, P. Brindel, G. Charlet: Submarine transmission systems using digital nonlinear compensation and adaptive rate forward error correction, J. Lightwave Technol. **34**(8), 1886–1895 (2016)

9.206 C.B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S.J. Savory, P. Bayvel: Polarization-mode dispersion aware digital backpropagation. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), Düsseldorf* (2016) pp. 1091–1093

9.207 K. Goroshko, H. Louchet, A. Richter: Overcoming performance limitations of digital back propagation due to polarization mode dispersion. In: *Proc. Internat. Conf. Transpar. Opt. Netw. (ICTON), Trento, paper Mo.B1.4* (2016)

9.208 C.B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S.J. Savory, P. Bayvel: Digital backpropagation accounting for polarization-mode dispersion, Opt. Express **25**, 1903–1915 (2017)

9.209 R. Holzlöhner, V.S. Grigoryan, C.R. Menyuk, W.L. Kath: Accurate calculation of eye diagrams and bit error rates in optical transmission systems using linearization, J. Lightwave Technol. **20**(3), 389–400 (2002)

9.210 D. Marsella, M. Secondini, E. Forestieri: Maximum likelihood sequence detection for mitigating nonlinear effects, J. Lightwave Technol. **32**(5), 908–916 (2014)

9.211 N.V. Irukulapati, H. Wymeersch, P. Johannisson, E. Agrell: Stochastic digital backpropagation, IEEE Trans. Commun. **62**(11), 3956–3968 (2014)

9.212 T.A. Eriksson, T. Fehenberger, P.A. Andrekson, M. Karlsson, N. Hanik, E. Agrell: Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments, J. Lightwave Technol. **34**(9), 2256–2266 (2016)

9.213 A. Napoli, M. Kuschnerov, C.-Y. Lin, B. Spinnler, M. Bohn, D. Rafique, V.A. Sleiffer, B. Schmauss: Adaptive digital back-propagation for optical communication systems. In: *Proc. Opt. Fiber Commun. Conf. 2014, paper M3C.4* (2014)

9.214 E.P. da Silva, R. Asif, K.J. Larsen, D. Zibar: Nonlinear compensation with modified adaptive digital backpropagation in flexigrid networks. In: *Proc. Conf. Lasers Electro-Opt. 2015, paper SM2M, San Jose* (2015)

9.215 W. Shieh, Y. Tang: Ultrahigh-speed signal transmission over nonlinear and dispersive fiber optic channel: the multicarrier advantage, IEEE Photonics J. **2**(3), 276–283 (2010)

9.216 L.B. Du, A.J. Lowery: Optimizing the subcarrier granularity of coherent optical communications systems, Opt. Express **19**(9), 8079 (2011)

9.217 M. Qiu, Q. Zhuge, X. Xu, M. Chagnon, M. Morsy-Osman, D.V. Plant: Subcarrier multiplexing using DACs for fiber nonlinearity mitigation in coherent optical communication systems. In: *Proc. OFC 2014, San Francisco* (2014), paper Tu3J.2

9.218 A. Bononi, N. Rossi, P. Serena: Performance dependence on channel baud-rate of coherent single-carrier WDM systems. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), London* (2013), paper Th.1.D.5

9.219 P. Poggiolini, A. Nespola, Y. Jiang, G. Bosco, A. Carena, L. Bertignono, S.M. Bilal, S. Abrate, F. Forghieri: Analytical and experimental results on system maximum reach increase through symbol rate optimization, J. Lightwave Technol. **34**(8), 1872–1885 (2016)

9.220 A. Carbó, J. Renaudier, P. Tran, G. Charlet: Experimental analysis of non linear tolerance dependency of multicarrier modulations versus number of WDM channels. In: *Proc. OFC 2013, Anaheim* (2016), paper Tu3A.6

9.221 C.S. Gardner, J.M. Greene, M.D. Kruskal, R.M. Miura: Method for solving the Korteweg–de Vries equation, Phys. Rev. Lett. **19**(19), 1095–1097 (1967)

9.222 V.E. Zakharov, A.B. Shabat: Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of wave in nonlinear media, Sov. J. Exp. Theor. Phys. **34**(1), 62–69 (1972)

9.223 S.V. Manakov: On the theory of two-dimensional stationary self-focusing of electromagnetic waves, Sov. J. Exp. Theor. Phys. **38**, 248–253 (1974)

9.224 M.J. Ablowitz, H. Segur: *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia 1981)

9.225 M.I. Yousefi, F.R. Kschischang: Information transmission using the nonlinear Fourier transform, Part I–III, IEEE Trans. Inf. Theory **60**(7), 4312–4369 (2014)

9.226 S.K. Turitsyn, J.E. Prilepsky, S. Le Thai, S. Wahls, L.L. Frumin, M. Kamalian, S.A. Derevyanko: Nonlinear Fourier transform for optical data processing and transmission: advances and perspectives, Optica **4**(3), 307–322 (2017)

9.227 A. Hasegawa, Y. Kodama: *Solitons in Optical Communications* (Clarendon, Oxford 1995)

9.228 A. Hasegawa, T. Nyu: Eigenvalue communication, IEEE J. Lightwave Technol. **11**(3), 395–399 (1993)

9.229 H. Terauchi, A. Maruta: Eigenvalue modulated optical transmission system based on digital coherent technology. In: *OptoElectron. Commun. Conf. (OECC), Kyoto, paper WR2–5* (2013)

9.230 J.E. Prilepsky, S.A. Derevyanko, K.J. Blow, I. Gabitov, S.K. Turitsyn: Nonlinear inverse synthesis and eigenvalue division multiplexing in optical fiber channels, Phys. Rev. Lett. **113**, 013901 (2014)

9.231 S.A. Derevyanko, J.E. Prilepsky, S.K. Turitsyn: Capacity estimates for optical transmission based on the nonlinear Fourier transform, Nat. Commun. **7**, 12710 (2016), https://doi.org/10.1038/ncomms12710

9.232 S. Civelli, E. Forestieri, M. Secondini: Why noise and dispersion may seriously hamper nonlinear frequency-division multiplexing, IEEE Photonics Technol. Lett. **29**(16), 1332–1335 (2017)

9.233 E.G. Turitsyna, S.K. Turitsyn: Digital signal processing based on inverse scattering transform, Opt. Lett. **38**(20), 4186–4188 (2013)

9.234 H. Bülow: Experimental demonstration of optical signal detection using nonlinear Fourier transform, J. Lightwave Technol. **33**(7), 1433–1439 (2015)

9.235 H. Bülow, V. Aref, W. Idler: Transmission of waveform determined by 7 eigenvalues with PSK-modulated spectral amplitude. In: *Europ. Conf. Opt. Commun. (ECOC), paper Tu.3.E.2, Germany* (2016)

9.236 S. Civelli, E. Forestieri, M. Secondini: Decision-feedback detection strategy for nonlinear frequency-division multiplexing, Opt. Express **26**(9), 12057–12071 (2018)

9.237 V. Aref, S. Le Thai, H. Bülow: Demonstration of fully nonlinear spectrum modulated system in the highly nonlinear optical transmission regime. In: *Proc. Europ. Conf. Opt. Commun. (ECOC), postdeadline paper, Germany* (2016)

9.238 A. Maruta, Y. Matsuda: Polarization division multiplexed optical eigenvalue modulation. In: *Int. Conf. Photonics Switch. (PS), Florence* (2015) pp. 265–267

9.239 S. Gaiarin, A.M. Perego, E.P. da Silva, F. Da Ros, D. Zibar: Experimental demonstration of dual polarization nonlinear frequency division multiplexed optical transmission system. In: *Proc. Europ. Conf. Opt. Commun., Sweden* (2017)

9.240 S. Wahls, H.V. Poor: Fast numerical nonlinear Fourier transforms, Trans. Inf. Theory **61**(12), 6957–6974 (2015)

9.241 M. Kamalian, J.E. Prilepsky, S.T. Le, S.K. Turitsyn: Periodic nonlinear Fourier transform for fiber-optic communications, Part I–II, Opt. Express **24**(16), 18353–18381 (2016)

9.242 I.T. Lima, T.D.S. DeMenezes, V.S. Grigoryan, M. O'Sullivan, C.R. Menyuk: Nonlinear compensation in optical communications systems with normal dispersion fibers using the nonlinear Fourier transform, J. Lightwave Technol. **35**(23), 5056–5068 (2017)

9.243 P. Serena, A. Ghazisaeidi, A. Bononi: A new fast and blind cross-polarization modulation digital compensator. In: *Proc. Europ. Conf. Opt. Commun. (ECOC)* (2012)

9.244 K. Kojima, D.S. Millar, T. Koike-Akino, K. Parsons: Constant modulus 4D optimized constellation alternative for DP-8QAM. In: *Proc. Europ. Conf. Opt. Commun. (ECOC)* (2014)

9.245 T. Fehenberger, A. Alvarado, G. Bocherer, N. Hanik: On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel, J. Lightwave Technol. **34**, 5063–5073 (2016)

9.246 O. Geller, R. Dar, M. Feder, M. Shtaif: A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems, J. Lightwave Technol. **34**, 3884–3889 (2016)

9.247 X. Liu, A. Chraplyvy, P. Winzer, R. Tkach, S. Chandrasekhar: Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit, Nat. Photonics **7**, 560–568 (2013)

9.248 H. Eliasson, P. Johannisson, M. Karlsson, P.A. Andrekson: Mitigation of nonlinearities using conjugate data repetition, Opt. Express **23**, 2392–2402 (2015)

9.249 C.E. Shannon: A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379–423 (1948)

9.250 S. Verdú: Fifty years of Shannon theory, IEEE Trans. Inf. Theory **44**(6), 2057–2078 (1998)

9.251 E. Agrell, M. Karlsson, A.R. Chraplyvy, D.J. Richardson, P.M. Krummrich, P. Winzer, K. Roberts, J.K. Fischer, S.J. Savory, B.J. Eggleton, M. Secondini, F.R. Kschischang, A. Lord, J. Prat, I. Tomkos, J.E. Bowers, S. Srinivasan, M. Brandt-Pearce, N. Gisin: Roadmap of optical communications, J. Opt. **18**(6), 063002 (2016)

9.252 R.G. Gallager: *Information Theory and Reliable Communication* (Wiley, Hoboken 1968)

9.253 F.D. Neeser, J.L. Massey: Proper complex random processes with applications to information theory, IEEE Trans. Inf. Theory **39**(4), 1293–1302 (1993)

9.254 G. Kramer, M. Yousefi, F. Kschischang: Upper bound on the capacity of a cascade of nonlinear and noisy channels. In: *Proc. IEEE Inf. Theory Workshop, Apr. 2015* (2015) pp. 1–4

9.255 E. Agrell, M. Karlsson: Influence of behavioral models on multiuser channel capacity, J. Lightwave Technol. **33**(17), 3507–3515 (2015)

9.256 J. Dauwels, H.-A. Loeliger: Computation of information rates by particle methods, IEEE Trans. Inf. Theory **54**, 406–409 (2008)

9.257 A.D. Ellis, J. Zhao, D. Cotter: Approaching the nonlinear Shannon limit, J. Lightwave Technol. **28**(4), 423–433 (2010)

9.258 R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel: Capacity limits of optical fiber networks, J. Lightwave Technol. **28**(4), 662–701 (2011)

9.259 G. Bosco, P. Poggiolini, A. Carena, V. Curri, F. Forghieri: Analytical results on channel capacity in uncompensated optical links with coherent detection, Opt. Express **19**(26), B438–B449 (2011)

9.260 M. Secondini, E. Forestieri: Scope and limitations of the nonlinear Shannon limit, J. Lightwave Technol. **35**(4), 893–902 (2017)

9.261 R. Dar, M. Shtaif, M. Feder: New bounds on the capacity of the nonlinear fiber-optic channel, Opt. Lett. **39**(2), 398–401 (2014)

9.262 D. Marsella, M. Secondini, E. Agrell, E. Forestieri: A simple strategy for mitigating XPM in nonlinear WDM optical systems. In: *Proc. Opt. Fiber Commun. Conf. Exhib. 2015, Paper Th4D.3* (2015)

9.263 E. Agrell: Conditions for a monotonic channel capacity, IEEE Trans. Commun. **63**(3), 738–748 (2015)

9.264 M. Sorokina, S. Sygletos, S. Turitsyn: Ripple distribution for nonlinear fiber-optic channels, Opt. Express **25**(3), 2228–2238 (2017)

9.265 K.S. Turitsyn, S.A. Derevyanko, I.V. Yurkevich, S.K. Turitsyn: Information capacity of optical fiber channels with zero average dispersion, Phys. Rev. Lett. **91**, 203901-1–203901-4 (2003)

9.266 M.I. Yousefi, F.R. Kschischang: On the per-sample capacity of nondispersive optical fibers, IEEE Trans. Inf. Theory **57**(11), 7522–7541 (2011)

9.267 M.I. Yousefi: The asymptotic capacity of the optical fiber, arXiv:1610.06458 [cs.IT] (2016)

Alberto Bononi

Dipt. di Ingegneria e Architettura
Università degli Studi di Parma
Lesignano de' Bagni (Parma), Italy
alberto.bononi@unipr.it

Alberto Bononi received his PhD from Princeton University in 1994. He was an assistant professor at SUNY Buffalo before joining Parma University, Italy, in 1996, where he now serves as Full Professor in Telecommunications. His main research is in optical networks and optical transmission performance modeling.

Ronen Dar

Optical transmission systems
Nokia Bell Labs
Holmdel, NJ, USA
ronendar@yahoo.com

Ronen Dar received his PhD from Tel-Aviv University in 2015. During 2015–2017 he was a Member of Technical Staff in the optical transmission systems group at Nokia Bell Labs. He is coauthor of several patents in the fields of communication and coding and the recipient of several academic awards, including the Israeli Academy of Sciences and Humanities Adam's Fellowship.

Marco Secondini

Institute of Communication, Information and Perception Technologies
Sant'Anna School of Advanced Studies
Pisa, Italy
marco.secondini@sssup.it

Marco Secondini received his PhD from Scuola Superiore Sant'Anna, Pisa, Italy, in 2006. He visited the Photonics Group of the University of Maryland Baltimore County in 2005. He is currently an assistant professor at Scuola Superiore Sant'Anna and collaborates with the Photonic National Laboratory of the CNIT. His research is focused on the theoretical aspects of optical fiber communications.

Paolo Serena

Dipt. di Ingegneria e Architettura
Università degli Studi di Parma
Parma, Italy
paolo.serena@unipr.it

Paolo Serena received his PhD from the University of Parma, Italy, in 2003. He has been Assistant Professor at the same university since 2005. His main research interest is in optical communications, including fiber-optic transmission modeling in the nonlinear regime, numerical methods for efficient simulations, and performance estimation methods.

Pierluigi Poggiolini

Dept. of Electronics and Telecommunications (DET)
Politecnico di Torino
Torino, Italy
pierluigi.poggiolini@polito.it

Pierluigi Poggiolini recevied his MS (1988) and his PhD (1993) from Politecnico di Torino. He did his postdoc at Stanford University (1994–1995). He is currently Full Professor at Politecnico di Torino and is co-recipient of two best paper awards (IEEE/OSA Journal of Lightwave Technology, 2013 and 2014). His current interest lies in modeling of nonlinear effects in WDM coherent optical systems and application of neural networks to optical communications.

10. Space-Division Multiplexing

Roland Ryf, Cristian Antonelli

Fiber-based optical communication networks are reaching a point where the capacity required on a single link can significantly exceed the capacity of a single-mode fiber, and at the same time conventional network architectures can no longer be scaled cost-effectively. Space-division multiplexing (SDM) addresses the capacity bottleneck imposed by the use of single-mode fibers within a completely new approach that relies on new fiber types, optical amplifiers, and optical switches capable of supporting multiple spatial channels.

The aim of this chapter is, on one hand, to provide an overview of the components that are necessary for the implementation of SDM transmission and, on the other hand, to review the modeling of the main propagation effects that occur in multimode and multicore fibers. The chapter also includes a description of the techniques that are used in SDM transmission experiments and an update on transmission records reported from around the globe. The chapter ends with the description of potential architectures supporting SDM networks.

Optical communications over fiber networks are the backbone of our current communication infrastructure, and the essential network components and subsystems have been described in detail in Part A of this Springer Handbook. Fiber-optic networks have scaled in capacity over the years by first adding multiple wavelength channels, and most recently by increasing the spectral efficiency of the transmitted signals through advanced modulation techniques enabled by the introduction of digital coherent transceivers [10.1]. This allowed to use pulse shaping in combination with multilevel complex formats to optimize the use of the optical spectrum, as well as transmitting independent signals on two orthogonal polarizations of the electric field propagating in a single-mode fiber (SMF) (See Chaps. 4–7 for a detailed description). In this way, a single SMF is able to provide around 50 Tb/s for distances up to 10 000 km. This high capacity, which

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_10

was once considered much higher than required for telecommunications, is now becoming the bottleneck of the cloud-centric information-technology communication networks. It is therefore of paramount importance to identify and develop new technologies to increase the capacity of fiber-optic communication systems by 2–3 orders of magnitude, while simultaneously reducing the cost per bit of the transmitted data. As current state-of-the-art fiber-optic systems operate close to the theoretical capacity limit [10.1], this can only be achieved by increasing the number of parallel spatial channels, which is the aim of research in space-division multiplexing (SDM) [10.2–4].

The simplest approach to overcome the capacity limit of SMFs is using traditional single-mode systems operated in parallel. However, this approach results in costly and inflexible optical networks.

Significant cost reductions can be expected by better integrating components at various levels, in particular by sharing components whenever possible or placing multiple devices in a single planar waveguide circuit or in a common free-space subsystem.

Integration offers great opportunities for cost reduction, most remarkably for transponders, which represent a significant expense in a transmission system and require large numbers of wavelength and spatial channels (presently an independent line card is used for each wavelength channel).

Optical amplifiers supporting multiple spatial channels also offer a great potential for integration: starting from simple parallel erbium-doped amplifiers sharing some components and control boards, to advanced pump-sharing schemes like for example cladding pumped amplifiers, where a common pump is guided in the fiber cladding and the signals to be amplified are guided in the fiber cores.

Optical switches, which are the basic building blocks of optical networks, can also be modified to support systems with multiple spatial channels. In particular, the concept of *joint switching* means a beam-steering-based switching element (for example microelectro-mechanical system (MEMS) mirror arrays, or liquid crystal on silicon (LCOS) pixel arrays) can be used to switch multiple spatial channels simultaneously.

At the physical transmission level, consisting of the optical fiber, it is also possible to utilize waveguide structures that support multiple spatial modes, either by using single cores that support multiple modes, or by adding multiple cores in the common cladding of an optical fiber. While the single-mode fiber transmission channel is very well known in terms of both the linear (chromatic dispersion, polarization-mode dispersion, and Rayleigh scattering) and nonlinear (Kerr effect, stimulated Raman scattering, and stimulated Brillouin scattering) effects, the characterization of the same propagation effects can become fairly complex if multiple modes are involved, as all possible intermodal interactions have to be accounted for. Modeling propagation effects in fibers supporting multiple modes is an imperative step towards the assessment of the transmission capacity in SDM systems.

10.1 Basic Concepts in Space-Division Multiplexed Optical Networks

Adding spatial channels to traditional wavelength-division multiplexed (WDM) networks can significantly affect the overall network complexity, as both the wavelength and space dimensions can be used for the routing of optical signals. Current network architectures are also often constrained by reliability requirements: for example, networks are designed such that a single failure at any component level can always be overcome by either a protection scheme (where a second independent physical path is available) or a restoration scheme (where a second independent path is configured when needed). Using multiple independent spatial paths provides more flexibility to overcome failure by appropriately designing and operating the network.

Finally, at a physical point-to-point link level, fibers supporting multiple light paths, like multimode and multicore fibers, can be utilized to reduce the component count by integrating functionalities in the spatial domain.

Routing in optical networks, network reliability, and point-to-point link engineering are not independent of each other, and the three domains have to be optimized jointly, which makes SDM networking a formidable challenge with no obvious winning network architecture, but rather with multiple possible solutions that depend on the specific network requirements.

In this chapter, we focus on SDM point-to-point links (Sect. 10.2) and routing in SDM networks (Sect. 10.8), as these two topics are key for understanding SDM networks.

10.2 SDM Point-to-Point Links

An SDM point-to-point link is schematically described in Fig. 10.1. It includes the following main components: A bank of transmitters $\mathrm{Tx}_{n,m}$, a multiplexing device that encodes N spatial channels and M wavelength channels into the physical spatial and wavelength channels, an SDM fiber possibly followed by K sections, each consisting of an optical amplifier and an SDM fiber span. At the end of the transmission fiber, a demultiplexing device splits all transmitted channels by spatial and wavelength channels, and the individual channels are received by a bank of single-mode receivers $\mathrm{Rx}_{n,m}$. The subscripts $n \leq N$ and $m \leq M$ in Fig. 10.1 are the indices of the spatial and wavelength channels, respectively.

Note that spatial and wavelength channels are not equivalent: Wavelength channels are formed by arbitrarily allocating portions of the continuous optical spectrum, whereas spatial channels are in a one-to-one correspondence with a discrete set of modes or cores of the transmission fiber. This has important implications for optical switching as we will see in Sect. 10.8. Additionally, there is also a fundamental difference between spatial and wavelength channels concerning crosstalk: Optical fibers do not introduce linear crosstalk between different wavelength channels (nonlinear effects can potentially do so, but we will neglect nonlinear effects in this stage), whereas linear crosstalk can be present between spatial channels, particularly if the spatial channels are arranged very closely, like in multicore fibers, where light can couple between neighboring cores, or in multimode fibers, where the modes spatially overlap and coupling can be produced by fiber imperfections. In practice, the presence of crosstalk between spatial channels, is not detrimental for optical transmission. In fact, the crosstalk between spatial channels can be undone using digital signal processing (DSP) techniques, which are similar to multiple-input-multiple-output (MIMO) algorithms used in wireless networks. A major advantage of optical fibers, as compared to the wireless channel, is that they have very low loss, and often also very low difference in loss between modes, which results in an almost unitary MIMO channel, which according to information theory, provides N times the capacity of a single channel, where N is the number of spatial channels. MIMO transmission is particularly effective if the received signals are detected using polarization-diverse coherent receivers (PD-Rxs), which are able to measure amplitude and phase for both polarizations of the received spatial channels. In fact, in the case where all modes guided by the fiber under test are detected using PD-Rxs, the optical field impinging upon the receiver is fully known and propagation-induced impairments (most effectively unitary linear effects) can be compensated for in the digital domain. The MIMO DSP that is necessary to undo linear impairments, is a generalization of the 2×2 MIMO implemented in commercially available digital coherent transceivers (Chap. 6), where $2N\times2N$ MIMO is required for N coupled spatial modes, and additionally a larger number of equalizer memory (taps) is needed owing to the fact that, as discussed in the next section, the group velocity difference between modes in multimode optical fibers is several orders of magnitude larger than the group velocity difference between the two polarizations of a single-mode fiber. The design of SDM fibers is therefore often optimized to make the magnitude of the group velocities of all the fiber modes as close as possible. Note that in the case of multimode fiber, this optimization is very similar to that performed on commercial OM3 and OM4 multimode fibers used for short-reach interconnects, where the difference in group velocities between modes can limit transmission reach and bandwidth [10.5–7].

Coherent MIMO transmission over SDM fiber can maximize the transmission capacity of the fiber close to the theoretical limit imposed by information theory under any crosstalk condition. The crosstalk in the n-th mode $X_\mathrm{t}(n)$ is defined as

$$X_\mathrm{t}(n) = \frac{\sum_{m\neq n} P_{m,n}}{P_{n,n}}\,, \tag{10.1}$$

where $P_{m,n}$ is the power transfer matrix between an excited mode m and a mode n received after transmission

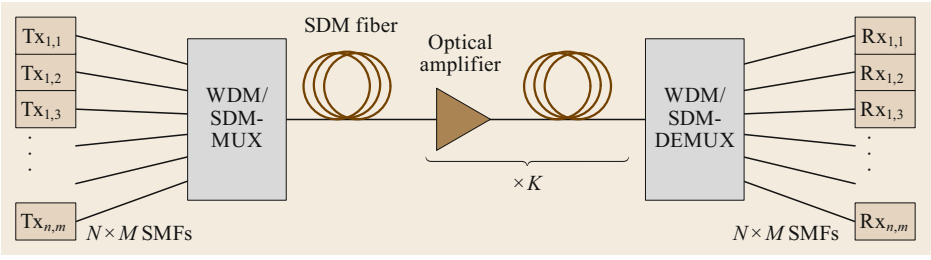


Fig. 10.1 Point-to-point combined WDM/SDM transmission

Table 10.1 Maximum acceptable crosstalk level for an additive system penalty obtained at a BER of 10^{-3} for quadrature phase-shift keying (QPSK) modulation and various quadrature-amplitude modulation (QAM) formats (after [10.8])

Acceptable system penalty	Max crosstalk (dB)			
	QPSK	16 QAM	64 QAM	256 QAM
1 dB	−17	−23	−29	−35
3 dB	−12	−18.5	−25	−31

through the fiber. The added complexity of the MIMO DSP, even if within the realm of modern ASIC technology, is often undesirable, and alternatively SDM fibers can be designed so that the crosstalk between spatial channels is reduced to levels where electronic crosstalk mitigation is no longer required. In Table 10.1 we report the maximum acceptable crosstalk level for an added system penalty (degradation of the quality factor Q of the transmitted signal) of 1 and 3 dB, respectively, at a bit-error ratio (BER) of 10^{-3} [10.8], which using state-of-the-art hard-decision forward error correction (FEC) is sufficient to obtain error-free performance (post-FEC BER $< 10^{-12}$).

This is of particular interest in the case of a system based on the use of multicore fibers where distances over 10 000 km have been achieved without MIMO processing. We will refer to this technique as transmission in the *low-crosstalk regime*. Note that the low-crosstalk regime is challenging not only for the fiber, but also for all the in-line optical components, connectors and splices, which also have to meet the low-crosstalk requirement. This is sometimes a limiting factor for the level of integration that can be achieved.

Alternatively, it is also possible to combine MIMO-based transmission with the low-crosstalk regime as not all modes will show the same amount of coupling between each other and the total number of spatial channels can then be separated into groups, where strong coupling is present within the groups and little crosstalk is observed between the groups. This technique is used in multimode fibers for distances smaller than 100 km and also in multicore fibers with few-mode cores (few-mode multicore fibers (FM-MFCs)), where MIMO is performed across the core modes of each core but not across the core modes of different cores, and distances up to 1000 km have been demonstrated [10.9, 10].

In the following sections, we discuss in more detail the components of a point-to-point SDM link.

10.2.1 SDM Fibers

The single-mode fiber has been the work horse of high-capacity long-distance communications for over three decades. It is also the most common and lowest-cost glass fiber (Chap. 2), mostly because of its favorable optical properties, like low loss and large bandwidth (tens of THz). A common way to increase the capacity of optical links is therefore the use of single-mode fiber ribbons. Fiber ribbons can be spliced with commercial ribbon splicer and containerized, therefore offering advantages in high fiber-count cables, and cables with over 3000 fibers are commercially available. Recently, it has been proposed to increase the capacity of the fiber-optic channel using the following fiber types:

- Multicore fibers: fibers with multiple cores
- Multimode fibers: fibers with cores supporting multiple modes.

Both fiber types are well known and have been proposed for various applications, like sensing, endoscopy, and short-reach interconnects, but a significant effort was recently devoted to optimizing the fiber design to support long-distance transmission.

Multicore Fibers

Since low-crosstalk multicore fibers (MCFs) have been successfully engineered, various techniques to further reduce the crosstalk between cores have been investigated. These include, for example, the use of trenches and holes around the cores and the use of heterogeneous cores, where neighboring cores are designed with different propagation constants. Additionally, bidirectional transmission can be used, where neighboring cores carry signals propagating in opposite directions. Nevertheless, the maximum number of cores in multicore fibers is limited by the cladding diameter. Increasing the cladding diameter substantially above the standard diameter of 125 µm results in more fragile fibers with reduced reliability, and practical cladding diameters for long-distance communication are limited to around 250 µm maximum cladding diameter. Additionally, larger core density can be achieved in multicore fibers by reducing the effective area of the cores, which however reduces the transmission performance of the fiber. Multicore fibers with more than 30 cores [10.11] have been demonstrated for a distance up to 1645 km ultrahigh-capacity transmission, and longer distances can be achieved by reducing the core count to 12 cores, where distances up to 8800 km have been demonstrated [10.12]. Multicore fibers can be spliced using splicers designed to support polarization-maintaining fibers, but the nonstandard cladding diameters require customized fiber holders. Also, the core alignment is typically less accurate than in the case of single-mode fiber splices, particularly for cores located far from the fiber center, which are susceptible to fiber rotation er-

rors, resulting in slightly larger splicing losses (up to 1 dB). Multicore fibers can be connectorized and prototype connectors have been demonstrated [10.13].

Alternative multicore fibers can be operated in the so-called *coupled-core* regime, where the distance between cores is made small, such that the cores become coupled. The coupled-core multicore fiber (CC-MCF) will therefore behave more like a microstructured fiber supporting multiple modes. It is important to point out that coupled-core fibers show an optimum core spacing, where the properties of the modes are such that the fiber perturbations cause strong random coupling between the modes, while the modal group-velocity difference is still moderately small. This effect is described in detail in Sect. 10.3.3.

Multimode Fibers

Multimode fibers (MMFs) consist of fibers with a single core that is either larger in diameter or has a larger refractive index compared to an SMF, such that the fiber can support more than one mode. The modes can then be potentially used as independent transmission channels, and fibers with 100 or more modes can be easily produced. Various refractive index designs of the core have been proposed to support SDM, like step-index and multistep index profiles, graded-index profiles, and ring-core fibers.

Two design strategies have been pursued for multimode fibers. The first strategy is to minimize the group delay difference between the supported fiber modes. For few modes this is possible by using multistep index fibers, like for example a depressed-cladding design in the case of three spatial modes [10.14]. For larger numbers of modes, cores with a graded-index profile provide an optimum solution, and fibers with up to 45 modes at 1550 nm wavelength have been demonstrated using 50 μm core diameters [10.5–7]. The fibers are very similar in design and manufacturing to commercially available graded-index fiber optimized for 850 nm wavelength and referred to as OM3 and OM4 fibers in commercial products, which are amply used for short-reach interconnects, for example in datacenters. Graded-index multimode fibers have a typical residual group delay difference of around 100 ps/km [10.15, 16], which for long-distance communication that can reach up to 10 000 km would result in up to 1 μs of maximum delay between modes. Fortunately, the crosstalk between modes strongly reduces the effect as explained in Sect. 10.5.1, such that impulse response widths in the order of tens of nanoseconds are experimentally observed. Using low group delay multimode fibers, transmission distances up to 4500 km have been demonstrated using 3 spatial modes [10.17], whereas experiments with up to 45 spatial modes [10.18, 19] have been reported for shorter fiber lengths. The second design strategy aims to reduce the coupling between modes and use the modes as independent transmission channels. The main goal in this case is to increase the difference in phase velocity between modes by optimizing the index profile. The phase velocity of a mode describes the speed at which the optical phase front travels in the fiber (a formal definition is given in Sect. 10.3). This approach can yield a significant crosstalk reduction, particularly between nearly degenerate modes, however fiber manufacturing imperfections, geometrical deformations like bending and twisting, and Rayleigh scattering limit the achievable crosstalk reduction, with the smallest reported crosstalk levels being of the order of −30 to −40 dB/km, and the longest transmission distance demonstrated being about 50 km. Low-crosstalk fibers are of interest for short-distance applications (< 100 km).

We note that alternative fiber designs aiming to avoid degeneracy between modes have been proposed, for example by using elliptical cores [10.20–22], or high-contrast ring-core fiber designs [10.23], which allows accessing and using all fiber modes without any MIMO digital signal processing.

10.2.2 Mode Multiplexers and Fiber Couplers

Mode multiplexers (MMUXs) are devices that couple multiple single-mode fibers into the modes of a multimode fiber. There are two types of MMUXs: The so-called *mode-selective device* directly couples a given single-mode fiber with a specific fiber mode, whereas the *nonmode-selective device*, associates different single-mode fiber with different orthogonal linear combinations of modes. For coherent MIMO transmission a nonmode-selective MMUX can be utilized with no disadvantage, whereas in some specific cases, like for example transmission with low-crosstalk between mode groups, or for modal delay compensation, mode-selective MMUXs with high modal selectivity are required.

Since fiber modes are orthogonal to each other (the definition of mode orthogonality is provided in Sect. 10.3), it is theoretically possible to separate modes in a lossless way, similarly to the way in which a polarizing beam splitter can separate two polarization components, or a diffraction grating can separate different wavelengths. Numerous techniques have been proposed and demonstrated for realizing mode multiplexers [10.25–33]. In what follows, we describe two of the most promising techniques.

The Photonic Lantern

Photonic lanterns (PLs) provide an adiabatic transition between N single-mode fibers and a step-index

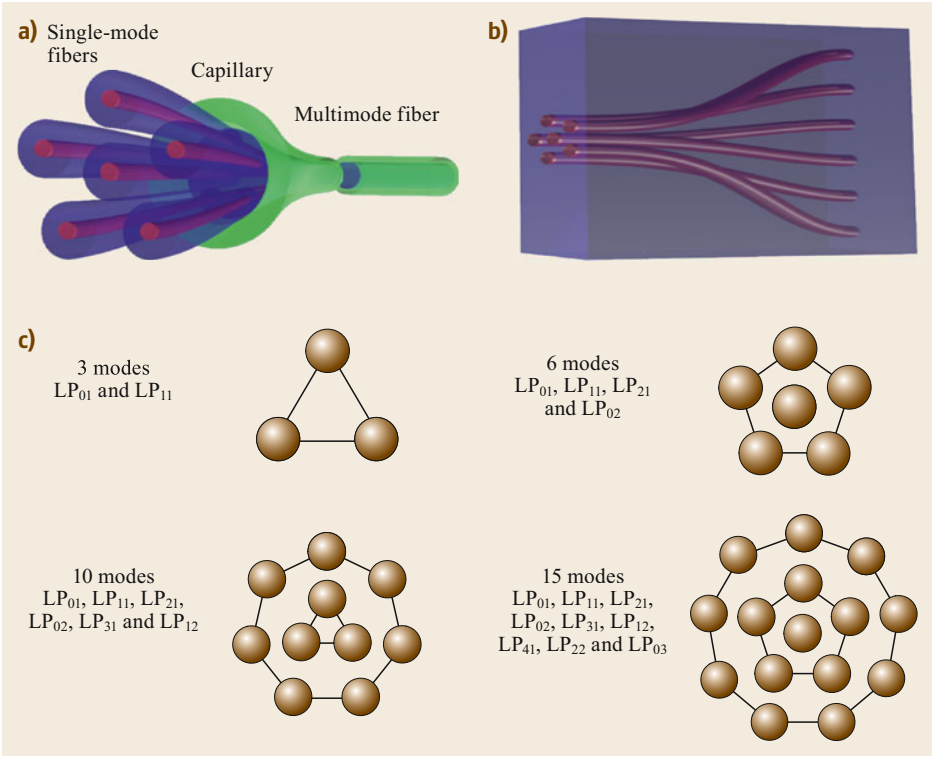


Fig. 10.2 (a) Principle of the fiber-based photonic lantern. (b) Photonic lantern implemented using laser-inscribed waveguides. (c) Core arrangement matching linearly polarized (LP) modes of multimode fibers with 3, 6, 10, and 15 spatial modes [10.24]

multimode fiber [10.24, 31, 32, 34–36]. This device is manufactured by using a glass processor starting from N single-mode fibers inserted in a low refractive index capillary. The composite structure is then continuously reduced in size (tapered) in the glass processor so that the cores of the single-mode fibers vanish and the claddings form the new core, whereas the low-refractive-index capillary forms the new cladding (Fig. 10.2a).

The end section of the PL can then be directly spliced to a multimode fiber with a matching core geometry. In order to map the correct modes between the single-mode and the multimode end, the SMFs have to be arranged into specific spatial patterns (Fig. 10.2c). Arranging the fibers according to these patterns is easily achieved for MMUXs up to six spatial modes, where the single-mode fibers self-arrange in the capillary. Photonic lanterns supporting a larger number of modes are fabricated more easily by using a drilled low-refractive-index preform instead of the capillary, where the fiber is held in the correct location by the drilled holes. This way MMUXs up to 15 spatial modes have been demonstrated [10.37]. Mode-selective PLs can be achieved by starting from nonidentical single-mode fibers, either by slightly changing the core diameter or the refractive index difference between the core and cladding of the single-mode fibers [10.31, 32, 38]. In this way, for example, the single-mode fibers with the lowest effective refractive index can be matched with the fiber mode that also has the lowest effective refractive index and a mapping between modes and the input SMF can be achieved. Photonic lanterns are of interest because devices with low insertion loss and low mode-dependent loss (MDL) can be achieved. For example, an MMUX supporting three spatial modes can have an insertion loss of less than 0.5 dB and less than 3 dB MDL (in short MDL is defined as the power ratio in dB between the least and most attenuated linear combinations of modes—see Sect. 10.5.2 for a more extended description). The corresponding values for six and ten modes are slightly larger, but PL technology is still the best performing in terms of loss.

Photonic lanterns can also be realized by using femto-second laser-inscribed three-dimensional waveguides (Fig. 10.2b), where waveguides are written into a glass substrate and brought close to each other arranged as shown in Fig. 10.2c, so that the waveguides almost merge [10.39]. In particular, there is an alternative design for the laser-inscribed PL, referred to as taper velocity couplers [10.40], where the inscribed waveguides are tapered within the coupling section. MMUXs fabricated this way are commercially available, and offer the advantage of being compatible with planar waveguide technology, as the input waveguide can be arranged in any desired geometry.

The Multiplane Light Conversion Mode Multiplexer

Multiplane light conversion has been proposed as a universal mode-converting device capable of transforming any input mode set into any output mode set [10.33]. This functionality is achieved by having the light traversing multiple phase-only transmission masks, each followed by a free-space transmission section. The number of required masks depends on the desired transformation, but fortuitously, one of the transformations of biggest interest in mode-division multiplexing, which is the transformation from spatially separated spots into Laguerre–Gaussian beams (which are in close approximation to the modes of a graded-index multimode fiber, as will be seen in the following sections), can be achieved with less than ten phase masks. Also, in practice, multiplane devices are implemented using a reflective geometry, where light is reflected multiple times between two planes: one plane consists of a substrate containing multiple phase masks, and the second is a high-reflectivity mirror. A schematic arrangement for a mode-multiplexer is shown in Fig. 10.3.

Using this technology, MMUXs with large mode selectivity have been realized with up to 45 spatial modes [10.41, 42], and proof-of-principle devices with up to 210 spatial modes using only 7 phase masks have been demonstrated [10.43]. Multiplane-based MMUXs offer great flexibility but up-to-date products and prototypes could not yet match the low-loss level of photonic lanterns. However, since theoretically they are lossless elements, lower loss devices can be expected by using a coating with higher reflectivity and improved mask fabrication and designs.

10.2.3 SDM Amplifiers

Optical amplifiers that support multiple spatial paths can be designed in several ways. A simple collocation of multiple erbium-doped fibers in a single module can already provide significant cost savings by sharing components and control electronics. Additionally, it is possible to share pump lasers and use arrays of integrated optics components like taps and detectors for power level monitoring. A higher level of integration is possible when using multimode or multicore erbium-doped fibers. It is then possible to extend the functionality of free-space optics-based components like for example optical isolators or gain-equalizing filters to support multiple modes or cores without the need for mode multiplexers or fan-in/fan-out devices in the case of multimode- and multicore fibers, respectively.

Multimode erbium-doped fibers, typically require a precisely controlled doping profile and/or specific modal excitation of the pump laser in order to achieve a similar gain for each mode of interest [10.44–47]. The situation is more advantageous in multicore doped fibers, where pump lasers are coupled into the individual doped core of the multicore fiber. The gain and noise figure of the core-pumped multicore amplifiers is expected to be comparable to traditional single-mode erbium-doped amplifiers, but the pump couplers will require at least one set of fan-in devices to couple the pump laser.

A particularly promising alternative approach to building optical amplifiers with multiple spatial channels is based on cladding pumping. In cladding pumping, a multimode high-power pump laser at 980 nm wavelength is coupled to the cladding of the amplifying fiber, which results in a homogeneous illumination of the entire fiber cross-section. The cores, which are erbium-doped, absorb the pump light and amplify the guided signals (Fig. 10.4).

The multimode pump can be coupled to the cladding by using the so-called *side-pumping* configuration, which consists of bringing a tapered multimode fiber carrying the pump light into contact with the external cladding of the amplifying fiber. Coupling efficiency of $> 80\%$ can be achieved without using traditional dichroic combiners.

Cladding pumping can be applied to both multicore and multimode fibers. In multicore fibers, all cores are illuminated by the pump light, which can produce homogeneous gain. This is advantageous compared to core pumping, where for each core the pump light has

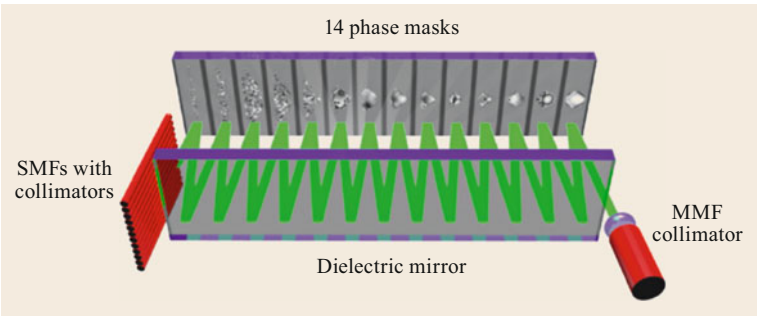


Fig. 10.3 Multiplane light converter configured as a mode multiplexer: Each single mode fiber (SMF) on the left-hand side will couple to a particular mode of the multimode fiber (MMF) on the right-hand side [10.41]

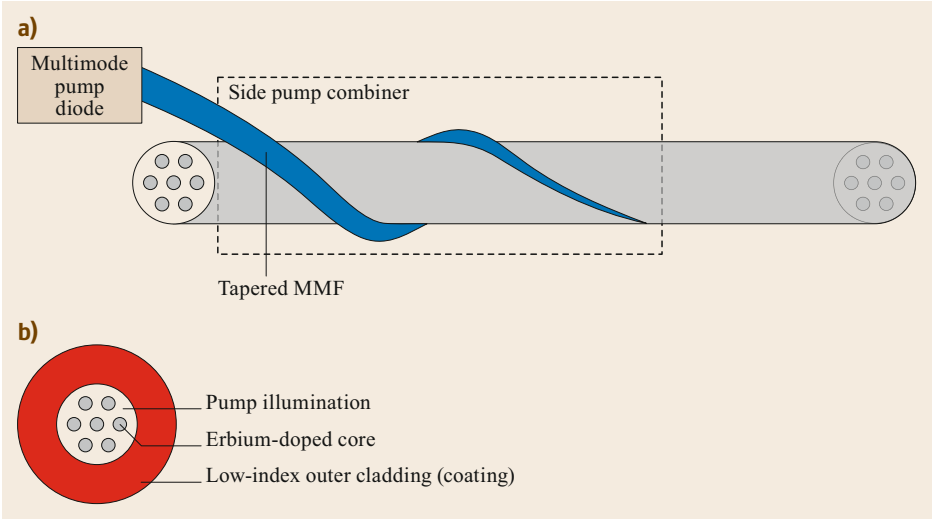


Fig. 10.4a,b Cladding pump amplifier scheme: **(a)** Side-pumping arrangement where a multimode pump is coupled into the cladding of a multicore fiber to homogeneously illuminate the cladding; **(b)** cross-section of multicore fiber showing the different fiber regions and the low-index coating used to confine the pump light

to be independently combined and coupled [10.48]. In order to minimize the gain variation between cores in cladding-pumped amplifiers, the core properties have to be precisely matched. Alternatively, variable gain attenuators (VOAs) acting on each core separately can be used in combination with a dynamic gain equalizing filter (DGEF), for example using an LCOS-based device [10.49]. Furthermore, cladding-pumped amplifiers, are typically operated in a nonsaturated output power regime, and therefore can naturally provide constant gain independently of the input power, which is desirable in optical networks to reduce the effect of transients caused by partial network failures. A drawback of operating the amplifier in a nonsaturated regime is however a reduced power efficiency of the cladding pump amplifier, which can in principle be overcome by means of pump-recycling schemes [10.50] able to reuse the pump light that is not absorbed by the cores at the end of the amplifying fiber (which otherwise is just dumped into free-space and absorbed).

Cladding pumping is advantageous also for multimode amplifiers: A homogeneous gain for all modes can be achieved by using a cladding pumping and in combination with a simple doping profile with constant erbium concentration across the whole core region [10.51, 52].

10.2.4 SDM Transceivers

Transceivers for SDM differentiate themselves from state-of-the-art single-mode digital coherent transceivers in two ways.

The first main difference is that additional DSP capabilities are required to support multiple modes, and a longer memory is necessary to accommodate larger group-delay differences in the MIMO DSP. Also, in order to simultaneously process multiple MIMO channels, a higher integration density of high-speed analog-digital converters (ADCs) is required. Note that the complexity of a conventional SMF digital coherent transceiver is dominated by the chromatic dispersion compensation (approximately 30%) and forward-error correction sections (approximately 25%), whereas the 2×2 MIMO section implies relatively little complexity (approximately 10%). Therefore, if the added complexity of the MIMO DSP grows hypothetically by one order of magnitude, the overall DSP complexity only increased by a factor of 2 per mode (see [10.53] for a detailed analysis).

The second main difference is that transceivers in a conventional WDM system require a tunable laser for each transceiver in order to cover the full wavelength channel map. Once the spatial dimension is added, it is in principle possible to share a single tunable laser among multiple spatial channels [10.54]. This is of particular interest in MIMO-based channels, where all spatial channels are treated as a single end-to-end MIMO transmission channel [10.55].

Additionally, the large count of basic components required in an SDM transceiver, like high-speed ADCs and digital-to-analog converters (DACs), modulators and optical coherent receivers, offers a great cost-saving potential through the integration of large pools of transceivers.

10.3 Fiber Modes

The modes of an optical fiber are the solutions of Maxwell's equations for electro-magnetic waves traveling along the propagation axis z of a dielectric waveguide characterized by some transversal refractive index profile. These solutions are invariant in the z-direction up to a phase term $\exp(\mathrm{j}\beta_n z)$, where n is the mode number and β_n is corresponding propagation constant. A single-mode electromagnetic field generated by a monochromatic source at the angular frequency ω_0 can be expressed as

$$\begin{aligned}\boldsymbol{E}_n(x,y,z,t) &= \mathrm{Re}\left[\boldsymbol{F}_n(x,y,\omega_0)\mathrm{e}^{\mathrm{j}(\beta_n z-\omega_0 t)}\right],\\ \boldsymbol{H}_n(x,y,z,t) &= \mathrm{Re}\left[\boldsymbol{G}_n(x,y,\omega_0)\mathrm{e}^{\mathrm{j}(\beta_n z-\omega_0 t)}\right],\end{aligned} \tag{10.2}$$

where $\boldsymbol{E}_n$ and $\boldsymbol{H}_n$ are real-valued three-dimensional vectors that represent the electric and magnetic field, respectively. The complex-valued vectors $\boldsymbol{F}_n$ and $\boldsymbol{G}_n$ are the corresponding lateral modal field distributions, and they are solutions of the Helmholtz equation in the form [10.56, 57]

$$\Delta_\mathrm{T}\boldsymbol{F}_n + \frac{\omega_0^2}{c^2}n^2(x,y)\boldsymbol{F}_n = \beta_n^2\boldsymbol{F}_n\,, \tag{10.3}$$

$$\Delta_\mathrm{T}\boldsymbol{G}_n + \frac{\omega_0^2}{c^2}n^2(x,y)\boldsymbol{G}_n = \beta_n^2\boldsymbol{G}_n\,, \tag{10.4}$$

where $n(x,y)$ is the transversal refractive index profile, c is the speed of light in a vacuum, and $\Delta_\mathrm{T} = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the transversal Laplace operator. Note that the first of the two equations is only exact if $\nabla\cdot\boldsymbol{E} = 0$, whereas in general $\nabla\cdot\boldsymbol{E} = -2\nabla\left[\log(n)\right]\cdot\boldsymbol{E}$, as follows from $\nabla\cdot\boldsymbol{D} = 0$ ($\boldsymbol{D}$ is the electric displacement vector). The condition $\nabla\cdot\boldsymbol{E} = 0$ is fulfilled exactly for step-wise constant refractive index profiles, and only approximately for slowly varying refractive index profiles.

In most cases that are of practical relevance, the fiber modes have to be calculated with numerical methods, and they fulfill the orthogonality condition [10.56]

$$\iint \mathrm{d}x\mathrm{d}y\left(\boldsymbol{F}_n\times\boldsymbol{G}_m^*\right)\cdot\hat{\boldsymbol{z}} = 2\mathcal{N}_n^2\delta_{n,m}\,, \tag{10.5}$$

where $\hat{\boldsymbol{z}}$ is a unit vector pointing in the propagation direction, $\delta_{n,m}$ is the Kronecker delta, and $\mathcal{N}_n$ is a normalization factor that is discussed later in the chapter. The orthogonality condition is used to decompose any electric and magnetic field at the fiber input characterized by the lateral profiles $\boldsymbol{\mathcal{F}}_\mathrm{fs}(x,y)$ and $\boldsymbol{\mathcal{G}}_\mathrm{fs}(x,y)$, respectively, as a linear combination of the fiber modes by using the following relations

$$\begin{aligned}\boldsymbol{\mathcal{F}}_\mathrm{fb}(x,y) &= \sum_{n=1}^{N} a_n\boldsymbol{F}_n(x,y)\,,\\ \boldsymbol{\mathcal{G}}_\mathrm{fb}(x,y) &= \sum_{n=1}^{N} a_n\boldsymbol{G}_n(x,y)\,,\\ a_n &= \iint \mathrm{d}x\mathrm{d}y\frac{\boldsymbol{\mathcal{F}}_\mathrm{fs}\times\boldsymbol{G}_n^*}{2\mathcal{N}_n^2}\cdot\hat{\boldsymbol{z}}\\ &= \iint \mathrm{d}x\mathrm{d}y\frac{\boldsymbol{F}_n^*\times\boldsymbol{\mathcal{G}}_\mathrm{fs}}{2\mathcal{N}_n^2}\cdot\hat{\boldsymbol{z}}\,,\end{aligned} \tag{10.6}$$

where the subscripts fs and fb refer to the input field in free space and to the guided field in the fiber, respectively.

Different modes have different propagation constants, unless they are degenerate. It is customary to represent each propagation constant as a Taylor expansion around the central frequency ω_0

$$\begin{aligned}\beta_n = \beta_{n,0} &+ \beta_{n,1}(\omega-\omega_0) + \frac{1}{2}\beta_{n,2}(\omega-\omega_0)^2\\ &+ \frac{1}{6}\beta_{n,3}(\omega-\omega_0)^3 + \ldots\end{aligned} \tag{10.7}$$

The different terms have the following physical interpretation. The first term is related to the phase velocity $v_{\mathrm{ph},n}$,

$$v_{\mathrm{ph},n} = \frac{\omega_0}{\beta_{n,0}}\,, \tag{10.8}$$

which describes the speed of the phase front of the propagating field. When two modes travel at the same phase velocity, energy can be exchanged and the coupled light is kept in phase over a long propagation distance. For pairs of modes with different phase velocities it is possible to define a beat length,

$$L_{n.m} = \frac{2\pi}{|\beta_{n,0}-\beta_{m,0}|}\,, \tag{10.9}$$

so that the accumulated phase difference between the two modes is $\Delta\phi = 2\pi z/L_{n,m}$, and the two modes are periodically in phase with the period being equal to the beat length. The second term of (10.7) yields the group velocity

$$v_{\mathrm{gr},n} = \frac{1}{\beta_{n,1}}\,, \tag{10.10}$$

which describes the speed of a light pulse traveling in mode n. The arrival time difference for two pulses trav-

eling in modes n and m, respectively, of a fiber of length ℓ is known as the differential group delay (DGD)

$$\mathrm{DGD}_{n,m} = \ell|\beta_{n,1} - \beta_{m,1}| \,. \tag{10.11}$$

The third term of (10.7) is related to the chromatic dispersion coefficient $\mathrm{CD}_n = (2\pi c/\lambda_0^2)\beta_{n,2}$, $\lambda_0 = 2\pi c/\omega_0$. Chromatic dispersion is responsible for the intramodal frequency dependence of the group velocity, which causes temporal pulse broadening even in the case when only a single mode is transmitted.

10.3.1 Modes of a Step-Index Fiber

The simplest optical fibers are based on a step-index profile, where the refractive index in polar coordinates is given by

$$n(r) = \begin{cases} n_c, & \text{if } 0 < r \le a \,, \\ n_0, & \text{if } r > a \,, \end{cases} \tag{10.12}$$

where a is the fiber core radius, and $r = \sqrt{x^2 + y^2}$. For step-index fibers the solution of (10.3) and (10.4) is known to be given by the Bessel function $J_n(r)$ in the core region, and by the modified Bessel function of the second kind $K_n(r)$ in the cladding region. Even if the solutions are known within each refractive index region, numerical methods have to be used to solve the boundary condition problem between core and cladding. The solutions for the first four mode groups of a step-index fiber are shown in Fig. 10.5.

The fundamental mode HE_{11} is degenerate in polarization, meaning that two modes with orthogonal polarizations, but with the same lateral profile and propagation constants are supported by the fiber. The next modes TM_{01} and TE_{01} are nondegenerate, and their polarization is position-dependent. The following HE_{21} mode is two-fold degenerate and can be represented as a linear combination of polarizations, by using circularly polarized light resulting in a ring-looking intensity profile (similar to TM_{01} and TE_{01}), but with a phase term of the form $\exp(\pm \mathrm{j}\varphi)$, where φ is the angle in polar coordinates. We note that fiber modes are often referred to as optical angular momentum (OAM) modes [10.23]. However, not all fiber modes can be represented as OAM modes, in particular TM_{0n} and TE_{0n} are not compatible with OAM modes [10.58], and also the concept of OAM modes breaks down when the effective index contrast is such as to break the degeneracy of the HE_{nm} modes [10.59, 60].

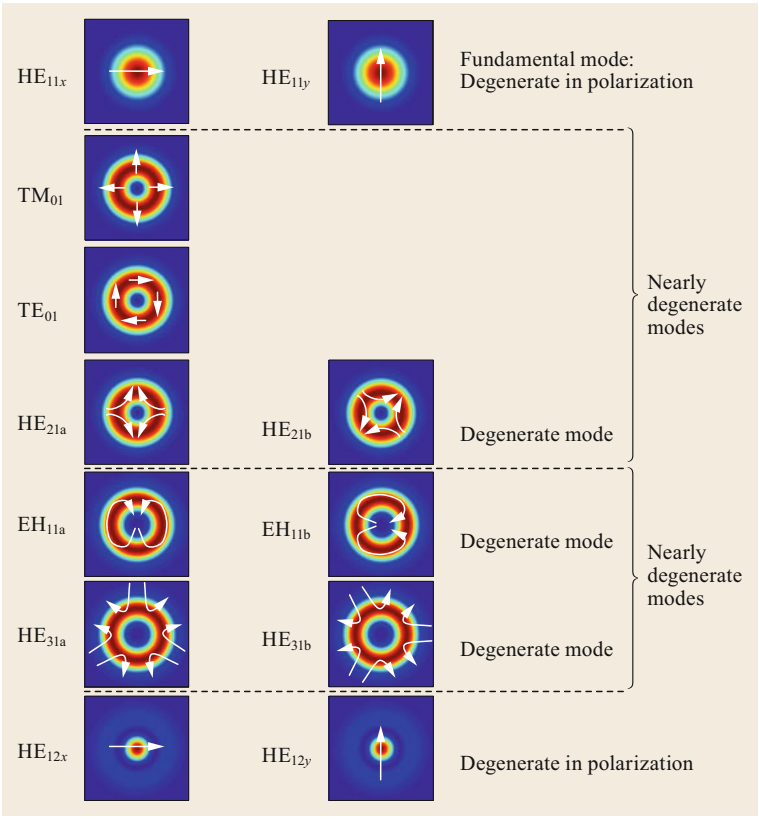


Fig. 10.5 First four mode groups of a step-index multimode fiber

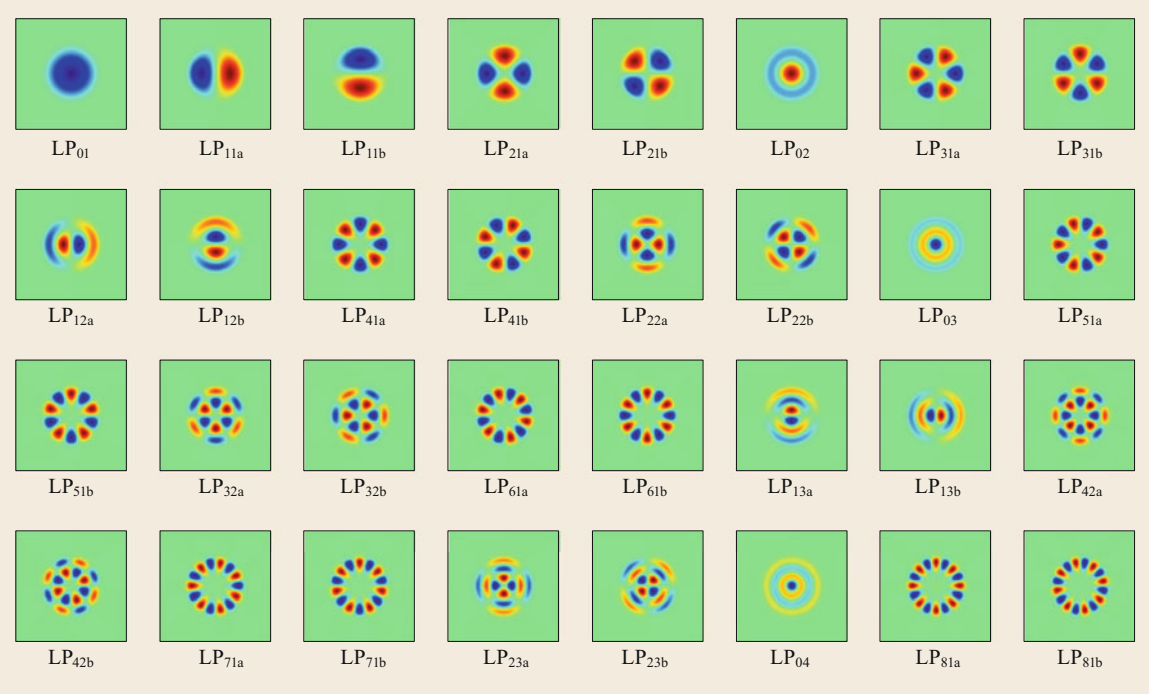


Fig. 10.6 First 32 modes of a step-index multimode fiber

Table 10.2 Cut-off frequencies of the modes of a step-index fiber

Mode	LP_{11}	LP_{21}	LP_{02}	LP_{31}	LP_{12}	LP_{41}	LP_{03}	LP_{32}	LP_{42}
V	2.4	3.8	3.9	5.1	5.5	6.4	7.1	8.4	9.8

In the relevant regime of weak guidance [10.61], where the index contrast is smaller or much smaller than 1%, the modes of a step-index fiber can be grouped in nearly degenerate mode groups (TM_{01}, TE_{01}, and the two-fold degenerate HE_{21} mode form a group). Modes belonging to the same group couple almost immediately due to fiber imperfections, and therefore do not propagate as individual modes in real fibers. Moreover, in the regime of weak guidance all modes are approximately linearly polarized (LP) [10.61]. Each LP mode is identified by two integers n and m and is denoted as LP_{nm}, where n characterizes the azimuthal dependence $\exp(\pm \mathrm{j}n\varphi)$ and sets the number of 2π phase changes for a rotation around the fiber axis, whereas m characterizes the radial dependence of the mode amplitude and sets the number of zero-crossings increased by one in the radial direction. Note that LP modes can equivalently be expressed either in terms of the functions $\exp(\pm \mathrm{j}n\varphi)$, or in terms of the functions $\cos(n\varphi)$ and $\sin(n\varphi)$. Figure 10.6 shows the first 32 modes of a weakly guiding step-index fiber in the $\cos(n\varphi)/\sin(n\varphi)$ representation. The main advantage of this representation is that the mode lateral profile functions are real-valued and can be plotted using a simple color map. In the figure red and blue indicate positive and amplitude values, respectively.

Modes with $n \neq 0$ are four-fold degenerate, as they are degenerate with respect to polarization and with respect to their azimuthal characteristics. Modes with $n = 0$ are two-fold degenerate, as they are only degenerate with respect to polarization. The number of LP modes that are guided by a step-index fiber depends on the normalized frequency V, which is defined as

$$V = \frac{2\pi a}{\lambda_0}\sqrt{n_\mathrm{c}^2 - n_0^2}\,. \tag{10.13}$$

Fibers with the same V number guide the same number of modes and the mode profiles are characterized by the same dependence on the normalized radial coordinate r/a. Step-index fibers are single-mode if $V < 2.4$. The cut-off frequencies of high-order modes are enumerated in Table 10.2.

According to (10.13), the number of modes of a step-index fiber can be increased either by increasing the core radius or by increasing the difference in refractive index $n_\mathrm{c} - n_0$.

10.3.2 Modes of a Graded-Index Fiber

Graded-index fibers have a nearly parabolic refractive index profile that is truncated at the core/cladding boundary. In order to reduce coupling between the last guided mode-group and the cladding modes, an additional low refractive index trench around the nearly parabolic core is added [10.5, 6]. The modes of a graded-index fiber with an ideally parabolic profile are Laguerre–Gaussian modes, where the radial term of the modal profile is a combination of a Gaussian function and a Laguerre polynomial, whereas the azimuthal term is still of the form $\exp(jn\varphi)$. As shown in Fig. 10.7, the modes of a graded-index fiber are divided into groups: all the modes of a given group are degenerate and each added group shows one additional

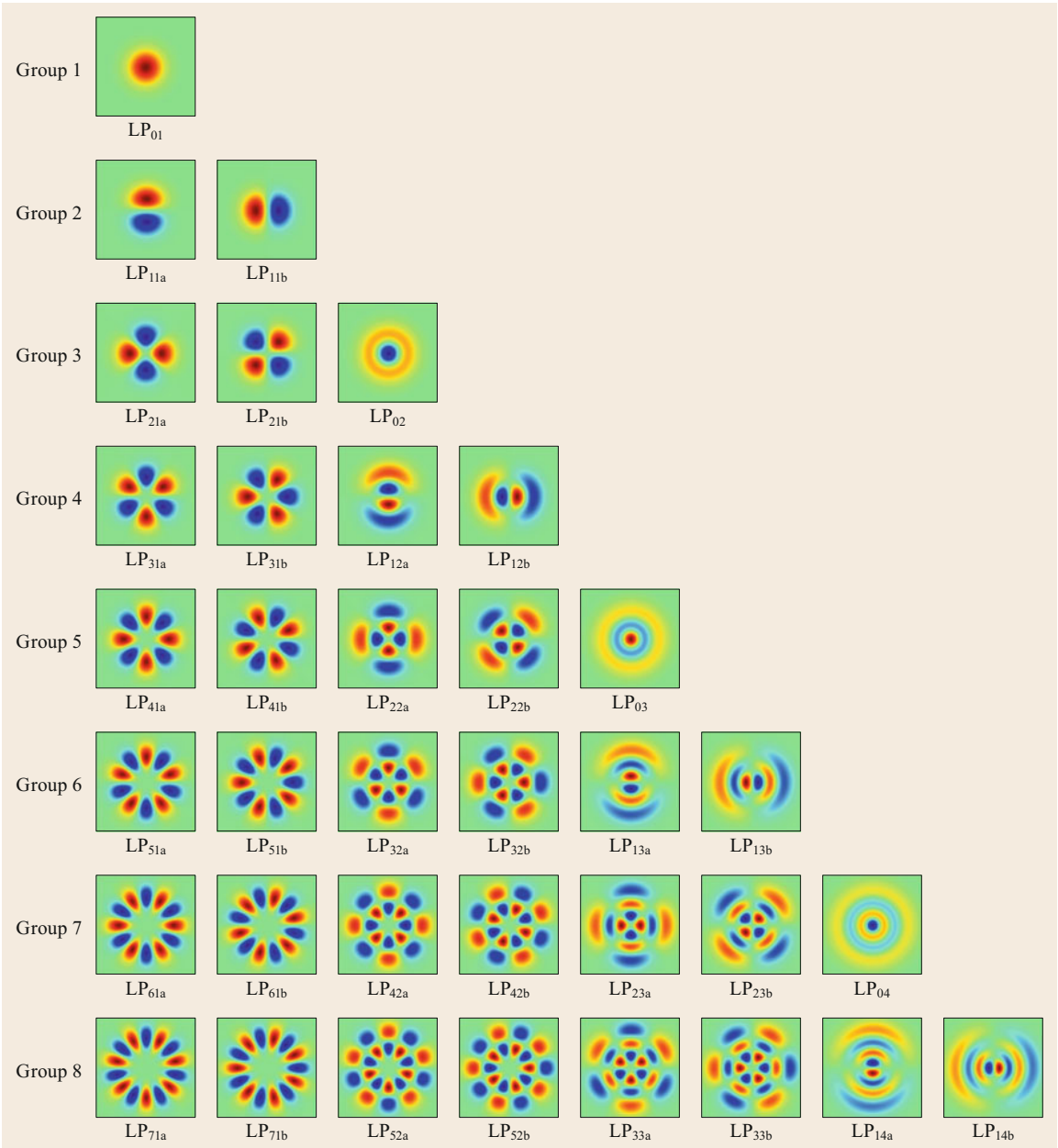


Fig. 10.7 The first 36 modes of a graded-index multimode fiber

degenerate mode. Although the modes of a graded-index fiber look similar to those of a step-index fiber, they have a different sequential order when sorted by the magnitude of their propagation constants. Also, the radial field extension, that in step-index fibers is mostly confined within the core for all modes, in graded-index fibers is proportional to the mode order, so that higher order modes have a significantly larger modal diameter. The propagation constants of the mode groups of an ideal graded-index fiber are equally spaced, and all modes have nearly the same group velocity, which makes the graded-index fiber the multimode fiber with the smallest theoretical DGD between pulses propagating in different modes. In practice, the DGD of a graded-index fiber is determined by the accuracy of the refractive index profile and by the dispersion properties of the core and cladding materials. Additionally, the modes of a graded-index fiber have the property of being invariant under spatial Fourier transformation. The spatial Fourier transform configuration can be achieved, for example, by placing the end face of the fiber in the back focal plain of a lens, so that the Fourier transform appears in the front focal plane of the same lens. For a graded-index fiber, this configuration reproduces a scaled version of the modes, which is not generally the case for other sets of guided modes like for example the modes of a step-index fiber.

10.3.3 Modes of Multicore Fibers: The Concept of Supermodes

Multicore fibers consist of multiple cores that are placed close to each other in a common cladding. The modes of a multicore fiber can be calculated by analyzing the modes of the ensemble of cores. The resulting modal field distributions spread across all cores and are referred to as supermodes. An example of supermodes for a three-core fiber is shown in Fig. 10.8.

When the coupling between cores is weak, the modes of a multicore fiber can be calculated using coupled-mode theory [10.62], where the supermodes are approximated by linear combinations of the modes of the individual cores (core modes), and the coupling coefficients depend on the overlap between the core modes [10.62, 63]. This approximation is useful to qualitatively study the systematic coupling between the cores. Note that both modal representations, supermodes and core modes, can be used to understand the coupling behavior of multicore fibers, however the supermodes representation has the advantage that it provides exact solutions for any core configuration including the case of strong core-mode overlap.

In practice, the modal properties of the multicore fibers depend strongly on the distance between the cores, and three distinct regimes can be identified:

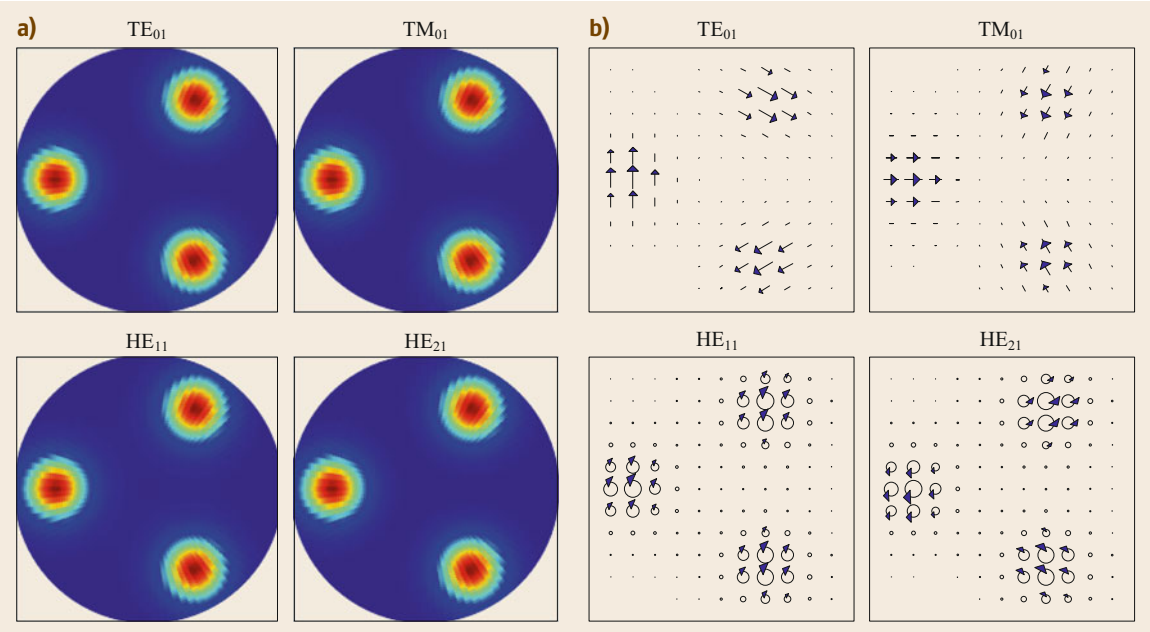


Fig. 10.8a,b Modes of a three-core multicore fiber supporting six vector modes. **(a)** The intensity profile, **(b)** direction of the electrical field indicated by *arrows*. *Arrows* arranged in circle indicates circular polarization and the position of the *arrow* relative to the circle indicates the overall relative phase

The first is the *weak coupling regime*, where the refractive index variations between the cores of the fiber structure produce an equivalent propagation-constant difference between the cores, that is bigger than the difference between the propagation constants of the supermodes. In this case, the cores behave as independent waveguides, showing a random-like coupling between neighbor cores, and the coupling between the cores is then described by the coupled power theory [10.64, 65]. The weak coupling regime is typically observed in multicore fibers where the core spacing is significantly larger than the core-mode field diameter.

The second regime is the *supermode regime*, where the difference in propagation constant between the supermodes is much bigger than the propagation constant variations introduced by the refractive index variations between the cores or external fiber perturbations like twists and bends. In this case, the supermodes are stable and can propagate unperturbed, and the fiber behaves similarly to a conventional multimode fiber, which shows a weak random-like coupling between the supermodes. The supermode regime is typically observed in multicore fibers where the core spacing is comparable or smaller than the core-mode field diameter.

The third regime can be observed when the difference in the propagation constant between the supermodes is comparable to the propagation constant variation induced by refractive index variation and the external fiber perturbations. In this case, the supermodes couple strongly and mix continuously. We refer to this regime as the *strong coupling regime*, and to the fibers operated in this regime as coupled-core multicore fibers. These fibers offer the advantage of showing very narrow impulse responses for all coupled cores, typically one order of magnitude (or more) narrower than impulse responses achievable in graded-index multimode fibers with an equivalent number of modes (Sect. 10.5.1). The strong coupling regime is typically observed in multicore fibers where the core spacing is comparable or slightly larger than the core-mode field diameter.

10.4 Representation of Modes in Fibers

We start by considering the case of a single-mode fiber, which for simplicity we assume to be a step-index fiber, where the fundamental mode is HE_{11}. We express the three-dimensional real-valued electric field vector $\boldsymbol{E}(x, y, z, t)$ introduced in the previous section as

$$\boldsymbol{E}(x,y,z,t) = \mathrm{Re}\left\{\left[\frac{\boldsymbol{F}_{\mathrm{HE}_{11x}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{HE}_{11}}(\omega_0)}E_x(z,t) + \frac{\boldsymbol{F}_{\mathrm{HE}_{11y}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{HE}_{11}}(\omega_0)}E_y(z,t)\right]\mathrm{e}^{-\mathrm{j}\omega_0 t}\right\}, \tag{10.14}$$

where $\boldsymbol{F}_{\mathrm{HE}_{11}x}$ and $\boldsymbol{F}_{\mathrm{HE}_{11}y}$ are the lateral profile functions of the fundamental mode aligned with the x- and y-directions, respectively. Note that the specific choice of the x- and y- directions is immaterial to this discussion, as any other pair of orthogonal directions would be equally suitable. The terms E_x and E_y are the corresponding complex envelopes of the field and the normalization coefficient $\mathcal{N}_{\mathrm{HE}_{11}}$ is introduced to ensure that the power in Watts that is carried by the x-oriented (y-oriented) mode is $|E_x|^2$ ($|E_y|^2$). We note that the mode lateral profile functions are evaluated at ω_0 owing to the fact that in all cases of practical relevance the bandwidth of the individual complex envelopes is sufficiently small to ignore the dependence of $\boldsymbol{F}_n$ on frequency. The form of (10.14) shows that the complex envelopes E_x and E_y are not exactly the x and y polarization components of the electric field, as follows from the fact that the HE_{11} lateral profile function possesses a nonzero component in the z-direction. The situation simplifies in the relevant case of weakly guiding fibers [10.61], where the fundamental mode LP_{01} is linearly polarized and its longitudinal component is negligible. In this case, E_x and E_y characterize fully and independently the x and y polarization components of the electric field (which is indeed a two-dimensional vector in the x–y plane) and the field vector can be expressed as

$$\begin{aligned}\boldsymbol{E}(x,y,z,t) &= \mathrm{Re}\left\{\left[\frac{\boldsymbol{F}_{\mathrm{LP}_{01x}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{LP}_{01}}(\omega_0)}E_x(z,t) + \frac{\boldsymbol{F}_{\mathrm{LP}_{01y}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{LP}_{01}}(\omega_0)}E_y(z,t)\right]\mathrm{e}^{-\mathrm{j}\omega_0 t}\right\}, \\ &= \frac{F_{\mathrm{LP}_{01}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{LP}_{01}}(\omega_0)}\mathrm{Re}\left\{[E_x(z,t)\hat{\boldsymbol{x}} + E_y(z,t)\hat{\boldsymbol{z}}]\mathrm{e}^{-\mathrm{j}\omega_0 t}\right\},\end{aligned} \tag{10.15}$$

where the second equality relies on the fact that the lateral profile of the fundamental mode is real-valued, and $\boldsymbol{F}_{\mathrm{LP}_{01x}} = F_{\mathrm{LP}_{01}}\hat{\boldsymbol{x}}$, $\boldsymbol{F}_{\mathrm{LP}_{01y}} = F_{\mathrm{LP}_{01}}\hat{\boldsymbol{y}}$, where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are unit vectors pointing in the x- and y-directions, respectively. It is convenient to introduce the bi-dimensional vector $\boldsymbol{E}(z,t)$ defined as

$$\boldsymbol{E}(z,t) = \begin{pmatrix} E_x(z,t) \\ E_y(z,t) \end{pmatrix}. \tag{10.16}$$

This vector is related to the Jones vector introduced in Sect. 10.4.1. It provides a full characterization of the optical field not only in the weakly guiding approximation, where its components are in a one-to-one correspondence with the x and y polarization components of the field, but also in the most general case where the longitudinal component is non-negligible. Note that the same vector notation is used, yet with no risk of confusion, to denote both the three-dimensional vectors $\boldsymbol{E}(x,y,z,t)$ and $\boldsymbol{F}(x,y,\omega_0)$, and the two-dimensional vector $\boldsymbol{E}(z,t)$. The extension of this representation to the case of higher order modes is straightforward, but requires some clarifications. Let us consider the group of the nearly degenerate modes $\mathrm{HE}_{21\mathrm{a}}$ and $\mathrm{HE}_{21\mathrm{b}}$, and the transverse modes TE_{01} and TM_{01}, whose excitation yields the field

$$\begin{aligned}\boldsymbol{E}(x,y,z,t) = \mathrm{Re}\Bigg\{\Bigg[&\frac{\boldsymbol{F}_{\mathrm{HE}_{21\mathrm{a}}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{HE}_{21}}(\omega_0)}E_{\mathrm{HE}_{21\mathrm{a}}}(z,t)\\ &+\frac{\boldsymbol{F}_{\mathrm{HE}_{21\mathrm{b}}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{HE}_{21}}(\omega_0)}E_{\mathrm{HE}_{21\mathrm{b}}}(z,t)\\ &+\frac{\boldsymbol{F}_{\mathrm{TE}_{01}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{TE}_{01}}(\omega_0)}E_{\mathrm{TE}_{01}}(z,t)\\ &+\frac{\boldsymbol{F}_{\mathrm{TM}_{01}}(x,y,\omega_0)}{\mathcal{N}_{\mathrm{TM}_{01}}(\omega_0)}E_{\mathrm{TM}_{01}}(z,t)\Bigg]\mathrm{e}^{-\mathrm{j}\omega_0 t}\Bigg\}.\end{aligned} \tag{10.17}$$

In the weakly guiding approximation, the lateral profile functions of the HE, TE, and TM modes can be used to express those of the LP_{11} group as follows

$$\frac{\boldsymbol{F}_{\mathrm{LP}_{11\mathrm{ax}}}}{\mathcal{N}_{\mathrm{LP}_{11}}} = \frac{1}{\sqrt{2}}\left(\frac{\boldsymbol{F}_{\mathrm{HE}_{21\mathrm{a}}}}{\mathcal{N}_{\mathrm{HE}_{21}}} + \frac{\boldsymbol{F}_{\mathrm{TM}_{01}}}{\mathcal{N}_{\mathrm{TM}_{01}}}\right), \tag{10.18}$$

$$\frac{\boldsymbol{F}_{\mathrm{LP}_{11\mathrm{ay}}}}{\mathcal{N}_{\mathrm{LP}_{11}}} = \frac{1}{\sqrt{2}}\left(\frac{\boldsymbol{F}_{\mathrm{HE}_{21\mathrm{b}}}}{\mathcal{N}_{\mathrm{HE}_{21}}} - \frac{\boldsymbol{F}_{\mathrm{TE}_{01}}}{\mathcal{N}_{\mathrm{TE}_{01}}}\right), \tag{10.19}$$

$$\frac{\boldsymbol{F}_{\mathrm{LP}_{11\mathrm{bx}}}}{\mathcal{N}_{\mathrm{LP}_{11}}} = \frac{1}{\sqrt{2}}\left(\frac{\boldsymbol{F}_{\mathrm{HE}_{21\mathrm{b}}}}{\mathcal{N}_{\mathrm{HE}_{21}}} + \frac{\boldsymbol{F}_{\mathrm{TE}_{01}}}{\mathcal{N}_{\mathrm{TE}_{01}}}\right), \tag{10.20}$$

$$\frac{\boldsymbol{F}_{\mathrm{LP}_{11\mathrm{by}}}}{\mathcal{N}_{\mathrm{LP}_{11}}} = \frac{1}{\sqrt{2}}\left(\frac{\boldsymbol{F}_{\mathrm{HE}_{21\mathrm{a}}}}{\mathcal{N}_{\mathrm{HE}_{21}}} - \frac{\boldsymbol{F}_{\mathrm{TM}_{01}}}{\mathcal{N}_{\mathrm{TM}_{01}}}\right), \tag{10.21}$$

(where we dropped the dependence on x, y, and ω_0 for ease of notation) with the result

$$\begin{aligned}\boldsymbol{E}(x,y,z,t) = \mathrm{Re}\Bigg\{\Bigg[&\frac{\boldsymbol{F}_{\mathrm{LP}_{11\mathrm{ax}}}}{\mathcal{N}_{\mathrm{LP}_{11}}}E_{\mathrm{LP}_{11\mathrm{ax}}}(z,t)\\ &+\frac{\boldsymbol{F}_{\mathrm{LP}_{11\mathrm{ay}}}}{\mathcal{N}_{\mathrm{LP}_{11}}}E_{\mathrm{LP}_{11\mathrm{ay}}}(z,t)\\ &+\frac{\boldsymbol{F}_{\mathrm{LP}_{11\mathrm{bx}}}}{\mathcal{N}_{\mathrm{LP}_{11}}}E_{\mathrm{LP}_{11\mathrm{bx}}}(z,t)\\ &+\frac{\boldsymbol{F}_{\mathrm{LP}_{11\mathrm{ay}}}}{\mathcal{N}_{\mathrm{LP}_{11}}}E_{\mathrm{LP}_{11\mathrm{by}}}(z,t)\Bigg]\mathrm{e}^{-\mathrm{j}\omega_0 t}\Bigg\}\end{aligned}$$

$$\begin{aligned}= \mathrm{Re}\Bigg(\Bigg\{&\frac{F_{\mathrm{LP}_{11\mathrm{a}}}}{\mathcal{N}_{\mathrm{LP}_{11}}}[E_{\mathrm{LP}_{11\mathrm{ax}}}(z,t)\hat{\boldsymbol{x}}\\ &+ E_{\mathrm{LP}_{11\mathrm{ay}}}(z,t)\hat{\boldsymbol{y}}]\\ &+ \frac{F_{\mathrm{LP}_{11\mathrm{b}}}}{\mathcal{N}_{\mathrm{LP}_{11}}}[E_{\mathrm{LP}_{11\mathrm{bx}}}(z,t)\hat{\boldsymbol{x}}\\ &+ E_{\mathrm{LP}_{11\mathrm{by}}}(z,t)\hat{\boldsymbol{y}}]\Bigg\}\mathrm{e}^{-\mathrm{j}\omega_0 t}\Bigg),\end{aligned} \tag{10.22}$$

where the complex envelopes of the linearly polarized modes are obtained from those of the true fiber modes through

$$\begin{bmatrix}E_{\mathrm{LP}_{11\mathrm{ax}}}(z,t)\\ E_{\mathrm{LP}_{11\mathrm{ay}}}(z,t)\\ E_{\mathrm{LP}_{11\mathrm{bx}}}(z,t)\\ E_{\mathrm{LP}_{11\mathrm{by}}}(z,t)\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1&0&1&0\\ 0&1&0&-1\\ 0&1&0&1\\ 1&0&-1&0\end{bmatrix} \times \begin{bmatrix}E_{\mathrm{HE}_{21\mathrm{a}}}(z,t)\\ E_{\mathrm{HE}_{21\mathrm{b}}}(z,t)\\ E_{\mathrm{TM}_{01}}(z,t)\\ E_{\mathrm{TE}_{01}}(z,t)\end{bmatrix} \tag{10.23}$$

as follows from using (10.18)–(10.21) into (10.22). The second equality in (10.22) shows that each of the complex envelopes in the LP representation provides a complete characterization of a space and polarization mode. All together, the four complex envelopes give a complete description of the field, and they are related to the generalized Jones representation presented in Sect. 10.4.3.

We can now move to the most general case of $2N$ modes, where the factor of two accounts either for the degeneracy of a spatial mode (as is the case for HE_{11} or HE_{21}), or for the fact that in all cases of practical relevance a fiber cannot guide only one out of two quasi-degenerate modes (as is the case for TE_{01} and TM_{01}). In the weakly guiding approximation, the factor of two accounts simply for polarization degeneracy. By suitably sorting the guided modes, we can express the electric field as

$$\boldsymbol{E}(x,y,z,t) = \mathrm{Re}\left[\sum_{n=1}^{2N}\frac{\boldsymbol{F}_n(x,y,\omega_0)}{\mathcal{N}_n(\omega_0)}E_n(z,t)\mathrm{e}^{-\mathrm{j}\omega_0 t}\right], \tag{10.24}$$

where the term $E_n(z,t)$ is the complex envelope of the field in the n-th mode and the vector $\boldsymbol{F}_n(x,y,\omega_0)$ is the corresponding mode lateral profile. As specified already in the examples illustrated above, the normalization coefficients $\mathcal{N}_n(\omega_0)$ are introduced to ensure that the power in Watts that is carried by the n-th mode

is given by $|E_n(z,t)|^2$ [10.66]. Note that a different normalization could be assumed (for instance, in [10.67] the power in Watts is given by $|E_n(z,t)|^2/(2Z_0)$, where $Z_0 = \sqrt{\mu_0/\epsilon_0}$ is the impedance of vacuum). Their expression follows from the mode orthogonality condition (10.5), which we recall here for convenience,

$$\int \mathrm{d}x\mathrm{d}y \left(\boldsymbol{F}_n \times \boldsymbol{G}_m^*\right) \cdot \hat{\boldsymbol{z}} = 2\mathcal{N}_n^2 \delta_{n,m} \,. \tag{10.25}$$

The orthogonality condition implies the following equation, which is often found in the literature [10.66–68],

$$\int \mathrm{d}x\mathrm{d}y \left(\boldsymbol{F}_n \times \boldsymbol{G}_m^* + \boldsymbol{F}_m^* \times \boldsymbol{G}_n\right) \cdot \hat{\boldsymbol{z}} = 4\mathcal{N}_n^2 \delta_{n,m} \,, \tag{10.26}$$

and in the weakly guiding approximation simplifies to

$$\frac{n_{\mathrm{eff}}}{2Z_0} \mathrm{Re}\left(\int \mathrm{d}x\mathrm{d}y\, \boldsymbol{F}_n \cdot \boldsymbol{F}_m^*\right) = \delta_{n,m}\mathcal{N}_n^2 \,, \tag{10.27}$$

where n_{eff} is the effective refractive index of the fundamental mode. With the field expression introduced in (10.24), in the ideal case of a perfectly circular fiber, in the absence of loss, mode coupling and nonlinear propagation effects, the complex envelopes evolve according to the simple evolution equation in the frequency domain

$$\frac{\partial \tilde{E}_n(z,\omega)}{\partial z} = \mathrm{j}\beta_n(\omega)\tilde{E}_n(z,\omega) \,, \tag{10.28}$$

where by the tilde we denote a (frequency) Fourier transform according to the definition

$$\tilde{E}_n(z,\omega) = \int E_n(z,t)\exp(\mathrm{j}\omega t)\mathrm{d}t \,. \tag{10.29}$$

An important feature of (10.24) is that the effect of perturbations is captured through the dependence of the complex envelopes on the longitudinal coordinate, while the modes used in the expansion are those of the unperturbed fiber. An alternative approach, that is not further discussed in this chapter, relies on using perturbed local modes [10.69].

It is worth pointing out that the true fiber modes form a complete orthogonal basis for representing locally the lateral profile of the field propagating in the fiber, and therefore any other orthogonal basis obtained from a unitary transformation of the true fiber modes lateral profile functions works as well [10.70]. However, since the resulting lateral profile functions are not in general fiber modes, the evolution of the corresponding complex envelopes is described by coupled equations even in the ideal case of an unperturbed fiber. Equations (10.18)–(10.21) and (10.23) can be interpreted as an example of this change of basis. Indeed, it is well known that LP modes are only true modes within the weakly guiding approximation, whereas in reality they couple during propagation, not only in fibers with high-index-contrast, but also in weakly guiding fibers where the accumulated effects of the small modal birefringence cannot be ignored [10.71].

In the remainder of this section we review the Jones and Stokes formalisms [10.72–74], which are widely used for the study of polarization-related phenomena in single-mode fibers, and discuss their generalization to the multimode case.

10.4.1 Jones and Stokes Formalism for Single-Mode Fibers

Jones calculus was originally proposed to describe polarized light by means of two-dimensional vectors [10.72]. Indeed, as discussed in the previous section, the vector $\boldsymbol{E}(z,t)$ defined in (10.16) provides a complete description of the electric field propagating in a single-mode fiber, and the physical interpretation of its two components is slightly different whether the fiber is weakly guiding or not. The corresponding Jones vector is defined as the Fourier transform of $\boldsymbol{E}(z,t)$ normalized to have unit modulus, namely

$$|e(z,\omega)\rangle = \frac{\tilde{\boldsymbol{E}}(z,\omega)}{|\tilde{\boldsymbol{E}}(z,\omega)|} = \begin{pmatrix} e_x(z,\omega) \\ e_y(z,\omega) \end{pmatrix} , \tag{10.30}$$

where we use the bra–ket notation to denote a Jones vector $|e\rangle$. By the bra $\langle e(z,\omega)|$ we denote the Hermitian adjoint of the field Jones vector (i.e., the complex conjugate row vector), so that the unit-modulus condition can be expressed as $\langle e|e\rangle = |e_x|^2 + |e_y|^2 = 1$, and the scalar product between two Jones vectors is given by $\langle u|e\rangle = u_x^* e_x + u_x^* e_y$.

We now move to introducing the Stokes representation of the electric field. This is an alternative description based on the use of real-valued three-dimensional vectors and it is isomorphic to the Jones representation [10.71]. If we denote by $\boldsymbol{e}$ the Stokes vector corresponding to the Jones vector $|e\rangle$, its three components are defined as

$$\begin{aligned} e_1 &= |e_x|^2 - |e_y|^2 \,, \\ e_2 &= 2\mathrm{Re}\left(e_x^* e_y\right) \,, \\ e_3 &= 2\mathrm{Im}\left(e_x^* e_y\right) \,. \end{aligned} \tag{10.31}$$

The length of the Stokes vector can be evaluated to be $|\boldsymbol{e}| = \sqrt{e_1^2 + e_2^2 + e_3^2} = \langle e|e\rangle = 1$, where the second equality follows from the normalization of $|e\rangle$ (in some cases the Jones vector is not normalized to have unit modulus, with the result that the length of the Stokes vector equals the optical power). The ensemble of all possible polarization states spans the surface of a sphere of unit radius in Stokes space, which is famously known as the Poincaré sphere. An alternative expression of the Stokes vector, which turns out to be highly convenient for the generalization of the Stokes formalism to the multimode fiber case is the one based on the use of Pauli spin matrices, which we recall here for convenience

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 0 & -\mathrm{j} \\ \mathrm{j} & 0 \end{pmatrix}. \tag{10.32}$$

Note that the original definition of the Pauli matrices in quantum mechanics differs from the above by a circular permutation of the matrix subscripts. With the use of the Pauli matrices, (10.31) can be re-expressed as

$$e_n = \langle e|\sigma_n|e\rangle\,, \quad n \in \{1, 2, 3\}\,, \tag{10.33}$$

and by formally collecting the Pauli matrices into a column vector which we denote by $\boldsymbol{\sigma}$, the Stokes vector can be expressed in the following compact form

$$\boldsymbol{e} = \langle e|\boldsymbol{\sigma}|e\rangle\,. \tag{10.34}$$

Another relevant relation between the Jones and Stokes representations has to do with the projection operator. This is a 2×2 matrix returning the projection on a given Jones vector $|e\rangle$ of the Jones vector to which is applied,

$$|e\rangle\langle e| = \frac{1}{2}\,(\mathbf{I} + \boldsymbol{e}\cdot\boldsymbol{\sigma})\,. \tag{10.35}$$

Here, by $\mathbf{I}$ we denote the 2×2 identity matrix (in what follows we will use the same symbol to denote the $M\times M$ identity matrix as well, where M is an arbitrary integer), and where the scalar product between a Stokes vector and the Pauli matrix vector stands for the linear combination $\boldsymbol{e}\cdot\boldsymbol{\sigma} = e_1\sigma_1 + e_2\sigma_2 + e_3\sigma_3$. The equivalence between (10.33) and (10.35) follows from the equality

$$\langle e|\sigma_n|e\rangle = \mathrm{tr}\,(\sigma_n|e\rangle\langle e|)\,, \tag{10.36}$$

and from the trace-orthogonality of the Pauli matrices, that is

$$\mathrm{tr}\,(\sigma_n\sigma_m) = 2\delta_{n,m}\,, \tag{10.37}$$

where $\delta_{n,m}$ is the Kronecker delta. A useful consequence of (10.35) is

$$|\langle u|v\rangle|^2 = \frac{1}{2}\,(1 + \boldsymbol{u}\cdot\boldsymbol{v})\,, \tag{10.38}$$

which shows that orthogonal states of polarization, for which $\langle u|v\rangle = 0$, are antiparallel in Stokes space, namely $\boldsymbol{u}\cdot\boldsymbol{v} = -1$.

10.4.2 Polarization Coupling and Unitary Propagation in Single-Mode Fibers

Manufacturing imperfections and mechanical stress that are always present in real fibers are responsible for the fact that orthogonal polarization modes couple during propagation in single-mode fibers. In the absence of polarization-dependent loss (PDL), polarization-mode coupling can be conveniently described in Jones space by means of a unitary matrix $\mathbf{U}$ defined through the following input-output relation,

$$|\tilde{e}(z,\omega)\rangle = \mathrm{e}^{-\frac{\alpha}{2}z}\mathbf{U}(z,\omega)|\tilde{e}(0,\omega)\rangle\,, \tag{10.39}$$

where the term $\exp(-\alpha z/2)$ describes polarization-averaged loss. This term is immaterial to the present analysis and will be dropped in what follows. The unitary property $\mathbf{U}(z,\omega)\mathbf{U}^\dagger(z,\omega) = \mathbf{I}$ implies the following form for the evolution equation of $\mathbf{U}(z,\omega)$,

$$\frac{\mathrm{d}\mathbf{U}(z,\omega)}{\mathrm{d}z} = \mathrm{j}\mathbf{B}(z,\omega)\mathbf{U}(z,\omega)\,, \tag{10.40}$$

where $\mathbf{B}(z,\omega)$ is a Hermitian matrix. Indeed, by differentiating both sides of the equality $\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$, one obtains $(\mathrm{d}\mathbf{U}/\mathrm{d}z)\mathbf{U}^\dagger = -[(\mathrm{d}\mathbf{U}/\mathrm{d}z)\mathbf{U}^\dagger]^\dagger$, which implies that $(\mathrm{d}\mathbf{U}/\mathrm{d}z)\mathbf{U}^\dagger$ is anti-Hermitian, and hence can be expressed as $\mathrm{j}\mathbf{B}$, where $\mathbf{B}$ is Hermitian. Since the Pauli matrices form a basis for traceless Hermitian matrices, the above can be conveniently re-expressed as

$$\frac{\mathrm{d}\mathbf{U}(z,\omega)}{\mathrm{d}z} = \mathrm{j}\left[\beta_0(z,\omega)\mathbf{I} + \frac{1}{2}\boldsymbol{\beta}(z,\omega)\cdot\boldsymbol{\sigma}\right]\mathbf{U}(z,\omega)\,, \tag{10.41}$$

where by β_0 we denote the propagation constant of the fundamental mode, whose third-order Taylor expansion defined in (10.7) yields the terms describing polarization-averaged phase delay, group delay, and chromatic dispersion. The traceless matrix $\boldsymbol{\beta}(z,\omega)\cdot\boldsymbol{\sigma}$ accounts for the local (z-dependent) polarization-mode coupling that occurs during propagation (including its frequency dependence), where $\boldsymbol{\beta}$ is a three-dimensional real-valued vector, which we refer to as the *birefringence vector* [10.73, 75] (this definition relaxes the use

of the term birefringence vector, which normally is restricted to the frequency derivative of $\boldsymbol{\beta}$). The simple relation between $\boldsymbol{\beta}$ and $\mathbf{B}$ is $\beta_n = \mathrm{tr}(\sigma_n \mathbf{B})$.

A closed-form solution for (10.41) in the general case does not exist, however the evolution of $\mathbf{U}$ over a short fiber section of length Δz is given by the following expression

$$\mathbf{U}(z+\Delta z,\omega) \simeq \exp\left[\mathrm{j}\beta_0(z,\omega)\Delta z\right] \times \exp\left[\frac{\mathrm{j}}{2}\boldsymbol{\beta}(z,\omega)\cdot\boldsymbol{\sigma}\,\Delta z\right]\mathbf{U}(z,\omega)\,, \tag{10.42}$$

which is customarily employed in numerical simulations, where a fiber is modeled as a concatenation of multiple short sections (waveplates). If the birefringence vector is z-independent, namely $\boldsymbol{\beta}(z,\omega) = \boldsymbol{\beta}(\omega)$, the above expression can be readily modified to evaluate the fiber transfer matrix $\mathbf{U}$, with the result

$$\mathbf{U}(z,\omega) = \exp\left[\mathrm{j}\int_0^z \beta_0(\zeta,\omega)\mathrm{d}\zeta\right] \times \exp\left[\frac{\mathrm{j}}{2}\boldsymbol{\beta}(\omega)\cdot\boldsymbol{\sigma}\,z\right], \tag{10.43}$$

which ensures the initial condition $\mathbf{U}(0,\omega) = \mathbf{I}$. This case is of no practical relevance (in reality the fiber birefringence is rapidly varying along the propagation axis), however the form of (10.43) is interesting since also in the most general case $\mathbf{U}$ can be expressed in the same form,

$$\mathbf{U}(z,\omega) = \exp\left[\mathrm{j}\phi_0(z,\omega)\right]\exp\left[\frac{\mathrm{j}}{2}\boldsymbol{r}(z,\omega)\cdot\boldsymbol{\sigma}\right], \tag{10.44}$$

where ϕ_0 accounts for the accumulated phase and $\boldsymbol{r}$ for the accumulated effect of the fiber birefringence from the input to the generic position z. Indeed, a unitary matrix can in general be expressed as $\mathbf{U} = \exp(\mathrm{j}\mathbf{H}/2)$, where $\mathbf{U}$ is a traceless Hermitian matrix that in turn can be expanded in terms of the Pauli matrices as $\mathbf{H} = \boldsymbol{r}\cdot\boldsymbol{\sigma}$. A useful alternative expression for $\mathbf{U}$ follows from the eigenvector analysis of the matrix exponential appearing in (10.44), which yields

$$\mathbf{U} = \mathrm{e}^{\mathrm{j}\phi_0}\left(\mathrm{e}^{-\mathrm{j}r/2}|r\rangle\langle r| + \mathrm{e}^{\mathrm{j}r/2}|r_\perp\rangle\langle r_\perp|\right), \tag{10.45}$$

where $r = |\boldsymbol{r}|$, and by $|r\rangle$ and $|r_\perp\rangle$, we denote the two orthogonal Jones vectors corresponding to the (unit-length) Stokes vectors $\pm\boldsymbol{r}/r$. Equation (10.45) shows that $|r\rangle$ and $|r_\perp\rangle$ are the two eigenstates of $\mathbf{U}$ and their eigenvalues are equal to $\exp(-\mathrm{j}r/2)$ and $\exp(\mathrm{j}r/2)$, respectively. Equation (10.45) is consistent with a general property of unitary matrices of any dimension of having orthogonal eigenvectors with unit-modulus eigenvalues. A detailed derivation of (10.45) can be found in [10.73]. The derivation relies essentially on two properties of the Pauli matrices, $\sigma_n^2 = \mathbf{I}$ and $\sigma_n\sigma_m = -\sigma_m\sigma_n$, which yield $(\boldsymbol{r}\cdot\boldsymbol{\sigma})^{2k} = |\boldsymbol{r}|^{2k}\mathbf{I}$. Use of the latter equality in the power expansion of the matrix exponential yields $\exp(\mathrm{j}\boldsymbol{r}\cdot\boldsymbol{\sigma}/2) = \cos(|\boldsymbol{r}|)\mathbf{I} + \mathrm{j}\sin(|\boldsymbol{r}|)(\boldsymbol{r}\cdot\boldsymbol{\sigma})/|\boldsymbol{r}|$. Equation (10.45) is finally obtained by considering (10.35) and the subsequent discussion.

The evolution of the field polarization state can be conveniently described also in Stokes space, where the overall effect of unitary fiber propagation is rotation of the field Stokes vector, as follows from the invariance of the Stokes vector length (this invariance is not trivially the consequence of the normalization involved in the definition of the Jones vectors but rather the result of power conservation during unitary propagation). If we denote by $\mathbf{R}$ the 3×3 rotation matrix isomorphic to the unitary Jones matrix $\mathbf{U}$, the input-output relation $\boldsymbol{e}_{\mathrm{out}} = \mathbf{R}\boldsymbol{e}_{\mathrm{in}} = \langle e_{\mathrm{in}}|\mathbf{U}^\dagger\boldsymbol{\sigma}\mathbf{U}|e_{\mathrm{in}}\rangle$ yields the following simple relation [10.73]

$$\mathbf{R}\boldsymbol{\sigma} = \mathbf{U}^\dagger\boldsymbol{\sigma}\mathbf{U}\,, \tag{10.46}$$

which connects $\mathbf{U}$ and $\mathbf{R}$. The matrix $\mathbf{R}$ is also referred to as the Müller matrix. The known evolution equation for the field Stokes vector is obtained by differentiating the expression $\boldsymbol{e} = \mathrm{tr}\,(\boldsymbol{\sigma}|e\rangle\langle e|)$, which yields

$$\frac{\partial\boldsymbol{e}}{\partial z} = \mathrm{tr}\left[\boldsymbol{\sigma}\,\mathrm{j}\frac{(\boldsymbol{\beta}\cdot\boldsymbol{\sigma})(\boldsymbol{e}\cdot\boldsymbol{\sigma}) - (\boldsymbol{e}\cdot\boldsymbol{\sigma})(\boldsymbol{\beta}\cdot\boldsymbol{\sigma})}{2}\right] = \boldsymbol{\beta}\times\boldsymbol{e}\,, \tag{10.47}$$

where the first equality follows from using (10.41) and the second requires using some of the Pauli matrices algebra. Equation (10.47) provides an intuitive interpretation of the local birefringence vector $\boldsymbol{\beta}$. Indeed, it shows $\boldsymbol{\beta}$ to be the local rotation axis that characterizes the trajectory drawn by the tip of the field Stokes vector on the Poincaré sphere, as the field propagates along the fiber, as illustrated in Fig. 10.9. In the case of uniform birefringence the trajectory simplifies to a circle and the motion on this circle is described by the matrix $\mathbf{R}(z) = \exp(z\boldsymbol{\beta}\times)$, where by $\boldsymbol{\beta}\times$ we denote the matrix operator that, if applied to the vector $\boldsymbol{s}$, performs the vector product $\boldsymbol{\beta}\times\boldsymbol{s}$, namely

$$\boldsymbol{\beta}\times = \begin{pmatrix} 0 & -\beta_3 & \beta_2 \\ \beta_3 & 0 & -\beta_1 \\ -\beta_2 & \beta_1 & 0 \end{pmatrix}. \tag{10.48}$$

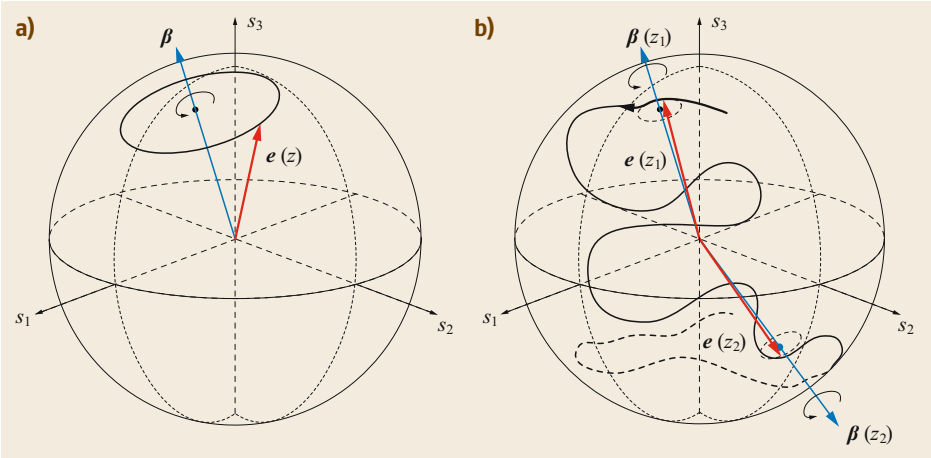


Fig. 10.9a,b Trajectory of the field Stokes vector on the Poincaré sphere. **(a)** If the birefringence vector $\boldsymbol{\beta}$ is constant along the fiber, the Stokes vector rotates around $\boldsymbol{\beta}$, namely the trajectory is a circle, the rotation axis is $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}/|\boldsymbol{\beta}|$, and the angular velocity is $|\boldsymbol{\beta}|$. **(b)** In the general case of varying birefringence, the Stokes vector trajectory can be approximated by means of infinitesimal rotations around the local birefringence vector $\boldsymbol{\beta}(z)$

The expression of $\mathbf{R}(z)$ shows that the rotation angle is $|\boldsymbol{\beta}|z$ and the rotation axis is $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}/|\boldsymbol{\beta}|$, thereby implying that the two orthogonal states whose Stokes vectors are parallel and antiparallel to $\boldsymbol{\beta}$ are propagation eigenstates. This argument is useful to clarify the isomorphic relation existing between the general unitary Jones matrix $\mathbf{U} = \exp(\mathrm{j}\phi_0)\exp(\mathrm{j}\boldsymbol{r}\cdot\boldsymbol{\sigma}/2)$ and the Müller matrix $\mathbf{R} = \exp(\boldsymbol{r}\times)$ (note that the phase shift ϕ_0 is immaterial in the Stokes representation, consistent with the definition of the Stokes vector itself).

10.4.3 Generalized Jones and Stokes Formalism

The example of the four-mode field discussed in Sect. 10.4 suggests that the complex envelopes $E_n(z,t)$ provide a complete description of the field, although their physical interpretation is slightly different whether the fiber is weakly guiding or not. The generalized Jones vector $|e(z,\omega)\rangle$, often referred to as the field *hyperpolarization* vector, is hence constructed by stacking the Fourier transform of the individual complex envelopes on top of each other, and by normalizing the resulting $2N$-dimensional column vector to have unit modulus, formally identically to the definition used in (10.30) for the single-mode case [10.76–78],

$$\boldsymbol{E}(z,t) = \begin{pmatrix} E_1(z,t) \\ E_2(z,t) \\ \vdots \\ E_{2N} \end{pmatrix}, \quad |e(z,\omega)\rangle = \frac{\tilde{\boldsymbol{E}}(z,\omega)}{|\tilde{\boldsymbol{E}}(z,\omega)|}\,. \tag{10.49}$$

The symbol $\boldsymbol{E}$, which was previously used to denote a two-dimensional column vector, here denotes a $2N$-dimensional column vector.

The generalization of the Stokes representation is less straightforward and entails a generalization of the Pauli matrix formalism. A convenient starting point is (10.35), which shows that the Stokes representation of a single-mode field is related to the expansion of the projection operator $|e\rangle\langle e|$ in terms of the Pauli matrices. Since $|e\rangle\langle e|$ is a $2N\times 2N$ Hermitian matrix for $N > 1$ as well as for $N = 1$, (10.35) can be generalized into

$$|e\rangle\langle e| = \frac{1}{2N}\left(\mathbf{I} + \boldsymbol{s}\cdot\boldsymbol{\Lambda}\right), \tag{10.50}$$

where $\boldsymbol{s}$ is the generalized Stokes vector and $\boldsymbol{\Lambda}$ is a vector collecting the generalized Pauli matrices Λ_n, which must be traceless Hermitian matrices fulfilling the following trace-orthogonality condition,

$$\mathrm{tr}\{\Lambda_m\Lambda_n\} = 2N\delta_{n,m}\,. \tag{10.51}$$

Matrices of this type form a basis for all $2N\times 2N$ traceless Hermitian matrices (a recursive algorithm to construct the matrices Λ_n for any number of modes is illustrated in the appendix of [10.78]). These have $D = 4N^2 - 1$ degrees of freedom, as follows from the fact that the elements on the main diagonal are real-valued and the off-diagonal elements are complex-conjugate in pairs. The subtraction of one accounts for the zero-trace constraint. These considerations imply that the generalized Stokes vectors are D-dimensional and real-valued, where D is hence the dimensionality of the generalized

Stokes space. Note, however, that the region spanned by the Stokes vectors is only $(4N-2)$-dimensional, like the hyperpolarization space (the $2N$ complex-valued entries of $|e\rangle$ minus the unit-magnitude constraint and the common phase of the hyperpolarization vector).

Equations (10.50) and (10.51) imply the following generalized properties

$$e_n = \langle e|\Lambda_n|e\rangle, \quad n \in \{1,\ldots,D\}\,, \tag{10.52}$$

$$\boldsymbol{e} = \langle e|\boldsymbol{\Lambda}|e\rangle\,, \tag{10.53}$$

$$|\langle u|v\rangle|^2 = \frac{1}{2N}\left(1+\boldsymbol{u}\cdot\boldsymbol{v}\right)\,. \tag{10.54}$$

Note that the length of the generalized Stokes vectors is given by $|\boldsymbol{e}| = \sqrt{2N-1}$, as follows from (10.54) for $u = v$. Another important consequence of the same equation is that Stokes vectors corresponding to orthogonal Jones vectors (for which $\langle u|v\rangle = 0$) are characterized by the relation $\boldsymbol{u}\cdot\boldsymbol{v} = -1$, which in the multidimensional case does not imply that $\boldsymbol{u}$ and $\boldsymbol{v}$ are antiparallel. In fact, since their magnitude is not 1, one can define the angle α formed by two Stokes vectors corresponding to orthogonal Jones vectors through the equality $\boldsymbol{u}\cdot\boldsymbol{v} = (2N-1)\cos(\alpha) = -1$. We note that this result does not change by normalizing the generalized Stokes vectors to have unit length, as is done in [10.79].

10.5 Mode Coupling and Unitary Propagation in SDM Fibers

In the absence of MDL, mode coupling in a fiber that supports $2N$ modes is described by a unitary $2N\times 2N$ matrix $\mathbf{U}(z,\omega)$, whose evolution obeys the equation,

$$\frac{\mathrm{d}\mathbf{U}(z,\omega)}{\mathrm{d}z} = \mathrm{j}\mathbf{B}(z,\omega)\mathbf{U}(z,\omega)\,, \tag{10.55}$$

which is identical to (10.40), provided that the symbol $\mathbf{B}$ denotes a $2N\times 2N$ Hermitian matrix. The individual terms of $\mathbf{B}$ account for the coupling between pairs of modes, whereas blocks of $\mathbf{B}$ describe the coupling within and between groups of degenerate modes. An illustration is presented in Fig. 10.10.

The matrix $\mathbf{B}$ can be expanded in terms of the generalized Pauli matrices, thereby rendering (10.50) into

$$\frac{\mathrm{d}\mathbf{U}(z,\omega)}{\mathrm{d}z} = \mathrm{j}\left[\beta_0(z,\omega)\mathbf{I} + \frac{1}{2N}\boldsymbol{\beta}(z,\omega)\cdot\boldsymbol{\Lambda}\right]\mathbf{U}(z,\omega)\,, \tag{10.56}$$

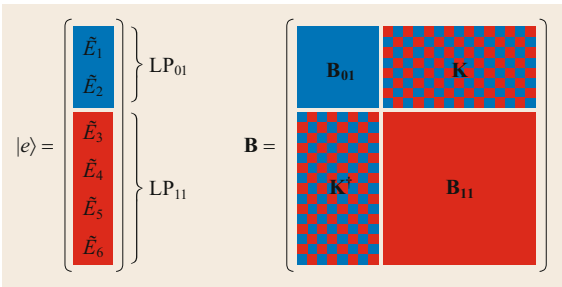


Fig. 10.10 The matrix $\mathbf{B}$ describing linear coupling in a fiber that supports propagation of LP_{01} and LP_{11} mode groups. The 2×2 block $\mathbf{B}_{01}$ accounts for polarization coupling within the fundamental mode, while the 4×4 block $\mathbf{B}_{11}$ accounts for mode coupling within the LP_{11} group. The 2×4 block $\mathbf{K}$ and its Hermitian adjoint $\mathbf{K}^\dagger$ describe intergroup mode coupling

where $\beta_0(z,\omega)$ has the meaning of the mode-averaged propagation constant, whereas the D-dimensional vector $\boldsymbol{\beta}(z,\omega)$ accounts for the mismatch between the various propagation constants, as well as for the local mode coupling [10.80, 81]. An alternative form of (10.55), which is often encountered in the literature, is obtained by accounting separately for the propagation constants of the individual modes,

$$\frac{\mathrm{d}\mathbf{U}(z,\omega)}{\mathrm{d}z} = \mathrm{j}\left[\mathbf{B}_0 + \frac{1}{2N}\boldsymbol{b}(z,\omega)\cdot\boldsymbol{\Lambda}\right]\mathbf{U}(z,\omega)\,, \tag{10.57}$$

where $\mathbf{B}_0$ denotes a diagonal matrix whose nonzero elements are the propagation constants of the individual modes, and where the vector $\boldsymbol{b}(z,\omega)$ only accounts for the mode coupling caused by the fiber perturbations. Clearly, this description is only appropriate in the case where the spatial modes used as a basis for the field lateral profile are also true fiber modes. In this case, in the absence of coupling ($\boldsymbol{b} = 0$) (10.57) yields $\mathbf{U} = \exp(\mathrm{j}\,\mathbf{B}_0 z)$. If the spatial modes assumed for the field lateral profile expansion are not true fiber modes (as is rigorously the case in the LP representation), then $\mathbf{B}_0$ is nondiagonal [10.71, 80, 82] and it accounts for the deterministic and periodic coupling that occurs between the spatial modes of the basis.

Similarly to the single-mode case, there is no closed-form solution for (10.57), except when the generalized birefringence vector is independent of z. In this situation (10.42)–(10.44) apply also to the case of multiple modes, provided that the quantity $\boldsymbol{\sigma}/2$ be replaced with $\boldsymbol{\Lambda}/2N$. A major difference between the single-mode and the multimode case stems from the fact that while the matrix $\boldsymbol{r}\cdot\boldsymbol{\sigma}$ admits two orthogonal eigenvectors, the matrix $\boldsymbol{r}\cdot\boldsymbol{\Lambda}$ admits $2N$ orthogonal eigenvectors,

and hence the matrix $\mathbf{U}$ can be expanded as

$$\mathbf{U} = \mathrm{e}^{\mathrm{j}\phi_0} \sum_{n=1}^{2N} \mathrm{e}^{\mathrm{j}\phi_n} |p_n\rangle\langle p_n| \tag{10.58}$$

where $|p_n\rangle$ and ϕ_n are the n-th eigenstate of $\boldsymbol{r}\cdot\boldsymbol{\Lambda}$ and the corresponding eigenvalue, respectively, and where $\sum_n \phi_n = 0$.

The Stokes-space representation of unitary evolution in the case of multiple-mode propagation is formally identical to the one discussed for single-mode fibers, and the main differences have to do with the increased dimensionality of the generalized Stokes space. Indeed, a unitary $4N\times 4N$ Jones matrix $\mathbf{U}$ corresponds to a norm-preserving transformation $\mathbf{R}$ in the $D\times D$ Stokes space, which can be still interpreted as a rotation, yet on a hypersphere, thereby failing to provide an intuitive description of the Stokes vector evolution. The relation connecting $\mathbf{U}$ and $\mathbf{R}$ is obtained from (10.46), by replacing the Pauli matrices with their generalized version, while the evolution equation for the generalized Stokes vector becomes

$$\begin{aligned}\frac{\partial \boldsymbol{e}}{\partial z} &= \mathrm{tr}\left[\boldsymbol{\Lambda}\,\mathrm{j}\frac{(\boldsymbol{\beta}\cdot\boldsymbol{\Lambda})(\boldsymbol{e}\cdot\boldsymbol{\Lambda}) - (\boldsymbol{e}\cdot\boldsymbol{\Lambda})(\boldsymbol{\beta}\cdot\boldsymbol{\Lambda})}{2N}\right] \\ &= \boldsymbol{\beta}\times\boldsymbol{e}\,,\end{aligned} \tag{10.59}$$

where the first equality follows from (10.56) and the second relies on the generalization of the vector product to the multidimensional case [10.78]. The k-th component of the generalized vector product between vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as

$$(\boldsymbol{A}\times\boldsymbol{B})_k = \sum_{i,j} f_{i,j,k} A_i B_j\,, \tag{10.60}$$

where by $f_{i,j,k}$ we denote the structure constants

$$f_{i,j,k} = \frac{\mathrm{j}}{(2N)^2}\mathrm{tr}\left[\Lambda_k(\Lambda_i\Lambda_j - \Lambda_j\Lambda_i)\right]. \tag{10.61}$$

Equation (10.59) is formally identical to the dynamic equation (10.47) obtained for single-mode propagation, and in principle it can be used for numerical simulations, just like in the single-mode case. Also, similarly to the case of a single-mode fiber, the Jones matrix $\mathbf{U} = \exp(\mathrm{j}\boldsymbol{r}\cdot\boldsymbol{\Lambda}/2N)$ is isomorphic to $\mathbf{R} = \exp(\boldsymbol{r}\times)$, where by $\boldsymbol{r}\times$ we denote the $D\times D$ matrix operator that returns the vector product $\boldsymbol{r}\times\boldsymbol{s}$, when applied to the vector $\boldsymbol{s}$. The expression of $\boldsymbol{r}\times$ follows from (10.60). It is interesting to note that the propagation matrix $\mathbf{R}$ cannot pull a legitimate Stokes vector out of the manifold of the legitimate Stokes vectors, and it can be shown that only Stokes vectors corresponding to the eigenstates of $\boldsymbol{r}\cdot\boldsymbol{\Lambda}$ are eigenstates of $\mathbf{R}$.

10.5.1 Modal Dispersion

The term modal dispersion is used to address two distinct phenomena. One is the modal dependence of the field group velocity, and the other is the frequency dependence of the random coupling process.

In the case of single-mode fibers, where the two polarizations of the fundamental mode are perfectly degenerate, modal dispersion is referred to as polarization-mode dispersion (PMD) and is a manifestation of the frequency dependence of the fiber random birefringence. In the case of multimode fiber structures, modal dispersion arises primarily from the group velocity mismatch existing between the fiber modes, but its properties are profoundly influenced by the regime of coupling that characterizes the multimode propagation. In all cases modal dispersion introduces a delayed channel response which needs to be equalized at the receiver end by means of MIMO techniques, thereby obviously increasing the complexity of the MIMO-DSP receiver. In what follows, we review the formalism developed for the study of PMD in single-mode fibers and discuss its generalization to the case of SDM fibers.

Polarization-Mode Dispersion in Single-Mode Fibers

The unitary condition $\mathbf{U}(z,\omega)\mathbf{U}^\dagger(z,\omega) = \mathbf{I}$ implies that the equation describing the frequency dependence of $\mathbf{U}$ is of the same form as (10.41) (which describes its z-dependence), namely

$$\frac{\partial \mathbf{U}(z,\omega)}{\partial\omega} = \mathrm{j}\left[\tau_0(z,\omega)\mathbf{I} + \frac{1}{2}\boldsymbol{\tau}(z,\omega)\cdot\boldsymbol{\sigma}\right]\mathbf{U}(z,\omega)\,. \tag{10.62}$$

The meaning of τ_0 and $\boldsymbol{\tau}$ is easily understood when they do not depend on frequency and hence (10.62) has the following simple solution

$$\mathbf{U}(z,\omega) = \exp(\mathrm{j}\tau_0\omega)\exp\left(\frac{\mathrm{j}}{2}\boldsymbol{\tau}\cdot\boldsymbol{\sigma}\,\omega\right)\mathbf{U}(z,0) \tag{10.63}$$

$$= \mathrm{e}^{\mathrm{j}\tau_0\omega}(\mathrm{e}^{\mathrm{j}\tau\omega/2}|\tau\rangle\langle\tau| + \mathrm{e}^{-\mathrm{j}\tau\omega/2}|\tau_\perp\rangle \times\langle\tau_\perp|)\mathbf{U}(z,0)\,, \tag{10.64}$$

where the second equality follows from the discussion related to (10.45). Here by $|\tau\rangle$ and $|\tau_\perp\rangle$, we denote the Jones vectors that correspond to the Stokes vectors $\pm\hat{\tau} = \pm\boldsymbol{\tau}/\tau$, with $\tau = |\boldsymbol{\tau}|$. This form indicates that a polarized input signal characterized by a state vector $|p\rangle$ such that $\mathbf{U}(z,0)|p\rangle = |\tau\rangle$ or by an orthogonal state $|p_\perp\rangle$ such that $\mathbf{U}(z,0)|p_\perp\rangle = |\tau_\perp\rangle$ is simply delayed by $\tau_0 + \tau/2$ or $\tau_0 - \tau/2$, respectively, at propagation dis-

tance z, namely

$$f(t)|p\rangle \rightarrow f\left(t-\tau_0-\frac{\tau}{2}\right)|\tau\rangle\,, \tag{10.65}$$

$$f(t)|p_\perp\rangle \rightarrow f\left(t-\tau_0+\frac{\tau}{2}\right)|\tau_\perp\rangle. \tag{10.66}$$

The polarization states $|p\rangle$ and $|p_\perp\rangle$ are known as *principal states of polarization* (PSP) and the relative delay τ that they accumulate during propagation is known as the *differential group delay* (DGD) (it is customary to refer to $|\tau\rangle$ and $|\tau_\perp\rangle$ as the slow and the fast PSPs, respectively, consistent with the fact that $|\tau\rangle$ is delayed with respect to $|\tau_\perp\rangle$). The vector $\boldsymbol{\tau}$, which as discussed provides a complete characterization of the PSPs, is famously known as the PMD vector.

The effect of PMD on arbitrarily polarized input states can be more conveniently described by introducing the distinction between *input* and *output* PSPs (this distinction is often erroneously ignored in the literature, however it becomes unnecessary if one assumes that no coupling occurs at $\omega = 0$ ($\mathbf{U}(z,0) = \mathbf{I}$), or equivalently if the Jones vectors are expressed in a rotating reference frame where this is the case). Indeed, $|p\rangle$ and $|p_\perp\rangle$ should be more correctly referred to as the *input* PSPs, whereas $|\tau\rangle$ and $|\tau_\perp\rangle$ should be referred to as the *output* PSPs. Using the simple relation existing between them, (10.64) can be re-expressed in the following form

$$\mathbf{U}(z,\omega) = \mathrm{e}^{\mathrm{j}\tau_0\omega}\left(\mathrm{e}^{\mathrm{j}\tau\omega/2}|\tau\rangle\langle p| + \mathrm{e}^{-\mathrm{j}\tau\omega/2}|\tau_\perp\rangle\langle p_\perp|\right), \tag{10.67}$$

which can be used to see that an input signal characterized by the state vector $|u\rangle$, during propagation splits into two replicas that are separated in time by the DGD,

$$f(t)|u\rangle \rightarrow \langle p|u\rangle f\left(t-\tau_0-\frac{\tau}{2}\right)|\tau\rangle + \langle p_\perp|u\rangle f\left(t-\tau_0+\frac{\tau}{2}\right)|\tau\rangle\,. \tag{10.68}$$

The two replicas are polarized along the output PSPs, whereas their amplitudes are equal to the projections of the input signal state vector onto the input PSPs. Equation (10.68) reduces to (10.65) or (10.66) if $|u\rangle = |p\rangle$ or $|u\rangle = |p_\perp\rangle$, respectively.

We recall that (10.67) and (10.68) were derived under the assumption that the PMD vector does not depend on frequency. The resulting description of PMD is hence an approximation usually referred to as a *first-order PMD picture*. Assessing the accuracy of this approximation requires studying the statistical properties of the PMD vector, which are briefly reviewed in what follows.

The PMD vector evolution equation is obtained in two steps. We first equate the two expressions for $\partial^2\mathbf{U}/\partial z\partial\omega$ obtained from (10.41) and (10.62), with the result

$$\frac{\partial\tau_0}{\partial z} = \frac{\partial\beta_0}{\partial\omega}\,, \tag{10.69}$$

$$\frac{\partial\boldsymbol{\tau}}{\partial z}\cdot\boldsymbol{\sigma} = \frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\sigma} + \mathrm{j}\frac{(\boldsymbol{\beta}\cdot\boldsymbol{\sigma})(\boldsymbol{\tau}\cdot\boldsymbol{\sigma}) - (\boldsymbol{\tau}\cdot\boldsymbol{\sigma})(\boldsymbol{\beta}\cdot\boldsymbol{\sigma})}{2}\,. \tag{10.70}$$

The first equation describes the accumulation of the polarization-averaged delay. The second can be further simplified by tracing out the Pauli matrices, with the same procedure illustrated in (10.47). The result is the famous PMD dynamic equation

$$\frac{\partial\boldsymbol{\tau}}{\partial z} = \frac{\partial\boldsymbol{\beta}}{\partial\omega} + \boldsymbol{\beta}\times\boldsymbol{\tau}\,. \tag{10.71}$$

The dependence of the birefringence vector on propagation distance renders the evolution of the PMD vector nontrivial. Most importantly, since $\boldsymbol{\beta}$ is random in nature (it describes random mode coupling), the PMD vector $\boldsymbol{\tau}$ is also random. The statistical properties of the birefringence vector of single-mode fibers have been accurately characterized in the past decade, and a well-established result is that its typical correlation length ranges from a few meters to a few hundreds of meters [10.75], implying that thousands of independent contributions accumulate over typical fiber lengths in metro and long-haul systems. This simple argument, in conjunction with the central-limit theorem, legitimates the description of the PMD vector evolution in terms of a three-dimensional Brownian motion [10.83]. That is, the three components of the PMD vector are independent and identically distributed Gaussian variables, and its length—the DGD—is characterized by a Maxwellian probability density function (plotted in Fig. 10.11a). The mean PMD vector length (or, equivalently, the mean DGD) is proportional to the square-root of the propagation distance

$$\langle\tau(z)\rangle = \kappa_{\mathrm{PMD}}\sqrt{z}\,, \tag{10.72}$$

where by angled brackets we denote ensemble averaging, and where the proportionality coefficient κ_{PMD} is the familiar PMD coefficient (note that the mean value of the DGD is frequency-independent, as it follows from the stationarity of the PMD process with respect to frequency). The PMD coefficient is customarily specified in units of $\mathrm{ps}/\sqrt{\mathrm{km}}$ and typical values range from $0.01\,\mathrm{ps}/\sqrt{\mathrm{km}}$ in modern low-PMD fibers to $0.5\,\mathrm{ps}/\sqrt{\mathrm{km}}$ in installed vintage systems [10.84]. We stress that the square-root growth of the mean DGD results from the random nature of the birefringence vector $\boldsymbol{\beta}$, while the details of the birefringence statistics

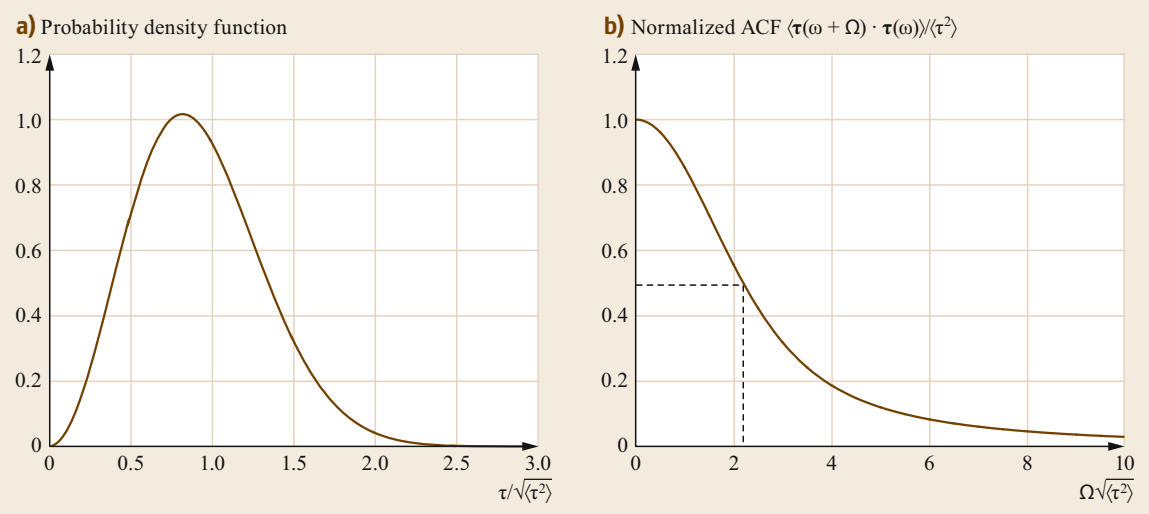


Fig. 10.11 **(a)** The probability density function of the DGD normalized to its root-mean-square value. **(b)** Normalized autocorrelation function of the PMD vector

are not relevant, as long as the fiber length exceeds by some orders of magnitude the birefringence correlation length. Just to mention one relevant example, it is worth pointing out that all the work carried out by Galtarossa's group [10.75, 82, 84, 85] relies on the assumption that circular birefringence is absent everywhere along the fiber, thereby implying that the third component of $\boldsymbol{\beta}$ vanishes. In this case, the second term at the right-hand side of (10.71) is the one responsible for lifting the PMD vector out of the equatorial plane in Stokes space, thereby making the assumption of vanishing circular birefringence immaterial, with the result that all of the described properties of the PMD vector are not affected by this detail of the model. Equation (10.72) can be expressed in the following equivalent form,

$$\langle \tau^2(z) \rangle = \kappa^2 z \,, \tag{10.73}$$

where $\kappa = \kappa_{\mathrm{PMD}} \sqrt{3\pi/8}$.

The random nature of PMD manifests itself also through the frequency dependence of the PMD vector, which is key to assessing the accuracy of the first-order approximation. This dependence is conveniently characterized by means of the two-frequency correlation function of the PMD vector [10.86–88], whose expression is

$$\langle \boldsymbol{\tau}(z,\omega) \cdot \boldsymbol{\tau}(z,\omega+\Omega) \rangle = 3 \frac{1 - \mathrm{e}^{-\frac{\Omega^2 \langle \tau^2(z) \rangle}{3}}}{\Omega^2} \tag{10.74}$$

The derivation of (10.74) is straightforward if one uses the tools of stochastic calculus [10.88]. The same result can also be obtained by approximating the fiber with a finite number N of constant-birefringence plates and then by taking the limit $N \to \infty$, which is the approach used in the work where (10.74) was first presented [10.86]. Note that the derivation of the autocorrelation function is performed by assuming a first-order expansion of the birefringence vector $\boldsymbol{\beta}(z, \omega + \Omega) \simeq \boldsymbol{\beta}(z,\omega) + \Omega(\partial\boldsymbol{\beta}/\partial\omega)(z,\omega)$. A similar assumption underpins the derivation of the generalized PMD vector autocorrelation function (ACF) in the multimode case. Inspecting the plot of (10.74) in Fig. 10.11 shows that the PMD vector ACF reduces to one half of its peak value at the angular frequency difference $\Omega_{3\mathrm{dB}} \simeq 2.18/\sqrt{\langle \tau^2 \rangle}$, which suggests that for smaller differences two PMD vectors are highly correlated with each other and hence the frequency dependence of the PMD vector is negligible. The corresponding frequency difference $B = \Omega_{3\mathrm{dB}}/2\pi \simeq 0.347/\sqrt{\langle \tau^2 \rangle}$ is often used as a definition of the PMD bandwidth, with the idea that the first-order PMD approximation only applies to the transmission of signals whose bandwidth does not exceed the PMD bandwidth. It is worth pointing out that in the case of single-mode fiber systems, this is almost always the case, for single-channel bandwidths of the order of a few tens of GHz. As an example consider a 1000 km link: for a legacy fiber with a PMD coefficient $\kappa_{\mathrm{PMD}} = 0.1\,\mathrm{ps}/\sqrt{\mathrm{km}}$, the PMD bandwidth is $B \simeq 100\,\mathrm{GHz}$, and it increases to $B \simeq 1\,\mathrm{THz}$ in the case of a low-PMD fiber with $\kappa_{\mathrm{PMD}} = 0.01\,\mathrm{ps}/\sqrt{\mathrm{km}}$. The situation is substantially different in the case of multimode fibers, as is discussed in the next section.

To conclude this section, we remind the reader that PMD is a unitary effect and hence, unlike PDL [10.89],

does not imply a fundamental system information capacity loss. For this reason, its effect can in principle be fully compensated for in the digital domain at the receiver of a polarization-multiplexed coherent system. The complexity of the necessary DSP scales with the magnitude of the system PMD (the differential delay that needs to be accommodated in time-domain equalization algorithms [10.90]), or equivalently with the PMD bandwidth (the resolution required in frequency-domain equalization algorithms [10.91]).

Generalization of the PMD Formalism

The derivation of (10.62) relies solely on the unitary nature of the Jones matrix $\mathbf{U}(z,\omega)$. Its extension to the case of multimode fiber structures is thereby straightforward, and the resulting equation can be expressed as

$$\frac{\partial \mathbf{U}(z,\omega)}{\partial \omega} = \mathrm{j}\left[\tau_0(z,\omega)\mathbf{I} + \frac{1}{2N}\boldsymbol{\tau}(z,\omega)\cdot\boldsymbol{\Lambda}\right]\mathbf{U}(z,\omega)\,, \tag{10.75}$$

where τ_0 is now the mode-averaged group delay, and $\boldsymbol{\tau}$ is a D-dimensional real-valued vector that generalizes the PMD vector and that is referred to as the *mode dispersion* (MD) vector [10.78]. Its evolution equation is also derived with the same procedure described in the single-mode case and the result is identical to (10.71), provided that the symbol $\times$ is used to denote the generalized vector product. A major difference with respect to the single-mode case is due to the phase and group velocity mismatch existing between the various fiber modes. As pointed out in the discussion of (10.56) and (10.57), this mismatch is captured by the generalized birefringence vector, which can be conveniently expressed as the sum of two contributions,

$$\boldsymbol{\beta}(z,\omega) = \boldsymbol{\beta}_\mathrm{d}(\omega) + \boldsymbol{\beta}_\mathrm{r}(z,\omega)\,, \tag{10.76}$$

where the term $\boldsymbol{\beta}_\mathrm{d}$ is the deterministic content of $\boldsymbol{\beta}$ accounting for the propagation constants mismatch (which is constant along the fiber, unless some specific special fiber design is considered), while the term $\boldsymbol{\beta}_\mathrm{r}$ models random coupling between modes. Moreover, if the spatial modes used for representing the field lateral profile are not true fiber modes, $\boldsymbol{\beta}_\mathrm{d}$ must also account for the deterministic coupling between them. With the formalism of (10.57), $\boldsymbol{\beta}_\mathrm{d}$ can be extracted using $\beta_{\mathrm{d},n} = \mathrm{tr}(\Lambda_n \mathbf{B}_0)/2N$, where $n = 1\ldots D$. As an example, Fig. 10.12 illustrates the case of a coupled-core three-core fiber where the spatial-modes basis consists of the fundamental modes of the individual cores (they are not true fiber modes—the true fiber modes are supermodes, as discussed in Sect. 10.3.3). In this case, one can compute [10.80] $\boldsymbol{\beta}_\mathrm{d} = 2b\sqrt{N}(\hat{e}_{10}+\hat{e}_{16}+\hat{e}_{18}+\hat{e}_{24}+\hat{e}_{26}+\hat{e}_{32})$, where $\hat{\boldsymbol{e}}_j, j = 1\ldots 35$ is a unit vector in the j-th direction of the generalized Stokes space.

Using (10.76), the MD vector evolution equation reads as

$$\frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\mathrm{d}\boldsymbol{\beta}_\mathrm{d}}{\mathrm{d}\omega} + \frac{\partial \boldsymbol{\beta}_\mathrm{r}}{\partial \omega} + (\boldsymbol{\beta}_\mathrm{d} + \boldsymbol{\beta}_\mathrm{r})\times\boldsymbol{\tau}\,. \tag{10.77}$$

The term $\partial\boldsymbol{\beta}_\mathrm{r}/\partial\omega$, which accounts for the frequency dependence of the perturbations, contributes to the evolution of the MD vector to a negligible extent as compared to the $\mathrm{d}\boldsymbol{\beta}_\mathrm{d}/\mathrm{d}\omega$, which accounts for the deterministic walk-off between nondegenerate modes, and hence can be ignored. The simplified evolution equation,

$$\frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\mathrm{d}\boldsymbol{\beta}_\mathrm{d}}{\mathrm{d}\omega} + (\boldsymbol{\beta}_\mathrm{d} + \boldsymbol{\beta}_\mathrm{r})\times\boldsymbol{\tau}\,, \tag{10.78}$$

shows that the local contribution to the MD vector $\mathrm{d}\boldsymbol{\beta}_\mathrm{d}/\mathrm{d}\omega$ is constant along the fiber, while the overall z-dependent birefringence vector $\boldsymbol{\beta}_\mathrm{d}(\omega) + \boldsymbol{\beta}_\mathrm{r}(z,\omega)$ rotates the MD vector as it accumulates along the fiber. This dynamics suggests that in the multimode case the statistics of the MD vector depend on the effectiveness with which the MD vector is randomized by the random birefringence, with different results in the two relevant regimes of weak and strong mode coupling.

Like in the single-mode fiber case, an intuitive interpretation of the MD vector can be gained from the first-order picture. In fact, the PSP expansion of the channel transfer matrix $\mathbf{U}$ in (10.67) is generalized to the multimode case in the following form,

$$\mathbf{U}(z,\omega) = \mathrm{e}^{\mathrm{j}t_0\omega}\sum_{n=1}^{2N}\mathrm{e}^{\mathrm{j}t_n\omega}|\tau_n\rangle\langle p_n|\,, \tag{10.79}$$

$$\boldsymbol{E} = \begin{bmatrix} E_{1x} \\ E_{1y} \\ E_{2x} \\ E_{1y} \\ E_{3x} \\ E_{3y} \end{bmatrix} \qquad \mathbf{B}_0 = \begin{bmatrix} 0 & 0 & b & 0 & b & 0 \\ 0 & 0 & 0 & b & 0 & b \\ b & 0 & 0 & 0 & b & 0 \\ 0 & b & 0 & 0 & 0 & b \\ b & 0 & b & 0 & 0 & 0 \\ 0 & b & 0 & b & 0 & 0 \end{bmatrix}$$

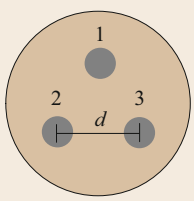


Fig. 10.12 State vector $\boldsymbol{E}$ in a three-core fiber, where the field is represented in the basis of the fundamental modes of the individual fiber cores, and the matrix $\mathbf{B}_0$ describing the deterministic coupling between them

where the output principal states (PSs) $|\tau_n\rangle$ are the $2N$ orthogonal eigenstates of the matrix $\boldsymbol{\tau}\cdot\boldsymbol{\Lambda}$ (they are related to the input PSPs $|p_n\rangle$ through $|\tau_n\rangle = \mathbf{U}(z,0)|p_n\rangle$), and the corresponding delays are referred to the mode-averaged group delay t_0, so that $\sum_n t_n = 0$. Thus, an input signal characterized by the state vector $|u\rangle$, as a result of propagation splits into $2N$ replicas, each delayed by $t_0 + t_n$,

$$f(t)|u\rangle \rightarrow \sum_{n=1}^{2N} \langle p_n|u\rangle f(t - t_0 - t_n)|\tau_n\rangle \,. \tag{10.80}$$

The analytical extraction of the mode delays, which are the eigenvalues of the matrix $(\boldsymbol{\tau}\cdot\boldsymbol{\Lambda})/2N$, in the multimode case is not as straightforward as in the single-mode case (where $t_1 = \tau/2$ and $t_2 = -\tau/2$), however the mode delays are related to the MD vector through the following simple relation

$$\tau^2 = 2N\sum_{n=1}^{2N} t_n^2 \,. \tag{10.81}$$

Within the first-order picture, the most relevant quantity is the largest differential group delay (LDGD), also referred to as the *delay spread* [10.92]), which is defined as the difference between the largest and the smallest of the $2N$ delays. The LDGD is the time interval that needs to be accommodated at the MIMO-DSP receiver, and it obviously affects the complexity of the MIMO-DSP receiver [10.93]. In this framework the statistics of the LDGD is of primary importance, as LDGD fluctuations might cause system outages if not properly accounted for in the receiver design. These considerations underpinned early studies of MD in SDM fibers, which were focused primarily on characterizing the probability density function of the LDGD [10.77, 78]. More recently, however, it became clear that the first-order picture can accurately describe MD in fibers with negligible mode coupling, whereas it is fundamentally inconsistent in the most relevant case of SDM fibers with strong mode coupling [10.80, 81, 94]. These two cases are discussed in what follows.

Modal Dispersion in the Regime of Weak Mode Coupling

Weak coupling between modes results from a large mismatch between the modes' propagation constants. In this regime modal dispersion manifests itself primarily in the form of modal walk-off, where distinct groups of quasi-degenerate modes accumulate a differential delay that increases proportionally to the propagation distance. Using the Stokes-space formalism, this result emerges from (10.78), which by setting $\boldsymbol{\beta}_\mathrm{r} = 0$, yields $\boldsymbol{\tau} = (\mathrm{d}\boldsymbol{\beta}_\mathrm{d}/\mathrm{d}\omega)z$ (this simple result follows from the fact that $\boldsymbol{\beta}_\mathrm{d}$ and $\mathrm{d}\boldsymbol{\beta}_\mathrm{d}/\mathrm{d}\omega$ are parallel vectors, as discussed in [10.80]). In this case, the first-order approximation is legitimate for signals within whose bandwidth the term $\mathrm{d}\boldsymbol{\beta}_\mathrm{d}/\mathrm{d}\omega$ does not vary significantly. In particular, in the familiar case of two uncoupled groups of degenerate modes, this expression of the MD vector can be shown to produce two distinct delays, whose absolute difference is equal to the differential group delay $L|v_{g1}^{-1} - v_{g,2}^{-1}|$, where $v_{g,1}$ and $v_{g,2}$ denote the group velocities of the two mode groups. Modal dispersion within the two groups of modes adds to the much larger intergroup dispersion, implying an almost negligible effect on the MIMO-DSP complexity, which depends primarily on the intergroup differential delay. This regime includes transmission in LP_{01} and LP_{11} mode groups of weakly guiding fibers, under the simplifying assumption of perfect degeneracy of the LP_{11} modes.

Obviously, the regime of weak mode coupling evolves into a regime of intermediate coupling, and eventually of strong coupling, as propagation distance increases. The analysis of this transition and its consequences for the fiber modal dispersion are rather complex and go beyond the purpose of this review. Recent studies on this subject can be found in [10.80, 94–97].

Modal Dispersion in the Regime of Strong Mode Coupling

Modes with similar propagation constants get strongly coupled over relatively short propagation distances, as a result of the fiber's perturbations. In this situation, the effect of the random birefringence vector $\boldsymbol{\beta}_\mathrm{r}$ is dominant and the most relevant properties of the MD vector can be derived by neglecting the deterministic birefringence vector $\boldsymbol{\beta}_\mathrm{d}$ in (10.78). The simplified equation,

$$\frac{\partial\boldsymbol{\tau}}{\partial z} = \frac{\mathrm{d}\boldsymbol{\beta}_\mathrm{d}}{\mathrm{d}\omega} + \boldsymbol{\beta}_\mathrm{r}\times\boldsymbol{\tau} \,, \tag{10.82}$$

differs from the PMD vector evolution equation in the forcing term, which is deterministic. Note that because of the many uncorrelated rotations of the accumulating MD vector driven by the random birefringence vector, the orientation of $\mathrm{d}\boldsymbol{\beta}_\mathrm{d}/\mathrm{d}\omega$ is immaterial, and the same argument used in the single-mode case can hence be used here to conclude that the MD vector evolves as a Gaussian vector too (indeed, direct measurements of the generalized birefringence vector statistics are not available yet, however the observed mode-coupling dynamics indicate that the modal content of the transmitted field in the regime of strong mode coupling is randomized over a few meters, suggesting that the correlation length of the generalized birefringence vector

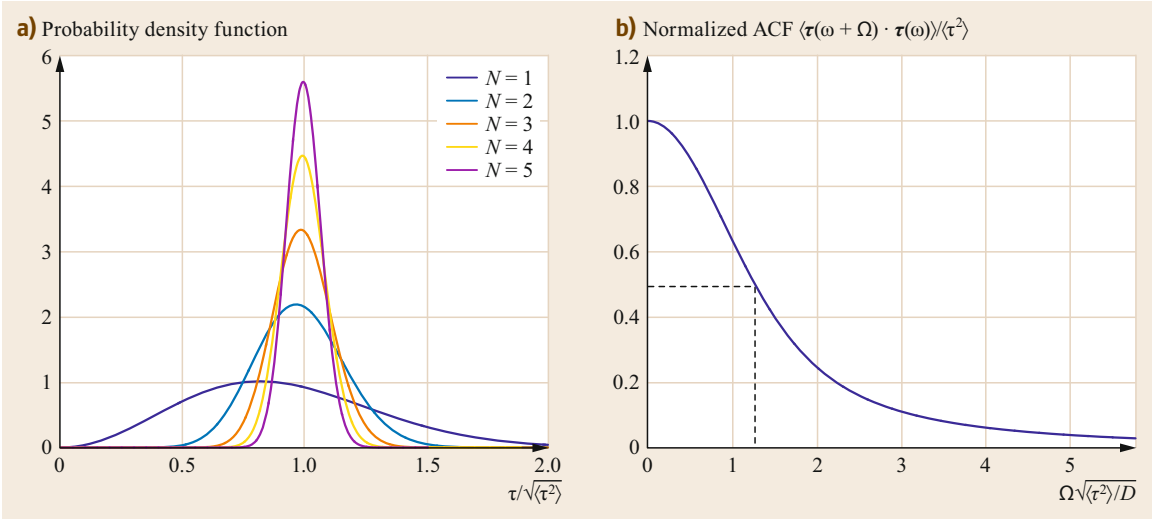


Fig. 10.13 **(a)** The probability density function of the MD vector modulus normalized to its root-mean-square value for various numbers of spatial modes. **(b)** Normalized autocorrelation function of the MD vector (the single-mode case corresponds to $D = 3$)

is of the same order or smaller than in single-mode fibers). Its modulus follows the chi distribution with D degrees of freedom (its square modulus follows the chi-squared distribution), and its mean-square value grows linearly with propagation distance, namely $\langle \tau^2 \rangle = \kappa^2 z$ (however, the dependence of the MD coefficient κ on the fiber design and perturbation statistics is rather complex [10.80]). We remind the reader that a random variable Y is chi-square-distributed with D degrees of freedom if it results from the sum of the squares of D identically distributed and zero-mean independent Gaussian variables X_n: $Y = \sum_{n=1}^{D} X_n^2$. The probability density function of τ is plotted for several values of N in Fig. 10.13a.

A major difference between the single-mode and the multimode case is in the fact that while the PMD vector length scales with the strength of the perturbations, the length of the MD vector scales with the modulus of the deterministic birefringence vector derivative $|\mathrm{d}\boldsymbol{\beta}_\mathrm{d}/\mathrm{d}\omega|$, which can be greater by orders of magnitude, depending on the deterministic walk-off between the fiber modes. An important consequence of this difference is that the MD bandwidth can be correspondingly smaller than the PMD bandwidth. Indeed, the MD vector autocorrelation function has the following form,

$$\langle \boldsymbol{\tau}(z, \omega + \Omega) \cdot \boldsymbol{\tau}(z, \omega) \rangle = \frac{D}{\Omega^2} \left[1 - \mathrm{e}^{-\frac{\Omega^2 \langle \tau^2(z) \rangle}{D}} \right], \tag{10.83}$$

and the MD bandwidth is $B_\mathrm{MD} \simeq 0.2\sqrt{D/\langle \tau^2 \rangle}$, as obtained by inspection of Fig. 10.13b (this expression can also be obtained by multiplying the PMD bandwidth by $\sqrt{D/3}$). It should be noted at this point that, while measurements of the PMD vector and its statistics are routinely performed in traditional single-mode systems, the experimental characterization of the MD vector in SDM systems is more involved [10.99] and therefore the system modal dispersion is typically characterized by exploiting the concept of the *intensity impulse response* (IIR). This is defined as the mode-averaged output power that is measured by exciting a single mode at the fiber input with a spectrally flat signal of bandwidth B. In formulae, we define the matrix $\mathbf{H}(t)$ whose (j, k) element $H_{j,k}(t)$ is the signal received in the j-th mode when the k-th mode was excited,

$$\mathbf{H}(t) = \int_{-B/2}^{B/2} \mathbf{U}(L, \omega) \mathrm{e}^{-\mathrm{j}\omega t} \frac{\mathrm{d}\omega}{2\pi}, \tag{10.84}$$

so that the IIR can be expressed as

$$I(t) = \frac{1}{2N} \sum_{j=1}^{2N} \sum_{k=1}^{2N} |H_{j,k}(t)|^2 . \tag{10.85}$$

Here the inner sum is the total output power that is measured when the j-th mode was excited, while the outer sum performs the mode averaging. If the probing signal bandwidth is sufficiently larger than the MD bandwidth (by one or more orders of magnitude), it can be shown [10.81] that the IIR is deterministic and practically independent of B. Most importantly, its temporal profile is Gaussian and the mean-square duration is very

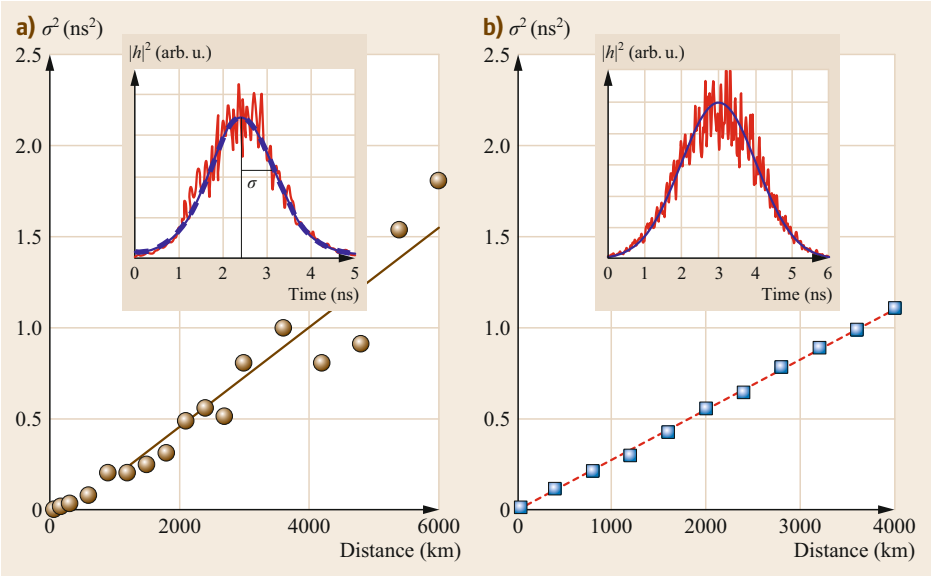


Fig. 10.14a,b The mean-square width of the intensity impulse response versus propagation distance for the coupled-core three-core fiber of [10.98]. The *inset* shows the intensity impulse response for the right-most data point. **(a)** and **(b)** present experimental [10.98] and simulation [10.81] results, respectively. The *dashed curve* in **(b)** is a plot of (10.86) (the relation between the MD coefficient κ and the fiber characteristics is discussed in [10.80])

simply related to the mean-square length of the MD vector (or, equivalently, to the MD bandwidth), namely

$$I(t) = I_0 \exp\left(-\frac{t^2}{2T^2}\right), \tag{10.86}$$

$$T^2 = \frac{\langle \tau^2 \rangle}{4N^2} = \frac{\kappa^2 z}{4N^2}, \tag{10.87}$$

where I_0 is a normalization coefficient immaterial to the present discussion. The Gaussian shape of the IIR has been observed in various experiments [10.53, 98, 100] and can be reproduced in simulations. Figure 10.14 presents a comparison between the measured and simulated IIR for the coupled-core three-core fiber used in [10.98]. The measured IIR mean-square duration of about 0.25 ns^2 at a propagation distance of 1000 km corresponds to an MD bandwidth of approximately 400 MHz, a value much smaller than typical WDM channel bandwidths used today in commercial systems. We remind the reader that, in contrast, typical PMD bandwidth values for the same link length are of the order of several hundreds of GHz (as seen in Sect. 10.5.1 *Polarization-Mode Dispersion in Single-Mode Fibers*).

The above argument shows the inadequacy of the first-order approximation to characterize the MD of SDM fibers for medium-to-long-reach transmission, where modes undergo strong coupling, and at the same time clarifies that a correct approach to designing the MIMO-DSP receiver must rely on the knowledge of the IIR duration. Strategies to reduce the receiver complexity include pursuing the reduction of the fiber MD through fiber design optimization. This approach means studying the dependence of the MD coefficient κ on the fiber characteristics (core number/geometry and/or refractive index profile), as well as on the statistics of the fiber perturbations. This is a rather challenging task and only a limited number of preliminary investigations are available in the literature [10.80, 85, 94].

10.5.2 Stokes-Space Analysis of Mode-Dependent Loss and its Impact on Information Capacity

Mode-dependent loss is a nonunitary propagation effect and as such it is responsible for impairing the capacity of SDM systems [10.55, 101–104]. The Stokes-space formalism has proven to be a convenient tool for the modeling of MDL and its impact on system performance. If we denote by S the average transmit power per mode and by $\mathbf{Q}$ the coherency matrix of the propagated amplification noise, the channel spectral efficiency in the absence of channel state information can be expressed as

$$C = \log_2\left[\det\left(\mathbf{I} + S\mathbf{Q}^{-1/2}\mathbf{U}\mathbf{U}^{\dagger}\mathbf{Q}^{-1/2}\right)\right]. \tag{10.88}$$

The matrix $\mathbf{U}\mathbf{U}^{\dagger}$ (which, in the absence of MDL, would equal the identity matrix) is Hermitian and can be expressed in terms of the generalized Pauli matrices,

$$\mathbf{U}\mathbf{U}^{\dagger} = \gamma_0 \left(\mathbf{I} + \boldsymbol{\Gamma} \cdot \boldsymbol{\Lambda}\right), \tag{10.89}$$

where γ_0 is the mode-averaged gain and the Stokes vector $\boldsymbol{\Gamma}$ is the MDL vector that generalizes the familiar PDL vector used in the single-mode fiber case. In the regime of strong mode coupling and large signal-to-noise ratio (SNR), the average spectral efficiency reduction per mode induced by MDL has been shown to be [10.103, 104]

$$\frac{C_0 - \langle C \rangle}{2N} = \frac{\langle \Gamma^2 \rangle}{3\ln(2)}, \tag{10.90}$$

where C_0 is the spectral efficiency of a perfect link, where the received SNR equals the ratio between the mode-averaged signal and the noise powers. The accuracy of (10.90) is excellent for SNR values larger than 10 dB [10.104]. A simple method for measuring $\langle \Gamma^2 \rangle$ is presented in [10.103].

A quantity which is often used as a figure of merit in the analysis of MDL is the power ratio between the least and the most attenuated hyperpolarization states, which is given by

$$\rho_{\mathrm{dB}} = 10 \log_{10}\left(\frac{1+\lambda_{\max}}{1+\lambda_{\min}}\right), \tag{10.91}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ denote the largest and smallest eigenvalues of $\boldsymbol{\Gamma} \cdot \boldsymbol{\Lambda}$ (note that the corresponding loss/gain values that are measured in experiments are $\gamma_0(1+\lambda_{\max})$ and $\gamma_0(1+\lambda_{\min})$, respectively). Interestingly, in the regime of small-to-moderate MDL the mean-square length of the MDL vector is related to this quantity by the following simple relation,

$$\langle \rho_{\mathrm{dB}}^2 \rangle = \frac{10^2}{\ln^2(10)} f(N) \langle \Gamma^2 \rangle \tag{10.92}$$

with

$$f(N) = 4\frac{(N-1)^2 + 24.7(N-1) + 16.14}{0.2532(N-1)^2 + 7.401(N-1) + 16.14}. \tag{10.93}$$

This connects the average MDL-induced spectral efficiency reduction per mode caused by MDL (10.90) with the mean-square MDL expressed in logarithmic units,

$$\frac{C_0 - \langle C \rangle}{2N} = \frac{\ln^2(10)}{300\ln(2) f(N)} \langle \rho_{\mathrm{dB}}^2 \rangle. \tag{10.94}$$

This expression does not depend on the specific way in which the in-line amplifiers are operated, as discussed in [10.104].

10.6 SDM Transmission Experiments

Numerous transmission experiments have been performed over multimode fibers with the numbers of spatial modes ranging from 3 to 45 [10.14, 18, 19, 105–109]. Also, multicore fibers have been studied experimentally in detail for many possible core arrangements up to 36 cores and spatial multiplicity (number of cores × number of modes) larger than 100.

Space-division multiplexed transmission experiments are very equipment-intensive: A typical SDM transmission experiment for six spatial channels is shown in Fig. 10.15.

The transmitter consists of a traditional WDM signal, where odd and even wavelength channels are modulated separately by two double-nested Mach–Zehnder (DN-MZM) modulators driven with four independent signals carrying the underlaying transmission pattern, like for example QPSK, 16-QAM, or 64-QAM, generated by high-speed digital-to-analog converters (DACs), where pseudo-random patterns are chosen such that the cross-correlation peaks between patterns are significantly smaller than the autocorrelation peaks. This is required to properly identify the timing of the received channels, and to evaluate their performance using digital signal processing [10.121]. Additional copies of the signal are generated and decorrelated using fiber delays such that each mode and polarization carries a locally independent signal. The decorrelated signals are then injected into a six-fold recirculating loop arrangement, which is used to emulate long-distance experiments (often in SDM experiments only limited lengths of prototype fibers are available).

The loop arrangement is similar to a traditional SMF loop, except that it consists of six loops which have to be adjusted to a path-length difference of typically within 1 cm, corresponding to a time delay of 50 ps. The loop contains amplifiers to overcome the fiber loss and the loss of the additional loop components, loop switches (that are used to open and close the loop during the loading and recirculation time, respectively), combiners and splitters (to inject and extract the light from the loop), and finally programmable gain equalizing filters (denoted as blockers in Fig. 10.15), to maintain a flat spectrum after each recirculation.

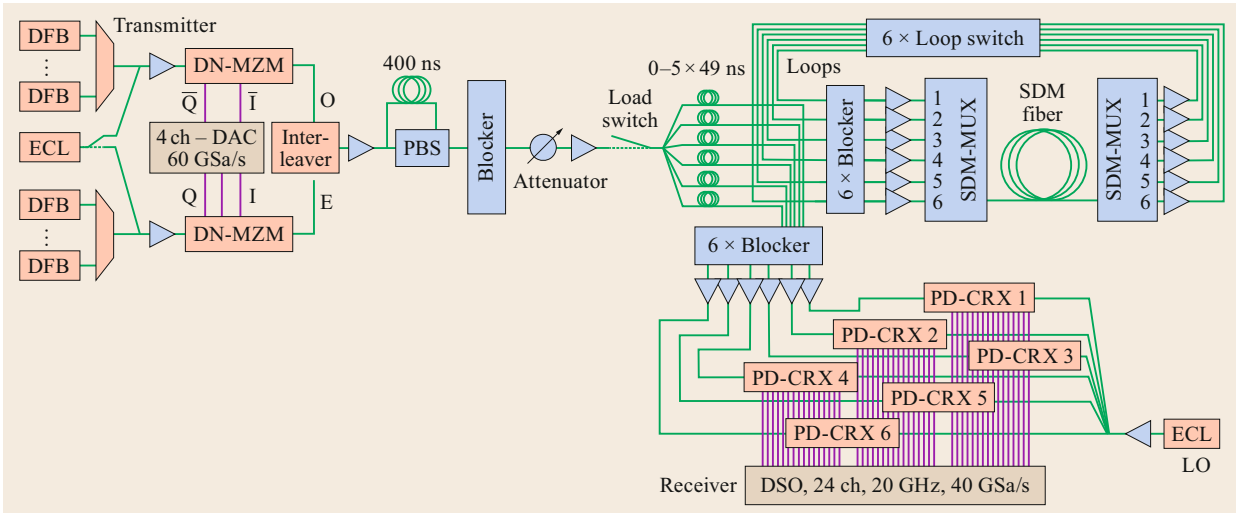


Fig. 10.15 Space-division multiplexed transmission experiment supporting six spatial channels. *Triangles* represent erbium-doped fiber amplifiers (EDFAs), PBS is a polarizing beam splitter, DSO is a digital storage oscilloscope, ECL is an external-cavity tunable laser, DFB is a distributed feedback laser, DN-MZM is a double-nested Mach–Zehnder modulator, and PD-CRX is a polarization-diverse coherent receiver

Table 10.3 Summary of relevant MIMO-based transmission results in SDM fibers

Fiber type	Nr spatial channels	Spectral efficiency ((bit/s)/Hz)	Distance (km)	Spectral efficiency distance ((bit/s)/(Hz km))	Capacity (Tb/s)	Reference
FMF	45	202	27	5454	101	[10.18]
FMF	36	72	2	144	3.6	[10.18]
FMF	15	43.63	22.8	995	17.5	[10.109]
FMF	10	58	87	5046	67.5	[10.110]
FMF	10	29	125	3625	115.2	[10.108]
FMF	6	36.7	90	3303	266	[10.111]
FMF	6	34.9	590	20 355	138	[10.112]
FMF	6	32	176	5632	24.6	[10.113]
CC-MCF	6	18	1705	30 690	18	[10.114]
FMF	6	16	708	11 328	6.1	[10.115]
FMF	6	10	74	740	41.6	[10.116]
CC-MCF	4	23	4400	101 400	11.5	[10.117]
CC-MCF	4	11.52	10 000	115 200	5.8	[10.117]
FMF	3	28.22	30	846	280	[10.118]
FMF	3	18.82	1045	19 663	159	[10.119]
FMF	3	17.3	2400	41 520	0.58	[10.17]
FMF	3	7.6	1000	7600	13.3	[10.44]
CC-MCF	3	4	4200	16 800	1.0	[10.98]
FMF	3	3	900	2700	9.6	[10.120]

The signals extracted from the loops are captured by an array of polarization-diverse coherent receivers (PD-CRXs), which extract the amplitude and phase of all modes and polarizations, so that the optical field after transmission is fully known. Note that it is necessary to measure all modes and polarization for the same time windows, therefore a digital storage oscilloscope (DSO) with 24 real-time channels is required for a transmission with 6 spatial modes (alternatively, time-multiplexed receiver schemes, where subsets of modes are delayed by single-mode fibers, have been proposed to reduce the number of ports that are necessary in the DSO [10.109, 122]).

The resulting signals are stored in the DSO, and subsequently processed by applying MIMO-DSP techniques, similar to the methods presented in Chap. 6.

Some representative results of MIMO-based transmission in multimode and coupled-core fibers are summarized in Table 10.3. By the terms *capacity* and *spectral efficiency* in the table, and more in general

Table 10.4 Summary of relevant SDM transmission in uncoupled multicore fibers

Fiber type	Nr Spatial channels	Spectral efficiency ((bit/s)/Hz)	Distance (km)	Spectral efficiency distance ((bit/s)/(Hz km))	Capacity (Tb/s)	Reference
FM-MCF	6 × 19	1100	11.3	12 430	10 160	[10.123]
FM-MCF	6 × 19	456	9.8	4469	2050	[10.124]
FM-MCF	3 × 12	247.9	40	9916	5.1	[10.125]
MCF	22	207	31	6417	2150	[10.126]
MCF	19	30.5	10.1	307	305	[10.127]
MCF[a]	14	109	3	327	1050	[10.128]
MCF	12	91.4	55	4753	1010	[10.129]
MCF[b]	12	73.6	1500	110 374	688	[10.130]
MCF	12	58.3	8830	515 000	520	[10.131]
MCF	7	28	7326	205 128	140.7	[10.132]
MCF	7	15	2688	40 320	9	[10.133]
MCF	7	14.4	6160	88 704	28.8	[10.134]
MCF	7	11.2	16.8	188.2	109	[10.135]

MCF: multicore fiber, FM-MCF multicore fiber with few-mode cores; [a] 12 single-mode cores and 2 few-mode cores with 3 spatial modes; [b] Core interleaved bidirectional transmission

in this review of experimental results, we refer to the largest achieved transmission rate, and to the same quantity divided by the total transmission bandwidth, respectively.

The longest transmission distances and highest spectral-efficiency-distance products were demonstrated in CC-MCFs, clearly confirming the advantages of the strong coupling regime. The maximum experimental capacity demonstrated in MIMO-SDM transmission clearly surpasses the largest reported values for single-mode fibers. In particular, the largest spectral efficiency demonstrated is as high as 202 (bit/s)Hz which is well above the nonlinear Shannon limit for single-mode fibers [10.1, 136] which is 26.5 (bit/s)Hz for a fiber length of 27 km, indicating that mode-multiplexed transmission over a few-mode fiber (FMF), that is, a fiber that supports 10 or fewer modes, has the technical potential to be considered as a replacement for single-mode fibers.

Transmission results for some representative multicore fiber transmission experiments are summarized in Table 10.4.

Multicore fibers, especially in combination with few-mode cores, can achieve spatial multiplicities larger than 100, providing an impressive transmission capacity in excess of 10 Pb/s, however only for distances shorter than 100 km. Longer distances up to 8800 km can be achieved using single-mode cores at a notable capacity of 520 Tb, which is of interest in particular for submarine transmission, where multiple parallel paths can achieve superior performance under a constraint of limited power [10.131].

10.7 Nonlinear Effects in SDM Fibers

In the previous sections we only considered linear effects in multimode fiber propagation. However, the transmission capacity of multimode systems, just like in the single-mode counterpart [10.1], is ultimately limited by nonlinear effects. The theory of nonlinearities in multimode fibers is challenging as all possible interactions between all involved modes have to be considered. Nonlinear multimode propagation is described by the coupled nonlinear Schrödinger equations [10.66, 67, 137]. If, for ease of discussion, we neglect loss and mode-dependent chromatic dispersion, the equations can be expressed as follows

$$\frac{\partial \boldsymbol{E}}{\partial z} = \mathrm{j}\mathbf{B}_0\boldsymbol{E} - \mathbf{B}_1\frac{\partial \boldsymbol{E}}{\partial t} - \mathrm{j}\frac{\beta_2}{2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} + \mathrm{j}\gamma\sum_{h,k,m,n=1}^{2N} C_{nhkm}E_h^* E_k E_m \hat{\boldsymbol{u}}_n \,, \tag{10.95}$$

where $\mathbf{B}_0 = \mathbf{B}(z,\omega_0)$ and $\mathbf{B}_1 = \partial\mathbf{B}(z,\omega_0)/\partial\omega$ account for random mode coupling and intermodal walk-off, respectively, β_2 is the mode-averaged chromatic disper-

sion coefficient, γ is the nonlinearity coefficient defined for single-mode fibers [10.138], and where by $\hat{\boldsymbol{u}}_n$ we denote a $2N$-dimensional column vector whose n-th component is equal to one and the others to zero. The nonlinearity coefficients C_{nhkm} involve overlap integrals between the modes, lateral profile functions and their expressions can be found in [10.66] and references therein. As can be seen in (10.95), the Kerr nonlinearity produces a total of $(2N)^4$ coefficients ($(2N)^3$ coefficients per mode) that have to be considered in the study of nonlinear effects. This can be a challenging task, especially when the modal properties vary strongly between modes, and in general only detailed numerical simulations will provide representative results [10.137]. In contrast, when all modal properties are similar, like in the case of strongly coupled fibers, theoretical results have indicated a significant advantage for strongly coupled SDM fibers over equivalent single-mode fibers [10.66, 139]. In the following two sections we briefly describe nonlinear experimental work performed in few-mode fibers and coupled-core multicore fibers.

10.7.1 Impact of Nonlinearities in the Strong-Coupling Regime

Coupled-core fibers are interesting for the study of nonlinear effects because all fiber modes have similar modal properties in terms of effective area and propagation coefficients, and hence the electric field propagates in the regime of strong mode coupling described earlier in this chapter. One important consequence of this situation is that the nonlinear term that appears in the propagation equation (10.95) can be drastically simplified by taking into account the fact that the length-scale on which random mode coupling is effective is by orders of magnitude smaller than typical nonlinear length-scales. The simplified propagation equation, which is known as the multicomponent Manakov equation [10.140], is in the form

$$\frac{\partial \boldsymbol{E}}{\partial z} = -\beta_1 \frac{\partial \boldsymbol{E}}{\partial t} - \mathrm{j}\frac{\beta_2}{2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} + \mathrm{j}\gamma\kappa|\boldsymbol{E}|^2\boldsymbol{E}\,, \tag{10.96}$$

where β_1 is the inverse group velocity common to all modes, and where the nonlinearity appears through the total optical power only, consistent with the fact that the electric field is isotropically distributed in the $2N$-dimensional hyperpolarization space. As can be seen, the $(2N)^4$ nonlinearity coefficients C_{nhkm} are replaced by a single coefficient κ, which is given by [10.140]

$$\kappa = \sum_{h,n} \frac{C_{nhhn} + C_{nhnh}}{2N(2N+1)}. \tag{10.97}$$

Equations (10.96) and (10.97) describe nonlinear propagation in the most general case of $2N$ strongly coupled modes. In the specific case of coupled-core fibers, which is considered in this section, (10.97) can be further simplified, with the result [10.66]

$$\kappa\gamma = \frac{1}{3}\frac{8}{2N+1}\gamma_0\,, \tag{10.98}$$

where γ_0 is the nonlinearity coefficient of a single-mode fiber with the same radius and refractive-index profile of the individual cores (for $N = 1$, (10.98) yields $\gamma\kappa = \frac{8}{9}\gamma_0$, the nonlinearity coefficient of the famous Manakov equation describing nonlinear propagation in single-mode fibers with random polarization coupling [10.141, 142]).

The scaling of κ with the number of modes is key to understanding the improved tolerance of coupled-core multicore fibers to nonlinear distortions. This can be easily seen by expressing the nonlinear term as $\gamma\kappa|\boldsymbol{E}|^2 \sim 4\gamma_0/3\sum_n |E_n|^2/2N$, which shows that the various modes can be considered as sources of nonlinear noise whose power is proportional to $1/(2N)^2$. Since they carry independent signals, the total nonlinear noise power results from the sum of the individual contributions and hence it scales like $\sim 2N\times 1/(2N)^2 = 1/2N$, thereby reducing with the number of strongly coupled modes supported by the fiber [10.66, 139]. A formal characterization of the nonlinear interference noise can be found in [10.143]. Note that while (10.98) is an analytical result derived specifically for coupled-core multicore fibers, the scaling $\gamma\kappa \sim 1/N$ is a more general characteristic of fibers operating in the regime of strong mode mixing. The simple argument underpinning this statement is that random mode coupling distributes the power transmitted in each mode equally between all modes, with the result that on average the nonlinearity must be proportional to the mode-averaged power, which is equal to $|\boldsymbol{E}|^2/2N$.

The superior tolerance of coupled-core multicore fibers to nonlinear distortions, as analytically predicted in [10.66, 139] and seen in early simulation work [10.144], has recently been confirmed in transmission experiments performed with a four-core fiber [10.117]. The results of an experimental comparison between a single-mode fiber [10.145] and a four-core coupled-core fiber [10.146] with nominally identical cores and the same span length are shown in Fig. 10.16.

The pure-silica core design realizes an ultralow loss and larger effective area high-performance fiber typically utilized in submarine links (see also Chap. 2 for more detail). Figure 10.16a shows the quality factor Q as a function of the launch power per wavelength

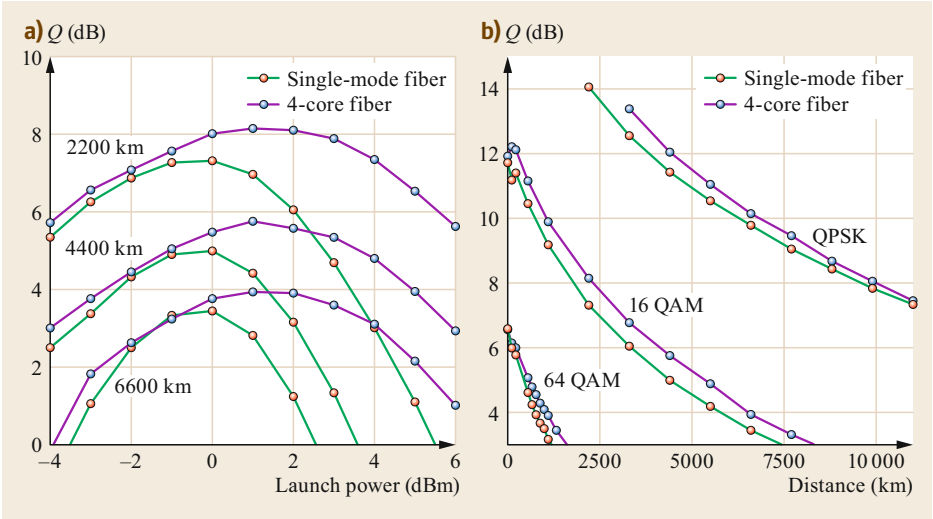


Fig. 10.16a,b Transmission performance comparison between a single-mode fiber and a 4-core coupled-core fiber with identical length and core design for a WDM signal with 15 channels at a baudrate of 30 GBd and a channel spacing of 33.33 GHz. **(a)** Quality factor Q as a function of the launch power for distances of 2200, 4400, and 6600 km, for a 16 QAM signal. **(b)** Q factor as a function of distance for QPSK, 16 QAM, and 64 QAM modulated signals

channel in a recirculating-loop system with 110-km-long spans. As can be clearly seen, the optimum launch power for the coupled-core fiber is about 2 dB larger, indicating a better tolerance to nonlinearities, which results in Q factors that are about 0.8 dB larger. Figure 10.16b compares the launch-power-optimized Q factors as functions of the propagation distance in the same recirculating-loop experiment, and for different modulation formats. The results clearly show that for all tested formats and distances up to 10 000 km the coupled-core fiber outperforms the equivalent single-mode fiber.

10.7.2 Impact of Nonlinearities in Few-Mode Fibers

Nonlinear effects in few-mode fibers are different than in single-mode fibers, as the modal properties allow for phase-matching conditions that are forbidden in single-mode fibers (see also Chap. 9 for a description of Kerr nonlinearities in single-mode fibers). For example, four-wave mixing is strongly suppressed in nonzero-dispersion single-mode fibers, because of the impact of chromatic dispersion on the phase-matching condition. In few-mode fibers, however, modal dispersion can compensate for chromatic dispersion, and therefore strong four-wave mixing can be observed. The effect can be better understood considering cross-phase modulation, where the intensity fluctuations of a signal traveling in one mode can imprint a phase on a second signal traveling in another mode. If both signals travel at the same group velocity, the interaction length for this effect becomes long, and a strong effect can be observed. As the group velocity depends on wavelength and mode, in low DGD few-mode fibers (like optimized GI fibers) it is possible to find conditions where two different modes at two different wavelengths have a matched group delay. This effect was experimentally observed in a fiber with three spatial modes [10.147] and a length of 5 km, confirming that the effect does not degrade significantly even in the presence of perturbations along the fiber. Similar experiments were also reported for fully nondegenerate four-wave mixing [10.148, 149], also confirming that four-wave mixing effects in few-mode fibers are non-negligible and can provide significant penalties for mode-multiplexed MIMO-based transmission.

As for the modeling of nonlinear propagation in few-mode fibers, we note that a similar simplification of the coupled NLSEs as in the case of coupled-core multicore fibers is obtained by taking into account the fact that modes belonging to the same group of quasi-degenerate modes mix strongly during propagation. The result is a set of coupled multicomponent Manakov equations, which in the case of two mode groups denoted a and b can be expressed in the following form,

$$\frac{\partial \boldsymbol{E}_\mathrm{a}}{\partial z} = -\beta_{1,\mathrm{a}}\frac{\partial \boldsymbol{E}_\mathrm{a}}{\partial t} - \mathrm{j}\frac{\beta_{2,\mathrm{a}}}{2}\frac{\partial^2 \boldsymbol{E}_\mathrm{a}}{\partial t^2} + \mathrm{j}\gamma\left(\kappa_\mathrm{a}|\boldsymbol{E}_\mathrm{a}|^2 + \kappa_\mathrm{ab}|\boldsymbol{E}_\mathrm{b}|^2\right)\boldsymbol{E}_\mathrm{a}\,, \tag{10.99}$$

$$\frac{\partial \boldsymbol{E}_\mathrm{b}}{\partial z} = -\beta_{1,\mathrm{b}}\frac{\partial \boldsymbol{E}_\mathrm{b}}{\partial t} - \mathrm{j}\frac{\beta_{2,\mathrm{b}}}{2}\frac{\partial^2 \boldsymbol{E}_\mathrm{b}}{\partial t^2} + \mathrm{j}\gamma\left(\kappa_\mathrm{ab}|\boldsymbol{E}_\mathrm{a}|^2 + \kappa_\mathrm{b}|\boldsymbol{E}_\mathrm{b}|^2\right)\boldsymbol{E}_\mathrm{b}\,, \tag{10.100}$$

where $\boldsymbol{E}_a$ and $\boldsymbol{E}_b$ are state vectors of dimensions $2N_a$ and $2N_b$, respectively, which describe the electric field in the two mode groups. The coupled Manakov equations, derived in [10.150] for arbitrary values of N_a and N_b, and in [10.137] for $N_a = N_b = 1$, can be used for the analytical study of intergroup nonlinear effects [10.143] under the assumption of negligible linear coupling between mode groups. However, if the propagating groups couple to a non-negligible extent, (10.99) and (10.100) must be supplemented with additional terms that account for linear intermodal crosstalk, which reduces their analytical tractability significantly [10.66, 95, 151–153].

10.8 Routing in SDM Networks

The signals transmitted over an SDM link, are typically associated with a spatial channel s_n and a wavelength channel λ_m, where the indices n and m identify the respective spatial and wavelength channel. Note that spatial channels are defined either as modes of an optical waveguide, as physically separated light-paths using multiple waveguides, or as a combination of the two.

In conventional wavelength-multiplexed networks, wavelength is used to optically route signals when traversing a network node. For each wavelength of an ingress fiber it is possible to select an egress fiber, as long as the wavelength channel of the egress fiber has not been assigned to another incoming signal at the same wavelength. This limitation is referred to as wavelength blocking and makes the initial network configuration and the subsequent channel provisioning (adding new channel routes in a live network) mathematically more complex, increases the blocking probability, and therefore reduces the capacity of the network [10.154].

For spatial channels similar limitations may occur, for example, when spatial channels are implemented by using distinct spatial modes. In contrast, if the spatial modes are carried by spatially separated waveguides, or separated by a spatial mode multiplexer, the individual spatial channels are all equivalent and can be switched between each other with no restriction.

In the general case, optical networks can be built based on nodes that are capable of switching any wavelength from any spatial channel coming from any direction, to any wavelength and to any spatial channel going to any direction (here we define directions as geographically separate routes and spatial channels as parallel-running channels, either in a single fiber like a multicore fiber, or multiple single-mode fibers hosted in a single conduit or cable). The complexity of such a node in terms of physical implementation and dynamic operation (traffic provisioning) is undesirably larger and is not cost-effective. It is therefore necessary to limit the complexity by forming logical units of switching, to reduce the logical channel number to around 100 channels for each direction. This can be achieved in various ways [10.155], and in the next sections the three most promising approaches to the bundling of wavelength/spatial channels are reviewed.

10.8.1 Parallel Single-Mode Systems

The first approach consists of duplicating conventional single-mode WDM systems and operating them in parallel. This approach in not cost-effective, but represents the baseline to be considered for alternative approaches. The most relevant limitation of this approach is that it is not possible to share resources between the duplicated systems, which may be responsible for significant blocking probability and under-utilization of resources.

10.8.2 Spatial Superchannels

In this approach the spatial channels are bundled together in a fixed number N, and components are used that can perform the equivalent single-mode operation on all N channels at the same time. The term *spatial superchannel* was coined [10.156] in reference to spectral superchannels where multiple subsequent wavelength channels are bundled to form a single spectrally wider transmission channel. Spatial superchannels look similar to a traditional single-mode system in terms of operation, except that the capacity is increased by a factor of N. This concept is particularly attractive because wavelength-selective switches supporting spatial superchannels, can be implemented using the joint switching architecture, where a single switching element can be reused to switch N channels in parallel, therefore effectively increasing the switch capacity of the switching element. This principle is shown in Fig. 10.17, where a tilt mirror is used to switch light between one input and two output superchannels, by tilting the mirrors such that the light is reflected from the input to the desired output.

In Fig. 10.17a a groupwise switching arrangement is shown which requires an N-times larger tilt angle compared to a traditional single-mode switch, whereas Fig. 10.17b shows the advantageous interleaved switching arrangement, where the superchannels can be switched by using tilt angles that are comparable

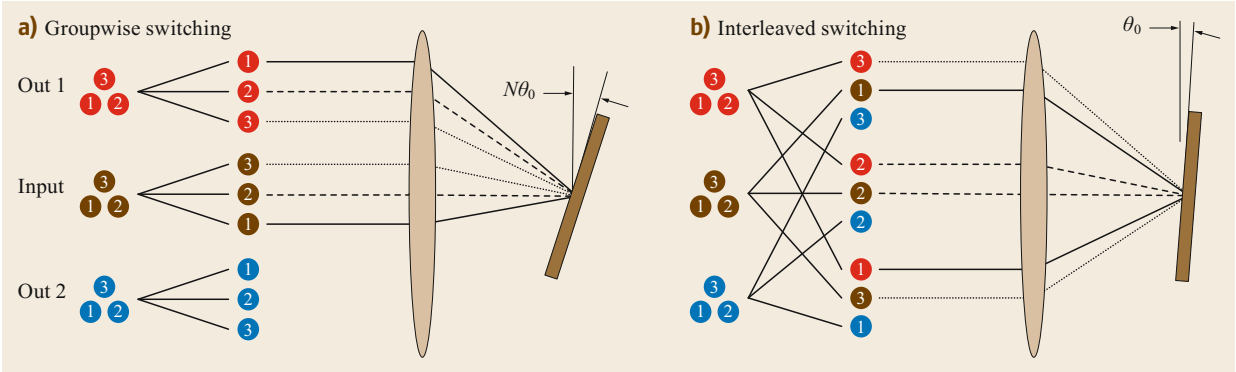


Fig. 10.17a,b Switching multiple spatial channels with a single switching element: Groupwise switching versus interleaved switching geometry

to the single-mode switch. Note that the overall switch size of a spatial superchannel switch is larger than the single-mode counterpart, as it needs to accommodate the larger required optical aperture and a more complex lens design. However, the switch element, which is often the main factor limiting the number of channels that can be switched, can stay the same size. Figure 10.17 shows the principle of a simple switch, but the same idea can also be used to build wavelength-selective switches, where light is separated in wavelength in the out-of-plane direction and the single mirror is replaced with a mirror array [10.155].

Optical networks based on the concept of spatial superchannels can be implemented with multicore fibers, where the number N of parallel channels coincides with the number of cores. The spatial superchannel architecture is of interest also for multimode fibers, where it is required to transmit all fiber modes in a common link, so that MIMO processing can be used to compensate for propagation-induced mode coupling.

Additionally, a spatial superchannel can also be used to logically bundle multiple single-mode fibers and therefore constitutes a very promising architecture for all possible SDM fiber types.

The main drawback of the spatial superchannel architecture is that there is no simple way to increase the number of parallel spatial channels N composing the superchannel once the network is deployed and operated.

10.8.3 Space-Routed Networks

An alternative way to build SDM optical networks is to completely drop the wavelength dimension for the switching domain and utilize pure spatial switching based on traditional switches that are wavelength transparent. This solution offers several advantages: All channels are equivalent and therefore no wavelength blocking is observed, which can dramatically simplify the network reconfiguration. Also, space switches are much easier to build and typically have lower loss compared to wavelength-selective switches. Furthermore, the local add/drop ports of the nodes become significantly simpler, as they are equivalent to ports carrying traffic from fibers coming from different directions.

The disadvantage of this solution is that it requires fibers with completely uncoupled spatial channels, and therefore it is not compatible with MIMO-based multimode transmission. The solution is therefore particularly attractive for single-mode and uncoupled multicore fibers. Space-routed networks also require transceivers capable of generating signals that occupy the whole transmission band. Potentially, such full-band transceivers are expected to be more economical as they offer a larger potential for integration. However, the network can suffer from granularity issues if the desired link capacity is small compared to the capacity of the full-band receiver.

10.9 Conclusion

Space-division multiplexing addresses the technologies needed to scale the link and network capacities of current optical communication systems. The main proposed solutions include new fiber types, optical amplifiers, and optical switches.

As an alternative to standard single-mode fibers, multimode fibers and multicore fibers offer effective ways to increase the spatial multiplicity of optical fibers, at the expense of more complex linear and nonlinear transmission effects that we reviewed in detail in the chapter. The

linear transmission effects in a system supporting N spatial modes can be mitigated by using $2N \times 2N$ MIMO digital signal processing, which is a generalization of the 2×2 MIMO processing used in conventional single-mode digital coherent transmission. Nonlinear impairments are typically moderately increased compared to the single-mode case, when the spatial modes are weakly coupled, whereas a reduction of nonlinear effects can be observed in the regime of strong mode mixing, like for example in the case of coupled-core multicore fibers.

Cladding- and core-pumped optical multicore and multimode amplifiers offer a sizeable potential for cost reduction by significantly reducing the number of required optical elements per amplified spatial channel.

Optical switches supporting multiple modes or spatial channels can be effectively implemented by using *joint-switching* architectures, which dramatically increase the switching capacity of the switching element by acting on all spatial channels at the same time.

Space-division multiplexing also enables multiple new network architectures. Even though no single one-size-fits-all SDM architecture is currently transpiring, technologies that are currently being investigated, have the potential to offer a significant advantage in terms of costs over parallelizing conventional single-mode fiber-based systems. The optimum solution will depend on the targeted application and in particular on the required link capacity and network granularity.

References

10.1 R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel: Capacity limits of optical fiber networks, J. Lightwave Technol. **28**(4), 662–701 (2010)

10.2 P. Winzer: Spatial multiplexing in fiber optics: The 10X scaling of metro/core capacities, Bell Labs Tech. J. **19**, 22–30 (2014)

10.3 D.J. Richardson, J.M. Fini, L.E. Nelson: Space-division multiplexing in optical fibres, Nat. Photonics **7**(5), 354–362 (2013)

10.4 G. Li, N. Bai, N. Zhao, C. Xia: Space-division multiplexing: The next frontier in optical communication, Adv. Opt. Photonics **6**(4), 413–487 (2014)

10.5 R. Ryf, N.K. Fontaine, H. Chen, B. Guan, B. Huang, M. Esmaeelpour, A.H. Gnauck, S. Randel, S. Yoo, A. Koonen, R. Shubochkin, Y. Sun, R. Lingle: Mode-multiplexed transmission over conventional graded-index multimode fibers, Opt. Express **23**(1), 235–246 (2015)

10.6 P. Sillard, D. Molin, M. Bigot-Astruc, A. Amezcua-Correa, K. de Jongh, F. Achten: 50 μm multimode fibers for mode division multiplexing, J. Lightwave Technol. **34**(8), 1672–1677 (2016)

10.7 P. Sillard, D. Molin, M. Bigot, A. Amezcua-Correa, K. de Jongh, F. Achten: DMGD-compensated links. In: *Opt. Fiber Commun. Conf.* (2017), https://doi.org/10.1364/OFC.2017.Tu2J.4

10.8 P. Winzer, A. Gnauck, A. Konczykowska, F. Jorge, J.-Y. Dupuy: Penalties from in-band crosstalk for advanced optical modulation formats. In: *37th Eur. Conf. Opt. Commun.* (2011), https://doi.org/10.1364/ECOC.2011.Tu.5.B.7

10.9 J. Sakaguchi, W. Klaus, J.M.D. Mendinueta, B.J. Puttnam, R.S. Luís, Y. Awaji, N. Wada, T. Hayashi, T. Nakanishi, T. Watanabe, Y. Kokubun, T. Takahata, T. Kobayashi: Large spatial channel (36-core $\times$ 3 mode) heterogeneous few-mode multicore fiber, J. Lightwave Technol. **34**(1), 93–103 (2016)

10.10 S. Matsuo, K. Takenaga, Y. Sasaki, Y. Amma, S. Saito, K. Saitoh, T. Matsui, K. Nakajima, T. Mizuno, H. Takara, Y. Miyamoto, T. Morioka: High-spatial-multiplicity multicore fibers for future dense space-division-multiplexing systems, J. Lightwave Technol. **34**(6), 1464–1475 (2016)

10.11 T. Mizuno, K. Shibahara, H. Ono, Y. Abe, Y. Miyamoto, F. Ye, T. Morioka, Y. Sasaki, Y. Amma, K. Takenaga, S. Matsuo, K. Aikawa, K. Saitoh, Y. Jung, D.J. Richardson, K. Pulverer, M. Bohn, M. Yamada: 32-core dense SDM unidirectional transmission of PDM-16QAM signals over 1600 km using crosstalk-managed single-mode heterogeneous multicore transmission line. In: *Opt. Fiber Commun. Conf.* (2016), https://doi.org/10.1364/OFC.2016.Th5C.3

10.12 A. Turukhin, H. Batshon, M. Mazurczyk, Y. Sun, C. Davidson, J. Chai, O. Sinkin, W. Patterson, G. Wolter, M. Bolshtyansky: Demonstration of 0.52 Pb/s potential transmission capacity over 8,830 km using multicore fiber. In: *42nd Eur. Conf. Opt. Commun.* (2016)

10.13 P. Pepeljugoski, F.E. Doany, D. Kuchta, B. Lee, C.L. Schow, L. Schares: Connector performance analysis for D-shaped multi-core multi mode fiber. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.Th4J.4

10.14 R. Ryf, S. Randel, A.H. Gnauck, C. Bolle, A. Sierra, S. Mumtaz, M. Esmaeelpour, E.C. Burrows, R.J. Essiambre, P.J. Winzer, D.W. Peckham, A.H. McCurdy, R. Lingle: Mode-division multiplexing over 96 km of few-mode fiber using coherent 6 x 6 MIMO processing, J. Lightwave Technol. **30**(4), 521–531 (2012)

10.15 L. Grüner-Nielsen, Y. Sun, J.W. Nicholson, D. Jakobsen, K.G. Jespersen, J.R. Lingle, B. Pálsdóttir: Few mode transmission fiber with low DGD, low mode coupling, and low loss, J. Lightwave Technol. **30**(23), 3693–3698 (2012)

10.16 P. Sillard, D. Molin, M. Bigot-Astruc, K. de Jongh, F. Achten, J.E. Antonio-López, R. Amezcua-Correa: Micro-bend-resistant low-differential-mode-group-delay few-mode fibers, J. Lightwave Technol. **35**(4), 734–740 (2017)

10.17 J. van Weerdenburg, R. Ryf, R. Alvarez-Aguirre, N.K. Fontaine, R.J. Essiambre, H. Chen, J.C. Alvarado-Zacarias, R. Amezcua-Correa, S. Gross, N. Riesen, M. Withford, D.W. Peckham, A. McCurdy, R. Lingle, T. Koonen, C. Okonkwo: Mode-multiplexed 16-QAM transmission over 2400-km large-effective-area depressed-cladding 3-mode fiber. In: *Opt. Fiber Commun. Conf.* (2018), https://doi.org/10.1364/OFC.2018.W4C.2

10.18 R. Ryf, N.K. Fontaine, H. Chen, S. Wittek, J. Li, J.C. Alvarado-Zacarias, R. Amezcua-Correa, J.E. Antonio-Lopez, M. Capuzzo, R. Kopf, A. Tate, H. Safar, C. Bolle, D.T. Neilson, E. Burrows, K. Kim, M. Bigot-Astruc, F. Achten, P. Sillard, A. Amezcua-Correa, J. Du, Z. He, J. Carpenter: Mode-multiplexed transmission over 36 spatial modes of a graded-index multimode fiber. In: *Eur. Conf. Opt. Commun.* (2018), https://doi.org/10.1109/ECOC.2018.8535431

10.19 R. Ryf, N.K. Fontaine, S. Wittek, K. Choutagunta, M. Mazur, H. Chen, J.C. Alvarado-Zacarias, R. Amezcua-Correa, M. Capuzzo, R. Kopf, A. Tate, H. Safar, C. Bolle, D.T. Neilson, E. Burrows, K. Kim, M. Bigot-Astruc, F. Achten, P. Sillard, A. Amezcua-Correa, J.M. Kahn, J. Schröder, J. Carpenter: High-spectral-efficiency mode-multiplexed transmission over graded-index multimode fiber. In: *Eur. Conf. Opt. Commun.* (2018), https://doi.org/10.1109/ECOC.2018.8535536

10.20 N. Riesen, J.D. Love, J.W. Arkwright: Few-mode elliptical-core fiber data transmission, IEEE Photonics Technol. Lett. **24**(5), 344–346 (2012)

10.21 L. Wang, S. LaRochelle: Design of eight-mode polarization-maintaining few-mode fiber for multiple-input multiple-output-free spatial division multiplexing, Opt. Lett. **40**(24), 5846–5849 (2015)

10.22 W. Wang, J. Zhao, H. Yu, Z. Yang, Y. Zhang, Z. Zhang, C. Guo, G. Li: Demonstration of 6 $\times$10 Gb/s MIMO-free polarization- and mode-multiplexed transmission, IEEE Photonics Technol. Lett. **30**(15), 1372–1375 (2018)

10.23 N. Bozinovic, Y. Yue, Y. Ren, M. Tur, P. Kristensen, H. Huang, A.E. Willner, S. Ramachandran: Terabit-scale orbital angular momentum mode division multiplexing in fibers, Science **340**(6140), 1545–1548 (2013)

10.24 N.K. Fontaine, R. Ryf, J. Bland-Hawthorn, S.G. Leon-Saval: Geometric requirements for photonic lanterns in space division multiplexing, Opt. Express **20**(24), 27123–27132 (2012)

10.25 J. Carpenter, T. Wilkinson: Characterization of multimode fiber by selective mode excitation, J. Lightwave Technol. **30**(10), 1386–1392 (2012)

10.26 M. Salsi, C. Koebele, D. Sperti, P. Tran, P. Brindel, H. Mardoyan, S. Bigo, A. Boutin, F. Verluise, P. Sillard, M. Bigot-Astruc, L. Provost, F. Cerou, G. Charlet: Transmission at 2 $\times$ 100Gb/s, over two modes of 40km-long prototype few-mode fiber, using LCOS based mode multiplexer and demultiplexer. In: *Opt. Fiber Commun. Conf.* (2011), https://doi.org/10.1364/NFOEC.2011.PDPB9

10.27 H. Chen, R. van Uden, C. Okonkwo, T. Koonen: Compact spatial multiplexers for mode division multiplexing, Opt. Express **22**(26), 31582 (2014)

10.28 Y. Ding, H. Ou, J. Xu, C. Peucheret: Silicon photonic integrated circuit mode multiplexer, IEEE Photonics Technol. Lett. **25**(7), 648–651 (2013)

10.29 K. Igarashi, D. Souma, K. Takeshima, T. Tsuritani: Selective mode multiplexer based on phase plates and Mach-Zehnder interferometer with image inversion function, Opt. Express **23**(1), 183–194 (2015)

10.30 S.H. Chang, H.S. Chung, N.K. Fontaine, R. Ryf, K.J. Park, K. Kim, J.C. Lee, J.H. Lee, B.Y. Kim, Y.K. Kim: Mode division multiplexed optical transmission enabled by all-fiber mode multiplexer, Opt. Express **22**(12), 14229–14236 (2014)

10.31 S.G. Leon-Saval, N.K. Fontaine, J.R. Salazar-Gil, B. Ercan, R. Ryf, J. Bland-Hawthorn: Mode-selective photonic lanterns for space-division multiplexing, Opt. Express **22**(1), 1036–1044 (2014)

10.32 T.A. Birks, S. Yerolatsitis, I. Gris-Sánchez: Fibre-based mode multiplexers. In: *Adv. Photonics* (2014), https://doi.org/10.1364/SOF.2014.SoW1B.1

10.33 G. Labroille, B. Denolle, P. Jian, P. Genevaux, N. Treps, J.-F. Morizur: Efficient and mode selective spatial mode multiplexer based on multi-plane light conversion, Opt. Express **22**(13), 15599–15607 (2014)

10.34 S.G. Leon-Saval, T.A. Birks, J. Bland-Hawthorn, M. Englund: Multimode fiber devices with single-mode performance, Opt. Lett. **30**(19), 2545–2547 (2005)

10.35 S. Yerolatsitis, I. Gris-Sánchez, T.A. Birks: Adiabatically-tapered fiber mode multiplexers, Opt. Express **22**(1), 608–617 (2014)

10.36 I. Giles, R. Chen, V. Garcia-Munoz: Fiber based multiplexing and demultiplexing devices for few mode fiber space division multiplexed communications. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.Tu3D.1

10.37 A.M. Velázquez-Benítez, J.E. Antonio-Lopez, J.C. Alvarado-Zacarias, G. Lopez-Galmiche, P. Sillard, D.V. Ras, C. Okonkwo, H. Chen, R. Ryf, N.K. Fontaine, R. Amezcua-Correa: Scaling the fabrication of higher order photonic lanterns using microstructured preforms. In: *Eur. Conf. Opt. Commun.* (2015), https://doi.org/10.1109/ECOC.2015.7341939

10.38 B. Huang, N.K. Fontaine, R. Ryf, B. Guan, S.G. Leon-Saval, R. Shubochkin, Y. Sun, R. Lingle, G. Li: All-fiber mode-group-selective photonic lantern using graded-index multimode fibers, Opt. Express **23**(1), 224–234 (2015)

10.39 B. Guan, B. Ercan, N.K. Fontaine, R. Ryf, H. Chen, R.P. Scott, Y. Zhang, S.J.B. Yoo: 15-spatial-mode photonic lanterns based on ultrafast laser inscription. In: *Eur. Conf. Opt. Commun* (2015), https://doi.org/10.1109/ECOC.2015.7341651

10.40 N. Riesen, S. Gross, J.D. Love, Y. Sasaki, M.J. Withford: Femtosecond laser written integrated spatial multiplexers for few-mode multicore fibre. In: *42nd Eur. Conf. Opt. Commun.* (2016) pp. 1064–1066

10.41 N.K. Fontaine, R. Ryf, H. Chen, D. Neilson, J. Carpenter: Design of high order mode-multiplexers using multiplane light conversion. In: *43rd Eur. Conf. Opt. Commun.* (2017), https://doi.org/10.1109/ECOC.2017.8346129

10.42 S. Bade, B. Denolle, G. Trunet, N. Riguet, P. Jian, O. Pinel, G. Labroille: Fabrication and characterization of a mode-selective 45-mode spatial multiplexer based on multi-plane light conversion. In: *Opt. Fiber Commun. Conf.* (2018), https://doi.org/10.1364/OFC.2018.Th4B.3

10.43 N.K. Fontaine, R. Ryf, H. Chen, D.T. Neilson, K. Kim, J. Carpenter: Scalable mode sorter supporting 210 Hermite-Gaussian modes. In: *Opt. Fiber Commun. Conf.* (2018), https://doi.org/10.1364/OFC.2018.Th4B.4

10.44 E. Ip, M.-J. Li, K. Bennett, Y.-K. Huang, A. Tanaka, A. Korolev, K. Koreshkov, W. Wood, E. Mateo, J. Hu, Y. Yano: $146\lambda \times 6 \times 19$-Gbaud wavelength-and mode-division multiplexed transmission over 10 $\times$ 50-km spans of few-mode fiber with a gain-equalized few-mode EDFA, J. Lightwave Technol. **32**(4), 790–797 (2014)

10.45 M. Salsi, R. Ryf, G. Le Cocq, L. Bigot, D. Peyrot, G. Charlet, S. Bigo, N.K. Fontaine, M.A. Mestre, S. Randel, X. Palou, C. Bolle, B. Guan, Y. Quiquempois: A six-mode erbium-doped fiber amplifier. In: *Eur. Conf. Exhib. Opt. Commun.* (2012), https://doi.org/10.1364/ECEOC.2012.Th.3.A.6

10.46 N. Bai, E. Ip, Y.-K. Huang, E. Mateo, F. Yaman, M.-J. Li, S. Bickham, S. Ten, J. Liñares, C. Montero, V. Moreno, X. Prieto, V. Tse, K.M. Chung, A.P.T. Lau, H.-Y. Tam, C. Lu, Y. Luo, G.-D. Peng, G. Li, T. Wang: Mode-division multiplexed transmission with inline few-mode fiber amplifier, Opt. Express **20**(3), 2668–2680 (2012)

10.47 R. Nasiri Mahalati, D. Askarov, J.M. Kahn: Adaptive modal gain equalization techniques in multimode erbium-doped fiber amplifiers, J. Lightwave Technol. **32**(11), 2133–2143 (2014)

10.48 K.S. Abedin, J.M. Fini, T.F. Thierry, B. Zhu, M.F. Yan, L. Bansal, F.V. Dimarcello, E.M. Monberg, D.J. DiGiovanni: Seven-core erbium-doped double-clad fiber amplifier pumped simultaneously by side-coupled multimode fiber, Opt. Lett. **39**(4), 993–996 (2014)

10.49 N. Fontaine, B. Guan, R. Ryf, H. Chen, A. Koonen, S. Ben Yoo, K. Abedin, J. Fini, T. Taunay, D. Neilson: Programmable gain equalizer for multi-core fiber amplifiers. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1109/OFC.2014.6887232

10.50 H. Takeshita, K. Matsumotc, E.L.T. de Gabory: Transmission of 200Gbps PM-16QAM signal through 7-core MCF and MC-EDFA using novel turbo cladding pumping scheme for improved efficiency of the optical amplification. In: *Eur. Conf. Opt. Commun.* (2018), https://doi.org/10.1109/ECOC.2018.8535546

10.51 Y. Jung, E.L. Lim, Q. Kang, T.C. May-Smith, N.H.L. Wong, R. Standish, F. Poletti, J.K. Sahu, S.U. Alam, D.J. Richardson: Cladding pumped few-mode EDFA for mode division multiplexed transmission, Opt. Express **22**(23), 29008–29013 (2014)

10.52 N.K. Fontaine, B. Huang, Z.S. Eznaveh, H. Chen, C. Jin, B. Ercan, A. Velázquez-Benítez, S.H. Chang, R. Ryf, A. Schülzgen, J.C. Alvarado, P. Sillard, C. Gonnet, E. Antonio-Lopez, R. Amezcua-Correa: Multi-mode optical fiber amplifier supporting over 10 spatial modes. In: *Opt. Fiber Commun. Conf.* (2016), https://doi.org/10.1364/OFC.2016.Th5A.4

10.53 S. Randel, P.J. Winzer, M. Montoliu, R. Ryf: Complexity analysis of adaptive frequency-domain equalization for MIMO-SDM transmission. In: *39th Eur. Conf. Opt. Commun.* (2013), https://doi.org/10.1049/cp.2013.1540

10.54 R. Ryf, S. Chandrasekhar, S. Randel, D.T. Neilson, N.K. Fontaine, M. Feuer: Physical layer transmission and switching solutions in support of spectrally and spatially flexible optical networks, IEEE Commun. Mag. **53**(2), 52–59 (2015)

10.55 P.J. Winzer, G.J. Foschini: MIMO capacities and outage probabilities in spatially multiplexed optical transport systems, Opt. Express **19**(17), 16680–16696 (2011)

10.56 H. Kogelnik: Theory of optical waveguides. In: *Guided-Wave Optoelectronics*, 2nd edn., ed. by T. Tamir (Springer, Berlin, Heidelberg 1988) pp. 7–88

10.57 D. Marcuse: *Theory of Dielectric Optical Waveguides* (Academic Press, San Diego 1991)

10.58 R.M. Nejad, L. Wang, J. Lin, S. Larochelle, L.A. Rusch: Parasitic effect of TE and TM modes in OAM-MDM transmission systems. In: *Conf. Lasers Electro-Optics* (2017), https://doi.org/10.1364/CLEO_SI.2017.SW4I.2

10.59 P. Gregg, P. Kristensen, A. Rubano, S. Golowich, L. Marrucci, S. Ramachandran: Spin-orbit coupled, non-integer OAM fibers: Unlocking a new Eigenbasis for transmitting 24 uncoupled modes. In: *Conf. Lasers Electro-Optics* (2016), https://doi.org/10.1364/CLEO_AT.2016.JTh4C.7

10.60 D.L.P. Vitullo, C.C. Leary, P. Gregg, R.A. Smith, D.V. Reddy, S. Ramachandran, M.G. Raymer: Observation of interaction of spin and intrinsic orbital angular momentum of light, Phys. Rev. Lett. **118**, 083601 (2017)

10.61 D. Gloge: Weakly guiding fibers, Appl. Opt. **10**(10), 2252–2258 (1971)

10.62 W.P. Huang, L. Li: Coupled-mode theory for optical waveguides: An overview, J. Opt. Soc. Am. A **11**(3), 963–983 (1994)

10.63 A.W. Snyder: Coupled-mode theory for optical fibers, J. Opt. Soc. Am. **62**(11), 1267–1277 (1972)

10.64 J.M. Fini, B. Zhu, T.F. Taunay, M.F. Yan: Statistics of crosstalk in bent multicore fibers, Opt. Express **18**(14), 15122–15129 (2010)

10.65 T. Hayashi, T. Taru, O. Shimakawa, T. Sasaki, E. Sasaoka: Design and fabrication of ultra-low crosstalk and low-loss multi-core fiber, Opt. Express **19**(17), 16576–16592 (2011)

10.66 C. Antonelli, M. Shtaif, A. Mecozzi: Modeling of nonlinear propagation in space-division mul-

tiplexed fiber-optic transmission, J. Lightwave Technol. **34**(1), 36–54 (2016)
10.67 F. Poletti, P. Horak: Description of ultrashort pulse propagation in multimode optical fibers, J. Opt. Soc. Am. B **25**(10), 1645–1654 (2008)
10.68 M. Kolesik, J.V. Moloney: Nonlinear optical pulse propagation simulation: From Maxwell's to unidirectional equations, Phys. Rev. E **70**, 036604 (2004)
10.69 A.A. Juarez, E. Krune, S. Warm, C.A. Bunge, K. Petermann: Modeling of mode coupling in multimode fibers with respect to bandwidth and loss, J. Lightwave Technol. **32**(8), 1549–1558 (2014)
10.70 C. Antonelli, A. Mecozzi, M. Shtaif, P.J. Winzer: Nonlinear propagation equations in fibers with multiple modes–transitions between representation bases, APL Photonics **4**, 022806 (2019)
10.71 H. Kogelnik, P.J. Winzer: Modal birefringence in weakly guiding fibers, J. Lightwave Technol. **30**(14), 2240–2245 (2012)
10.72 R.C. Jones: A new calculus for the treatment of optical systems. Description and discussion of the calculus, J. Opt. Soc. Am. **31**(7), 488–493 (1941)
10.73 J.P. Gordon, H. Kogelnik: PMD fundamentals: Polarization mode dispersion in optical fibers, Proc. Natl. Acad. Sci. U.S.A. **97**(9), 4541–4550 (2000)
10.74 N. Frigo: A generalized geometrical representation of coupled mode theory, IEEE J. Quantum Electron. **22**(11), 2131–2140 (1986)
10.75 A. Galtarossa, L. Palmieri, M. Schiano, T. Tambosso: Measurement of birefringence correlation length in long, single-mode fibers, Opt. Lett. **26**(13), 962–964 (2001)
10.76 S. Fan, J.M. Kahn: Principal modes in multimode waveguides, Opt. Lett. **30**(2), 135–137 (2005)
10.77 K.-P. Ho, J.M. Kahn: Statistics of group delays in multimode fiber with strong mode coupling, J. Lightwave Technol. **29**(21), 3119–3128 (2011)
10.78 C. Antonelli, A. Mecozzi, M. Shtaif, P.J. Winzer: Stokes-space analysis of modal dispersion in fibers with multiple mode transmission, Opt. Express **20**(11), 11718–11733 (2012)
10.79 I. Roudas, J. Kwapisz: Stokes space representation of modal dispersion, IEEE Photonics J. **9**(5), 1–15 (2017)
10.80 C. Antonelli, A. Mecozzi, M. Shtaif: The delay spread in fibers for SDM transmission: Dependence on fiber parameters and perturbations, Opt. Express **23**(3), 2196–2202 (2015)
10.81 A. Mecozzi, C. Antonelli, M. Shtaif: Intensity impulse response of SDM links, Opt. Express **23**(5), 5738–5743 (2015)
10.82 L. Palmieri, A. Galtarossa: Coupling effects among degenerate modes in multimode optical fibers, IEEE Photonics J. **6**(6), 1–8 (2014)
10.83 F. Curti, B. Daino, G.D. Marchis, F. Matera: Statistical treatment of the evolution of the principal states of polarization in single-mode fibers, J. Lightwave Technol. **8**(8), 1162–1166 (1990)
10.84 A. Galtarossa, L. Palmieri, A. Pizzinat, M. Schiano, T. Tambosso: Measurement of local beat length and differential group delay in installed single-mode fibers, J. Lightwave Technol. **18**(10), 1389 (2000)
10.85 L. Palmieri, A. Galtarossa: Intramodal dispersion properties of step-index few-mode spun fibers, J. Lightwave Technol. **34**(2), 303–313 (2016)
10.86 M. Karlsson, J. Brentel: Autocorrelation function of the polarization-mode dispersion vector, Opt. Lett. **24**(14), 939–941 (1999)
10.87 M. Shtaif, A. Mecozzi, J.A. Nagel: Mean-square magnitude of all orders of polarization mode dispersion and the relation with the bandwidth of the principal states, IEEE Photonics Technol. Lett. **12**(1), 53–55 (2000)
10.88 M. Shtaif, A. Mecozzi: Study of the frequency autocorrelation of the differential group delay in fibers with polarization mode dispersion, Opt. Lett. **25**(10), 707–709 (2000)
10.89 M. Shtaif: Performance degradation in coherent polarization multiplexed systems as a result of polarization dependent loss, Opt. Express **16**(18), 13918–13932 (2008)
10.90 S.J. Savory: Digital filters for coherent optical receivers, Opt. Express **16**(2), 804–817 (2008)
10.91 M.S. Faruk, K. Kikuchi: Adaptive frequency-domain equalization in digital coherent optical receivers, Opt. Express **19**(13), 12789–12798 (2011)
10.92 S. Arik, D. Askarov, J.M. Kahn: Effect of mode coupling on signal processing complexity in mode-division multiplexing, J. Lightwave Technol. **31**(3), 423–431 (2013)
10.93 S.O. Arik, K.-P. Ho, J.M. Kahn: Group delay management and multiinput multioutput signal processing in mode-division multiplexing systems, J. Lightwave Technol. **34**(11), 2867–2880 (2016)
10.94 S.O. Arik, K.-P. Ho, J.M. Kahn: Delay spread reduction in mode-division multiplexing: Mode coupling versus delay compensation, J. Lightwave Technol. **33**(21), 4504–4512 (2015)
10.95 C. Antonelli, A. Mecozzi, M. Shtaif, P.J. Winzer: Random coupling between groups of degenerate fiber modes in mode multiplexed transmission, Opt. Express **21**(8), 9484–9490 (2013)
10.96 F.M. Ferreira, D. Fonseca, H.J.A. da Silva: Design of few-mode fibers with M-modes and low differential mode delay, J. Lightwave Technol. **32**(3), 353–360 (2014)
10.97 F.M. Ferreira, C.S. Costa, S. Sygletos, A.D. Ellis: Semi-analytical modelling of linear mode coupling in few-mode fibers, J. Lightwave Technol. **35**(18), 4011–4022 (2017)
10.98 R. Ryf, R. Essiambre, A. Gnauck, S. Randel, M.A. Mestre, C. Schmidt, P. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, T. Hayashi, T. Taru, T. Sasaki: Space-division multiplexed transmission over 4200 km 3-core microstructured fiber. In: *Opt. Fiber Commun. Conf.* (2012), https://doi.org/10.1364/OFC.2012.PDP5C.2
10.99 C. Antonelli, A. Mecozzi, M. Shtaif, P.J. Fontaine, H. Chen, R. Ryf: Stokes-space analysis of modal dispersion of sdm fibers with mode-dependent loss: theory and experiments, J. Lightwave Technol. **38**(7), 1668–1677 (2020)

10.100 R. Ryf, J.C. Alvarado, B. Huang, J. Antonio-Lopez, S.H. Chang, N.K. Fontaine, H. Chen, R.J. Essiambre, E. Burrows, R. Amezcua-Correa, T. Hayashi, Y. Tamura, T. Hasegawa, T. Taru: Long-distance transmission over coupled-core multi-core fiber. In: *42nd Eur. Conf. Opt. Commun.* (2016)

10.101 K.-P. Ho, J.M. Kahn: Mode-dependent loss and gain: statistics and effect on mode-division multiplexing, Opt. Express **19**(17), 16612–16635 (2011)

10.102 K.P. Ho, J.M. Kahn: Frequency diversity in mode-division multiplexing systems, J. Lightwave Technol. **29**(24), 3719–3726 (2011)

10.103 A. Andrusier, M. Shtaif, C. Antonelli, A. Mecozzi: Assessing the effects of mode-dependent loss in space-division multiplexed systems, J. Lightwave Technol. **32**(7), 1317–1322 (2014)

10.104 C. Antonelli, A. Mecozzi, M. Shtaif, P.J. Winzer: Modeling and performance metrics of MIMO-SDM systems with different amplification schemes in the presence of mode-dependent loss, Opt. Express **23**(3), 2203–2219 (2015)

10.105 R. Ryf, S. Randel, A.H. Gnauck, C. Bolle, R.-J. Essiambre, P.J. Winzer, D.W. Peckham, A. McCurdy, R. Lingle: Space-division multiplexing over 10 km of three-mode fiber using coherent 6 x 6 MIMO processing. In: *Opt. Fiber Commun. Conf.* (2011), https://doi.org/10.1364/OFC.2011.PDPB10

10.106 E. Ip, N. Bai, Y.-K. Huang, E. Mateo, F. Yaman, M.-J. Li, S. Bickham, S. Ten, J. Linares, C. Montero, V. Moreno, X. Prieto, V. Tse, K.M. Chung, A. Lau, H.-Y. Tam, C. Lu, Y. Luo, G.-D. Peng, G. Li: 88 x 3 x 112-Gb/s WDM transmission over 50 km of three-mode fiber with inline few-mode fiber amplifier. In: *37th Eur. Conf. Exhib. Opt. Commun.* (2011)

10.107 R. Ryf, N.K. Fontaine, M.A. Mestre, S. Randel, X. Palou, C. Bolle, A.H. Gnauck, S. Chandrasekhar, X. Liu, B. Guan, R.-J. Essiambre, P.J. Winzer, S. Leon-Saval, J. Bland-Hawthorn, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, L. Grüner-Nielsen, R.V. Jensen, R. Lingle: 12 × 12 MIMO transmission over 130-km few-mode fiber. In: *Front. Opt.* (2012), https://doi.org/10.1364/FIO.2012.FW6C.4

10.108 R. Ryf, H. Chen, N.K. Fontaine, A.M. Velazquez-Benitez, J. Antonio-Lopez, C. Jin, B. Huang, M. Bigot-Astruc, D. Molin, F. Achten, P. Sillard, R. Amezcua-Correa: 10-mode mode-multiplexed transmission over 125-km single-span multimode fiber. In: *Eur. Conf. Opt. Commun.* (2015), https://doi.org/10.1109/ECOC.2015.7341687

10.109 N.K. Fontaine, R. Ryf, H. Chen, A.V. Benitez, B. Guan, R. Scott, B. Ercan, S.J.B. Yoo, L.E. Grüner-Nielsen, Y. Sun, R. Lingle, E. Antonio-Lopez, R. Amezcua-Correa: 30×30 MIMO transmission over 15 spatial modes. In: *Opt. Fiber Commun. Conf.* (2015), https://doi.org/10.1364/OFC.2015.Th5C.1

10.110 H. Chen, R. Ryf, N.K. Fontaine, A.M. Velázquez-Benítez, J. Antonio-López, C. Jin, B. Huang, M. Bigot-Astruc, D. Molin, F. Achten, P. Sillard, R. Amezcua-Correa: High spectral efficiency mode-multiplexed transmission over 87-km 10-mode fiber. In: *Opt. Fiber Commun. Conf.* (2016)

10.111 Y. Wakayama, D. Soma, S. Beppu, S. Sumita, K. Igarashi, T. Tsuritani: 266.1-Tbit/s repeatered transmission over 90.4-km 6-mode fiber using dual C+L-band 6-mode EDFA. In: *Opt. Fiber Commun. Conf.* (2018), https://doi.org/10.1364/OFC.2018.W4C.1

10.112 J. van Weerdenburg, R. Ryf, J.C. Alvarado-Zacarias, R.A. Alvarez-Aguirre, N.K. Fontaine, H. Chen, R. Amezcua-Correa, Y. Sun, L. Grüner-Nielsen, R.V. Jensen, R. Lingle, T. Koonen, C. Okonkwo: 138-Tb/s mode- and wavelength-multiplexed transmission over six-mode graded-index fiber, J. Lightwave Technol. **36**(6), 1369–1374 (2018)

10.113 R. Ryf, S. Randel, N.K. Fontaine, M. Montoliu, E. Burrows, S. Chandrasekhar, A.H. Gnauck, C. Xie, R.-J. Essiambre, P. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, L. Gruner-Nielsen, R.V. Jensen, R. Lingle: 32-bit/s/Hz spectral efficiency WDM transmission over 177-km few-mode fiber. In: *Opt. Fiber Commun. Conf.* (2013), https://doi.org/10.1364/OFC.2013.PDP5A.1

10.114 R. Ryf, N.K. Fontaine, B. Guan, R.-J. Essiambre, S. Randel, A.H. Gnauck, S. Chandrasekhar, A. Adamiecki, G. Raybon, B. Ercan, R.P. Scott, S.J.B. Yoo, T. Hayashi, T. Nagashima, T. Sasaki: 1705-km transmission over coupled-core fibre supporting 6 spatial modes. In: *Eur. Conf. Opt. Commun.* (2014), https://doi.org/10.1109/ECOC.2014.6964273

10.115 R. Ryf, S. Randel, N. Fontaine, X. Palou, E. Burrows, S. Corteselli, S. Chandrasekhar, A. Gnauck, C. Xie, R.-J. Essiambre, P. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, L. Gruner-Nielsen, R. Jensen, R. Lingle: 708-km combined WDM/SDM transmission over few-mode fiber supporting 12 spatial and polarization modes. In: *39th Eur. Conf. Opt. Commun.* (2013), https://doi.org/10.1049/cp.2013.1420

10.116 Y. Chen, A. Lobato, Y. Jung, H. Chen, V. Sleiffer, M. Kuschnerov, N. Fontaine, R. Ryf, D. Richardson, B. Lankl, N. Hanik: 41.6 Tbit/s C-band SDM OFDM transmission through 12 spatial and polarization modes over 74.17 km few mode fiber, J. Lightwave Technol. **33**(7), 1440–1444 (2015)

10.117 R. Ryf, N.K. Fontaine, S.H. Chang, J.C. Alvarado, B. Huang, J. Antonio-Lopez, H. Chen, R.J. Essiambre, E. Burrows, R.W. Tkach, R. Amezcua-Correa, T. Hayashi, Y. Tamura, T. Hasegawa, T. Taru: Long-haul transmission over multi-core fibers with coupled cores. In: *Eur. Conf. Opt. Commun.* (2017), https://doi.org/10.1109/ECOC.2017.8345874

10.118 G. Rademacher, R.S. Luis, B.J. Puttnam, R. Ryf, H. Furukawa, R. Maruyama, K. Aikawa, A. Maruta, Y. Awaji, N. Wada: 93.34 Tbit/s/mode (280 Tbit/s) transmission in a 3-mode graded-index few-mode fiber. In: *Opt. Fiber Commun. Conf.* (2018), https://doi.org/10.1364/OFC.2018.W4C.3

10.119 G. Rademacher, R. Ryf, N.K. Fontaine, H. Chen, R.J. Essiambre, B.J. Puttnam, R.S. Luís, Y. Awaji, N. Wada, S. Gross, N. Riesen, M. Withford, Y. Sun, R. Lingle: Long-haul transmission over few-mode

fibers with space-division multiplexing, J. Lightwave Technol. **36**(6), 1382–1388 (2018)

10.120 R. Ryf, N.K. Fontaine, M. Montoliu, S. Randel, B. Ercan, H. Chen, S. Chandrasekhar, A. Gnauck, S.G. Leon-Saval, J. Bland-Hawthorn, J.R.S. Gil, Y. Sun, R. Lingle: Photonic-lantern-based mode multiplexers for few-mode-fiber transmission. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.W4J.2

10.121 A. Sierra, S. Randel, P.J. Winzer, R. Ryf, A.H. Gnauck, R.J. Essiambre: On the use of delay-decorrelated I/Q test sequences for QPSK and QAM signals, IEEE Photonics Technol. Lett. **24**(12), 1000–1002 (2012)

10.122 R.G. Van Uden, C.M. Okonkwo, H. Chen, H. de Waardt, A.M. Koonen: Time domain multiplexed spatial division multiplexing receiver, Opt. Express **22**(10), 12668–12677 (2014)

10.123 D. Soma, Y. Wakayama, S. Beppu, S. Sumita, T. Tsuritani, T. Hayashi, T. Nagashima, M. Suzuki, M. Yoshida, K. Kasai, M. Nakazawa, H. Takahashi, K. Igarashi, I. Morita, M. Suzuki: 10.16-Peta-B/s dense SDM/WDM transmission over 6-mode 19-core fiber across the C+L band, J. Lightwave Technol. **36**(6), 1362–1368 (2018)

10.124 D. Soma, K. Igarashi, Y. Wakayama, K. Takeshima, Y. Kawaguchi, N. Yoshikane, T. Tsuritani, I. Morita, M. Suzuki: 2.05 Peta-bit/s super-Nyquist-WDM SDM transmission using 9.8-km 6-mode 19-core fiber in full C band. In: *Eur. Conf. Opt. Commun.* (2015), https://doi.org/10.1109/ECOC.2015.7341686

10.125 T. Mizuno, T. Kobayashi, H. Takara, A. Sano, H. Kawakami, T. Nakagawa, Y. Miyamoto, Y. Abe, T. Goh, M. Oguma, T. Sakamoto, Y. Sasaki, I. Ishida, K. Takenaga, S. Matsuo, K. Saitoh, T. Morioka: 12-core x 3-mode dense space division multiplexed transmission over 40 km employing multi-carrier signals with parallel MIMO equalization. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.Th5B.2

10.126 B.J. Puttnam, R.S. Luís, W. Klaus, J. Sakaguchi, J.M.D. Mendinueta, Y. Awaji, N. Wada, Y. Tamura, T. Hayashi, M. Hirano, J. Marciante: 2.15 Pb/s transmission using a 22 core homogeneous single-mode multi-core fiber and wideband optical comb. In: *Eur. Conf. Opt. Commun.* (2015), https://doi.org/10.1109/ECOC.2015.7341685

10.127 J. Sakaguchi, B.J. Puttnam, W. Klaus, Y. Awaji, N. Wada, A. Kanno, T. Kawanishi, K. Imamura, H. Inaba, K. Mukasa: 305 Tb/s space division multiplexed transmission using homogeneous 19-core fiber, J. Lightwave Technol. **31**(4), 554–562 (2013)

10.128 D. Qian, E. Ip, M.-F. Huang, M.-J. Li, A. Dogariu, S. Zhang, Y. Shao, Y.-K. Huang, Y. Zhang, X. Cheng, Y. Tian, P. Ji, A. Collier, Y. Geng, J. Linares, C. Montero, V. Moreno, X. Prieto, T. Wang: 1.05Pb/s transmission with 109b/s/Hz spectral efficiency using hybrid single- and few-mode cores. In: *Front. Opt.* (2012), https://doi.org/10.1364/FIO.2012.FW6C.3

10.129 H. Takara, A. Sano, T. Kobayashi, H. Kubota, H. Kawakami, A. Matsuura, Y. Miyamoto, Y. Abe, H. Ono, K. Shikama, Y. Goto, K. Tsujikawa, Y. Sasaki, I. Ishida, K. Takenaga, S. Matsuo, K. Saitoh, M. Koshiba, T. Morioka: 1.01-Pb/s (12 SDM/222 WDM/456 Gb/s) crosstalk-managed transmission with 91.4-b/s/Hz aggregate spectral efficiency. In: *Eur. Conf. Exhib. Opt. Commun.* (2012), https://doi.org/10.1364/ECEOC.2012.Th.3.C.1

10.130 T. Kobayashi, H. Takara, A. Sano, T. Mizuno, H. Kawakami, Y. Miyamoto, K. Hiraga, Y. Abe, H. Ono, M. Wada, Y. Sasaki, I. Ishida, K. Takenaga, S. Matsuo, K. Saitoh, M. Yamada, H. Masuda, T. Morioka: 2 × 344 Tb/s propagation-direction interleaved transmission over 1500-km MCF enhanced by multicarrier full electric-field digital back-propagation. In: *39th Eur. Conf. Exhib. Opt. Commun.* (2013), https://doi.org/10.1049/cp.2013.1691

10.131 A. Turukhin, H.G. Batshon, M. Mazurczyk, Y. Sun, C.R. Davidson, J.X. Chai, O.V. Sinkin, W. Patterson, G. Wolter, M.A. Bolshtyansky, D.G. Foursa, A. Pilipetskii: Demonstration of 0.52 Pb/s potential transmission capacity over 8,830 km using multicore fiber. In: *42nd Eur. Conf. Opt. Commun.* (2016)

10.132 K. Igarashi, T. Tsuritani, I. Morita, Y. Tsuchida, K. Maeda, M. Tadakuma, T. Saito, K. Watanabe, K. Imamura, R. Sugizaki, M. Suzuki: Super-Nyquist-WDM transmission over 7,326-km seven-core fiber with capacity-distance product of 1.03 Exabit/s/km, Opt. Express **22**(2), 1220–1228 (2014)

10.133 S. Chandrasekhar, A.H. Gnauck, X. Liu, P.J. Winzer, Y. Pan, E.C. Burrows, T. Taunay, B. Zhu, M. Fishteyn, M.F. Yan, J.M. Fini, E. Monberg, F. Dimarcello: WDM/SDM transmission of 10 x 128-Gb/s PDM-QPSK over 2688-km 7-core fiber with a per-fiber net aggregate spectral-efficiency distance product of 40,320 km.b/s/Hz, Opt. Express **20**(2), 706–711 (2012)

10.134 H. Takahashi, T. Tsuritani, E.L.T. de Gabory, T. Ito, W.R. Peng, K. Igarashi, K. Takeshima, Y. Kawaguchi, I. Morita, Y. Tsuchida, Y. Mimura, K. Maeda, T. Saito, K. Watanabe, K. Imamura, R. Sugizaki, M. Suzuki: First demonstration of MC-EDFA-repeatered SDM transmission of 40 x 128-Gbit/s PDM-QPSK signals per core over 6,160-km 7-core MCF, Opt. Express **21**(1), 789–795 (2013)

10.135 J. Sakaguchi, Y. Awaji, N. Wada, A. Kanno, T. Kawanishi, T. Hayashi, T. Taru, T. Kobayashi, M. Watanabe: Space division multiplexed transmission of 109-Tb/s data signals using homogeneous seven-core fiber, J. Lightwave Technol. **30**(4), 658–665 (2012)

10.136 P.J. Winzer: Energy-efficient optical transport capacity scaling through spatial multiplexing, IEEE Photonics Technol. Lett. **23**(13), 851–853 (2011)

10.137 S. Mumtaz, R.J. Essiambre, G.P. Agrawal: Nonlinear propagation in multimode and multicore fibers: Generalization of the Manakov equations, J. Lightwave Technol. **31**(3), 398–406 (2013)

10.138 G. Agrawal: Pulse propagation in fibers. In: *Nonlinear Fiber Optics*, 5th edn., ed. by G. Agrawal (Academic Press, Boston 2013) pp. 27–56

10.139 C. Antonelli, A. Mecozzi, M. Shtaif: Scaling of inter-channel nonlinear interference noise and capacity with the number of strongly coupled modes in SDM systems. In: *Opt. Fiber Commun. Conf.* (2016), https://doi.org/10.1364/OFC.2016.W4I.2

10.140 A. Mecozzi, C. Antonelli, M. Shtaif: Nonlinear propagation in multi-mode fibers in the strong coupling regime, Opt. Express **20**(11), 11673–11678 (2012)

10.141 S.V. Manakov: On the theory of two-dimensional stationary self-focusing of electromagnetic waves, Sov. J. Exp. Theor. Phys. **38**, 505–516 (1974)

10.142 D. Marcuse, C.R. Manyuk, P.K.A. Wai: Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence, J. Lightwave Technol. **15**(9), 1735–1746 (1997)

10.143 C. Antonelli, O. Golani, M. Shtaif, A. Mecozzi: Nonlinear interference noise in space-division multiplexed transmission through optical fibers, Opt. Express **25**(12), 13055–13078 (2017)

10.144 S. Mumtaz, R.J. Essiambre, G.P. Agrawal: Reduction of nonlinear penalties due to linear coupling in multicore optical fibers, IEEE Photonics Technol. Lett. **24**(18), 1574–1576 (2012)

10.145 M. Hirano, T. Haruna, Y. Tamura, T. Kawano, S. Ohnuki, Y. Yamamoto, Y. Koyano, T. Sasaki: Record low loss, record high FOM optical fiber with manufacturable process. In: *Opt. Fiber Commun. Conf.* (2013), https://doi.org/10.1364/OFC.2013.PDP5A.7

10.146 T. Hayashi, Y. Tamura, T. Hasegawa, T. Taru: Record-low spatial mode dispersion and ultra-low loss coupled multi-core fiber for ultra-long-haul transmission, J. Lightwave Technol. **35**(3), 450–457 (2017)

10.147 R. Essiambre, M. Mestre, R. Ryf, A. Gnauck, R. Tkach, A. Chraplyvy, Y. Sun, X. Jiang, R. Lingle: Experimental observation of inter-modal cross-phase modulation in few-mode fibers, IEEE Photonics Technol. Lett. **25**(6), 535–538 (2013)

10.148 R.J. Essiambre, M.A. Mestre, R. Ryf, A.H. Gnauck, R.W. Tkach, A.R. Chraplyvy, Y. Sun, X. Jiang, R. Lingle: Experimental investigation of inter-modal four-wave mixing in few-mode fibers, IEEE Photonics Technol. Lett. **25**(6), 539–542 (2013)

10.149 O.F. Anjum, M. Guasoni, P. Horak, Y. Jung, P. Petropoulos, D.J. Richardson, F. Parmigiani: Polarization insensitive four wave mixing based wavelength conversion in few-mode optical fibers, J. Lightwave Technol. **36**(17), 3678–3683 (2018)

10.150 A. Mecozzi, C. Antonelli, M. Shtaif: Coupled Manakov equations in multimode fibers with strongly coupled groups of modes, Opt. Express **20**(21), 23436–23441 (2012)

10.151 A. Mecozzi, C. Antonelli, M. Shtaif: Nonlinear equations of propagation in multi-mode fibers with random mode coupling. In: *39th Eur. Conf. Opt. Commun.* (2013), https://doi.org/10.1049/cp.2013.1453

10.152 F.M. Ferreira, C.S. Costa, S. Sygletos, A.D. Ellis: Nonlinear performance of few-mode fiber links with intermediate coupling, J. Lightwave Technol. **37**(3), 989–999 (2019)

10.153 S. Buch, S. Mumtaz, R.-J. Essiambre, A.M. Tulino, A.D. Agrawal: Averaged nonlinear equations for multimode fibers valid in all regimes of random linear coupling, Opt. Fiber Technol. **48**, 123–127 (2019)

10.154 C. Nuzman, J. Leuthold, R. Ryf, S. Chandrasekhar, C.R. Giles, D.T. Neilson: Design and implementation of wavelength-flexible network nodes, J. Lightwave Technol. **21**(3), 648 (2003)

10.155 D.M. Marom, P.D. Colbourne, A. D'errico, N.K. Fontaine, Y. Ikuma, R. Proietti, L. Zong, J.M. Rivas-Moscoso, I. Tomkos: Survey of photonic switching architectures and technologies in support of spatially and spectrally flexible optical networking, J. Opt. Commun. Netw. **9**(1), 1–26 (2017)

10.156 L. Nelson, M. Feuer, K. Abedin, X. Zhou, T. Taunay, J. Fini, B. Zhu, R. Isaac, R. Harel, G. Cohen, D. Marom: Spatial superchannel routing in a two-span ROADM system for space division multiplexing, J. Lightwave Technol. **32**(4), 783–789 (2014)

Roland Ryf

Advanced Photonics Research
Nokia Bell Labs
Holmdel, NJ, USA
roland.ryf@nokia-bell-labs.com

Roland Ryf is the director of the Advanced Photonics Research department at Nokia Bell Labs, Holmdel, NJ, where since 2000 he has been working on optical switches and wavelength filters, multimode components, and optical amplifiers, and numerous first experimental demonstrations of long-distance high-capacity space-division multiplexed transmission over multimode fibers and coupled-core multicore fibers. He is a fellow of the IEEE and the Optical Society of America.

Cristian Antonelli

Dept. of Physical and Chemical Sciences
University of L'Aquila
L'Aquila, Italy
cristian.antonelli@univaq.it

Cristian Antonelli received his PhD in 2006 from the University of L'Aquila, where he joined the Department of Physical and Chemical Sciences in 2014. Previously he was with AT&T Labs, NJ, the Research Laboratory of Electronics at MIT, and CNISM, the Italian inter-university consortium for the physics of matter. His work focuses on the modeling and characterization of fiber-optic communication systems.

Part B Core Networks

Core Netw

Part B Core Networks

Ed. by Biswanath Mukherjee

Optical networks form the high-capacity backbone of our global Internet. They provide the worldwide information superhighways that are essential to our economies and personal lives. Optical networks transport signals encoded onto light over fiber-optic cables made from glass fibers. These signals are transmitted across a telecommunications network consisting of various optical switching and processing nodes that are connected together in appropriate network architectures. Optical networks can transmit signals at extremely high bandwidths (in the Tbps regime). Since the 1980s, these networks have been a key enabling technology for cloud networks and the Internet, which are currently used to perform the vast majority of all human-to-human and machine-to-machine information data transmission.

Optical networks are used for a huge range of applications. For example, they are employed to link continents using submerged optical cables, and to connect different countries and the major cities of each country with one another using terrestrial cables. They are also applied in metropolitan area and access networks to link to end users and individual homes (i.e., fiber-to-the-home or FTTH networks), and optical networking has recently been used in datacenters.

This part (B) of the Handbook focuses on optical core networks, which provide long-distance, wide-area network coverage. Note that Part A (Optical Subsystems for Transmission and Switching) dealt with the enabling technologies upon which optical networks are built, while Part C (Datacenter and Supercomputer Networks) focuses on the application of optical networking in datacenters and supercomputers and Part D (Access Networks) covers optical access networks that connect to the core network and provide coverage to individual homes and businesses.

The material in Part B is organized as described below.

Core (backbone) network architectures as deployed by telecommunication carriers are studied in Chapter 11. While the focus is on core networks, the chapter also provides invaluable information on how the core network interworks with various types of access networks, such as FTTH networks, hybrid fiber-coaxial networks used by cable network operators, and radio access networks used by cellular network operators. This chapter also outlines the resilience strategies used by telecom carriers in their networks to ensure that network connectivity is rapidly recovered if there is a network outage. This chapter is a must-read, as it provides a clear perspective on practical core network architectures for beginners as well as those experienced in this field.

In a core network, when a node needs to communicate with another node, an efficient (and typically inexpensive) route needs to be selected from among many possible routes that may be available and then established. Also, the huge capacity of an optical fiber is carved up into smaller bandwidth chunks (typically wavelengths) to address the optoelectronic bandwidth mismatch, so an electronic transmitter can transit into a wavelength. This gives rise to the routing and wavelength assignment (RWA) problem and the light path concept for connecting two end points through a wavelength across the optical core network. Recently, to address the extensive and diverse bandwidth capabilities of transmitters and the requirements of applications/services, it is has become possible to assign bandwidth spectra of various sizes, leading to the routing and spectrum assignment (RSA) problem. Such networks are also called elastic optical networks. These topics are covered in Chapter 12.

As optical networks have matured, their design and operational standards have also evolved. These issues are covered in Chapter 13.

Chapter 14 addresses a companion topic: traffic grooming, where a connection that is set up between any two end nodes can not only carry a single service between the end points but also 'groom' (i.e., appropriately multiplex, switch at intermediate nodes, and demultiplex) multiple lower-capacity services onto higher-capacity light paths.

Chapter 15 investigates automation of the control and management of optical networks for fast optical service provisioning, as well as automated recovery from failures. It covers network discovery, dynamic service provisioning, and seamless fault recovery for optical control planes ranging from fully distributed to more recent software-defined network (SDN) frameworks.

Chapter 16 consists of two parts: (1) cross-layer design, focusing on the interactions between the physical/optical and network layers; (2) design aspects relating to the interactions between the network and application layers. Given that optical signals degrade in quality as they travel through the fiber, it is important to understand physical-layer impairments (PLIs) so that the signal can be detected at the destination. PLI-aware RWA is therefore a key issue. Also, the optical network should ultimately serve the needs of the application, so application-aware optical networking is very important too.

Virtualization is a major topic in networking, as the vast resources of a physical network can be carved up into smaller 'slices,' where each slice can, for example, be operated as an independent virtual network, pro-

vide a different service, or be dedicated to a different operation/operator. Aspects of optical network virtualization such as methods of isolating network slices from each other and the roles of SDNs and network function virtualization (NFV) technologies are discussed in Chapter 17.

Metropolitan area networks (MANs) span cities, regions, districts, and municipalities, and they connect core optical networks to access networks as well as providing broadband networking to enterprises. Chapter 18 describes MAN architectures as well as relevant technologies such as SONET/SDH, OTN, WDM, IP/MPLS, and carrier ethernet from a MAN perspective.

Energy efficiency is very important for the sustainability of our planet. Chapter 19 reviews various aspects relating to optical network energy efficiency, including the energy used by various hardware components, energy-efficient software control strategies for cross-layer design and operation, and network survivability strategies.

Strategies such as RWA, RSA, and traffic grooming mentioned above are based on optical circuit switching (OCS), which is very practical and widely deployed, particularly since OCS can also accommodate the opto-electronic bandwidth mismatch. However, client traffic can be highly dynamic and bursty, so the past few decades have seen significant interest in all-optical switching of a client's optical packets or group of packets (called a burst), giving rise to optical packet switching (OPS) and optical burst switching (OBS), respectively. While the jury is still out on the practicality of these methods today, they are included in Chapter 20 of this Handbook, as unforeseen developments may eventually make these methods more practical.

11. Carrier Network Architectures and Resiliency

Robert D. Doverspike

In this chapter we investigate carrier network architectures, and how and where resilience is provided by commercial telecommunications carriers in today's optical networks. Much of the content is generally unpublished and we provide the reader a unique insight into this topic through our privilege of working for decades on the inside of different telecommunications carriers. To provide a fuller understanding of this complex topic, we first describe the typical partitioning of terrestrial networks into their metro-access, metro-core, and intermetro segments and then describe the multilayered structure within each of these segments (Sect. 11.2). Within these constructs, we describe where and how network resiliency is provided against a modeled set of potential outages and other network impacting events (Sect. 11.3). To better understand how the resiliency techniques deployed in various layers and segments are engineered, we discuss how end-to-end services are pieced together across these segments to provide their needed network quality of service and availability. Finally, to even better understand why today's network resiliency techniques have been deployed, we take the reader through a historical evolution of how and why key resiliency technologies and methodologies were developed and deployed, including why some phased out (Sect. 11.4).

There are many published articles about optical networks, their architectures, and potential restoration and resiliency techniques. However, there is a great disparity between the published world and the reality of installed, commercial networks. One of the reasons for this disparity is that the exact structure and operation of networks in a large telecommunication carrier (or *carrier*) are generally unpublished. In general, network planners, architects, provisioners, operations, and maintenance personnel do not disclose details about their business. Furthermore, internal organizations are careful about maintaining the proprietary nature of how they run their networks. However, to understand restoration methods and operations that are vital to network resiliency, one must understand both the network architecture and how it operates. Fortunately, we have had

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_11

the privilege of working in or with some of these organizations over many decades and, as a result, the reader will enjoy a realistic development of this topic. In terms of scope, we will concentrate on terrestrial carrier-based optical (and their related) networks.

We note that there is a parallel topic of resiliency within datacenters. Networks inside large datacenters (*intradatacenter* networks) transport orders of magnitude more traffic than the networks that connect those datacenters (*interdatacenter* networks), often via privately owned and operated networks; therefore, they have a substantially different set of architectures, technologies, economics, and quality of service objectives. As such, the topic of intradatacenter networks is a major topic unto itself and is beyond the scope of our chapter. However, there are many private enterprise interdatacenter networks. These networks are usually leased overlays or cutouts from larger commercial networks and, as such, have some of the same resiliency aspects as commercial networks. We will discuss these private networks briefly in Sect. 11.3.6, *Large Enterprise Private Networks*.

Following the script of many historical movies, we will give a present-day view (snapshot) of architectures and their resulting network resiliency and then step backwards in time to discuss how these particular architectures came to be:

- Sect. 11.1 broadly discusses today's network architectures
- Sect. 11.2 examines where and how resiliency is provided in today's network architectures
- Sect. 11.3 presents why networks evolved to their present architectural state, plus looks forward to how they might evolve going forward
- Sect. 11.4 provides a summary and future directions.

As a final introductory comment, we note that there is a plethora of network restoration and resiliency papers in the published literature. It is not our objective to attempt to cover all these different ideas and proposals because there is a large gap between an idea for network restoration or resiliency applied to an abstracted, mathematical network model versus a usable and cost-effective implementation of that idea in a commercial network. For, implementing a restoration method in commercial networks that is reliable, dependable, and cost-effective is technically difficult and fraught with potential hazards. Implementing even the simplest restoration method can take many years of fine-tuning and adaptation and even after years of attempts, some of these ideas never get over the "bar" of operational reliability and performance. We will illustrate some of these real-life experiences as we go along.

11.1 Today's Network Architectures

We begin by describing how most carrier networks are architected today. Most publications study optical network availability and resiliency within a narrow context, often dominated by the abstraction within a contiguous, simplified network model. However, to provide a broader understanding of how resiliency in optical networks is related to overall service, it is important to give a more comprehensive view.

11.1.1 Network Segments and Layer

Figure 11.1 illustrates how we conceptually segment a large national terrestrial network. As mentioned in the introduction, this chapter will focus on *terrestrial* networks and therefore we will suppress the term going forward. We organize the networks of large telecommunications carriers into different segments. A national *intermetro segment* interconnects many metropolitan (metro) areas. Each of these metro areas is further decomposed into a *metro-core segment* and a *metro-access segment*. The intermetro segment directly interfaces to the metro-core segments. In the metro-core segments and intermetro segments, large carriers place most of their equipment in buildings called *central offices* (*COs*). Alternatively, smaller carriers (or sometimes large carriers in less dense areas) may instead use a large *hut* or lease portions of a larger carrier's CO (i.e., they are *colocated*). Today most central offices in developed countries are interconnected by optical fiber. The metro-access segment of the network refers to the portion of the network between a customer location and its first (serving) central office. Note that the term *customer* could include another carrier who purchases wholesale services from an incumbent carrier. While this chapter is focused on optical networks, end-to-end services must often be transported over mixed media in the metro-access segment and therefore we will also consider other media, such as coax and wireless.

The networks in each of these segments are further organized into network layers that consist of nodes (switching or cross-connect equipment) and links (logical adjacencies between the equipment), which we can visually depict as network graphs vertically stacked on top of one another. Links (capacity) of a higher layer network are provided as *connections* in lower layer networks (also called traffic, demand, or circuits, de-

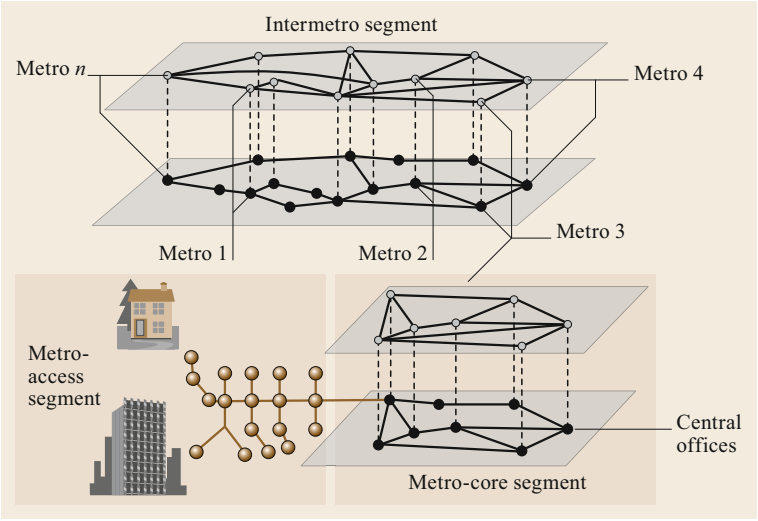


Fig. 11.1 Terrestrial network layers and segmentation

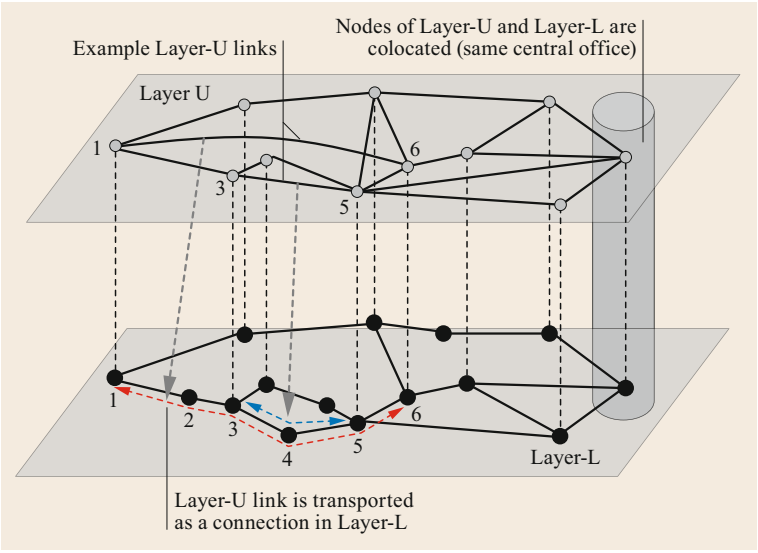


Fig. 11.2 Example of multilayered network graphs

pending on the layer). See [11.1] and [11.2] for more details about the networking and business context of this segmentation.

Before we detail how each of these segments is architected, we provide some background on a simple graph model of multilayer networking and its effect on the topic of network resiliency. This graphical network-layering model is shown in Fig. 11.2, which depicts two layers. This figure and description is adapted from [11.2]. The abstract units of this graph model are nodes, edges (aka links), and connections (aka demands). Nodes represent the cross-connect or switching equipment of a network technology layer. Here, Layer-U (*upper layer*) connects such equipment to their counterpart nodes in Layer-L (*lower layer*) by interlayer links, depicted as lightly dashed vertical lines. While this model has no specific geographical correlation, we note that the switching or cross-connect equipment represented in Layer-U usually are colocated in the same buildings/locations (e.g., COs) as their lower layer counterparts in Layer-L. In such representations, the interlayer links are called *intraoffice* links. The links of Layer-U are transported as connections in lower Layer-L. For example, Fig. 11.2 highlights a link (edge) between nodes 1 and 6 of layer U. This link is transported via a connection between nodes 1 and 6 of Layer-L. The path of this connection is shown through nodes (1, 2, 3, 4, 5, 6) at Layer-L. Another example is given by the link between nodes 3 and 5 of Layer-U. This routes over nodes (3,4,5) in Layer-L. From

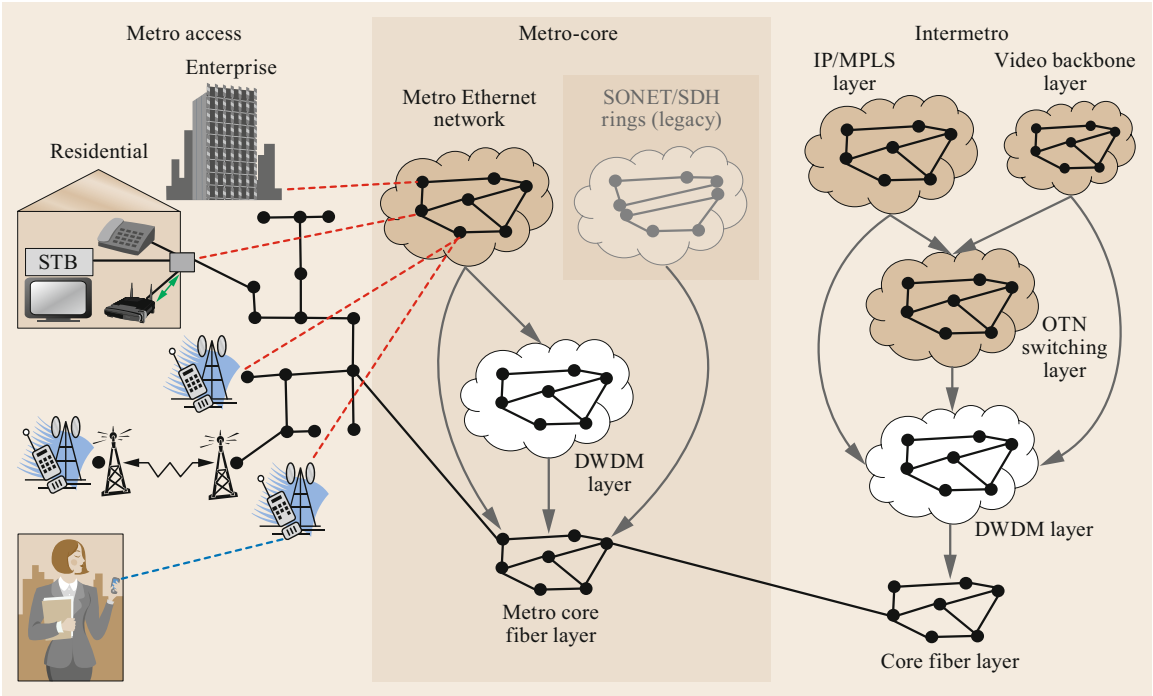


Fig. 11.3 Network layers and segmentation: more details

the perspective of network resiliency, we identify some interesting characteristics in Fig. 11.2 that are also common to real networks:

1. There are more nodes in Layer-L than Layer-U.
2. When viewed as separate abstract graphs, the degree of logical connectivity in Layer-L is less than Layer-U. For example, there are at most three edge-diverse paths between nodes 1 and 6 in Layer-U. However, there are only at most two edge-diverse paths between that same pair of nodes in Layer-L.
3. When we project the links of Layer-U onto their paths in Layer-L, we see some overlap. For example, the paths of the two logical links highlighted in Layer U overlap on links (3,4) and (4,5) in Layer-L. This example highlights the concept of a *shared risk link group* (*SRLG*). For example, if link 3–4 of Layer-L fails, then both links 1–6 and 3–5 of Layer-U fail. More generally, failures of links of lower layer networks often cause multiple link-failures in higher layer networks. If one applies these concepts to the segments and multilayer networks that we will describe, it will help to demonstrate the complexity of providing network resiliency.

Figure 11.3 gives a more detailed (but still simplified) view of the context of network layers and segmentation for large carriers. This figure is derived from [11.3]. As one observes, each segment can be a complex combination of layers and technologies. Furthermore, as we will see, network resiliency is achieved differently segment-by-segment and layer-by-layer. That is why it is important to garner an understanding of this abstraction before launching into descriptions of resiliency methodologies. We will describe the various aspects of Fig. 11.3 in the next few sections.

11.1.2 The Metro–Access Segment

Figure 11.4 expands the metro-access portion of Fig. 11.3, wherein it shows five different example architectures. These are still simplified diagrams and many more complex variations and hybrids of these architectures can be found in practice. Most carriers further segment a metro-access segment network into a *local access* segment and a *feeder* or *distribution* segment. In most large carriers, fiber is used in the distribution segment but as mentioned previously, the type of media is mixed in the local access segment. Furthermore, we try to avoid some of the common terminology found in many published articles of *Fiber to the X*, where X is curb, street, home, premises etc., because we feel this unnecessarily stereotypes the architectures.

Residential Broadband Service – PON

The access architecture shown in Fig. 11.4a shows a residential broadband service for a *single dwelling unit* (*SDU*) connected by a *passive optical network*

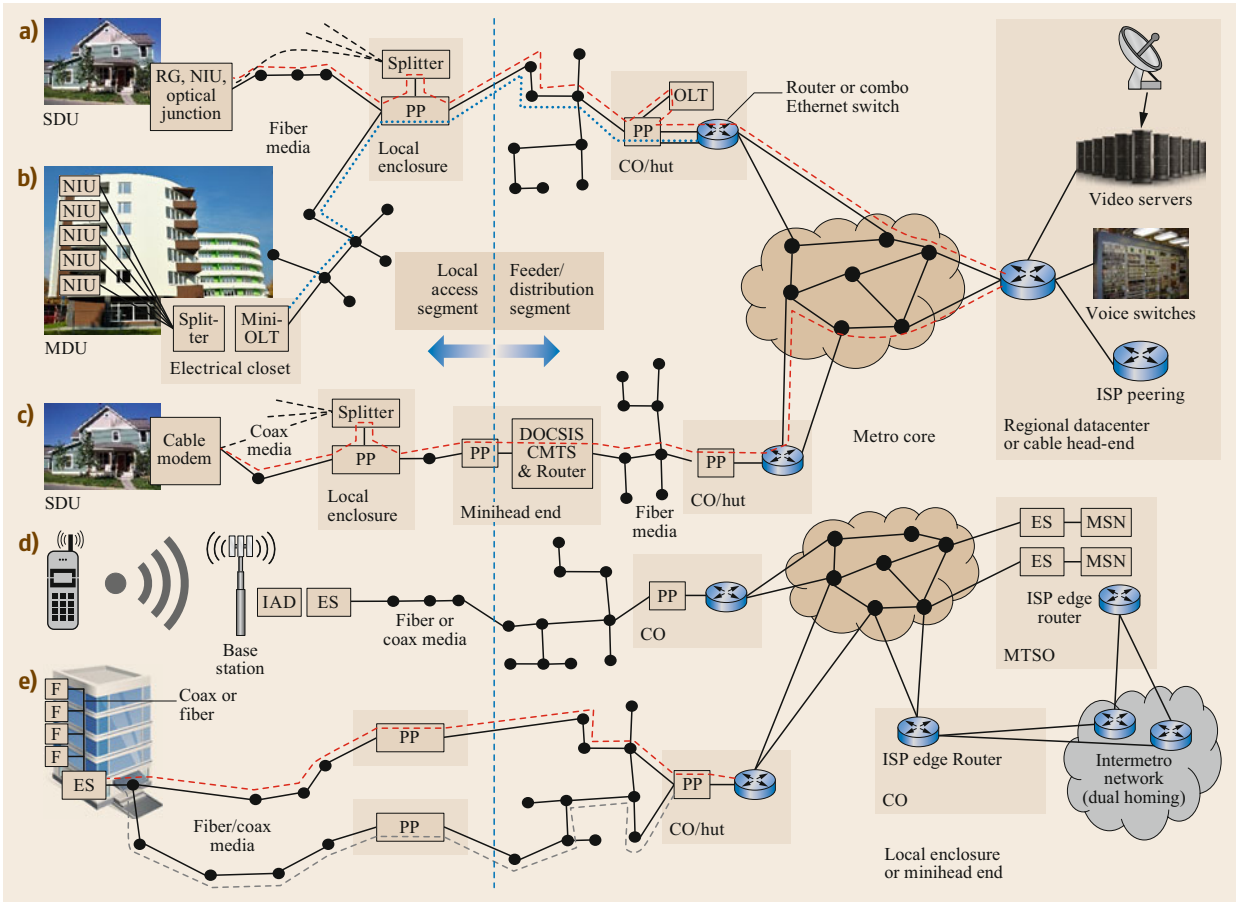


Fig. 11.4a–e The metro-access segment and connections to other segments (CO: Central office; MDU: Multiple dwelling unit; SDU: Single dwelling unit; RG: Residential gateway; OLT: (PON) optical line terminal; NIU: (PON) network interface unit; PP: Patch panel; ES: Ethernet switch; ISP: Internet service provider; MSN: Multiservice node; IAD: Intelligent access device; MTSO: Mobile telephone switching office). **(a)** Residential broadband service for an SDU via PON access network technology, **(b)** residential broadband service for an MDU via PON access network technology, **(c)** residential broadband service via hybrid coax/fiber access network technology, **(d)** residential broadband service via cellular wireless network technology, **(e)** enterprise Ethernet access service via fiber and/or coax access network technology

(*PON*) architecture, such as *GPON* [11.4]. There are different instantiations of the customer premises equipment for a PON architecture. A typical configuration on the customer premises is one or more set-top boxes connected to (functionally) a *residential gateway* and *network interface unit* (*NIU*). The NIU communicates directly with an *optical line terminal* (*OLT*). The inside wiring (sometimes called *inside plant* – *ISP* – however this term conflicts with other uses of the acronym *ISP*, so we avoid it) can be provided by a combination of coax cable, Ethernet cable, or Wi-Fi. This combination differs by individual residence and carrier policy. However, the connection from the outside of the house all the way to the OLT is provided over optical fiber. The outside fiber often interconnects with the inside wiring via a fiber junction box that is installed on the outside of the house. Often, this box in turn connects via coax or Ethernet cable to a small, combo *black box* that serves as a hub to the set-top boxes.

An outside-hardened form of thin fiber leaves the house and is routed (either aerially or buried) to a local cross-connect enclosure, often called a *local connection point* (*LCP*), where the GPON splitter is located. Some carriers have retrofitted their copper plant to support PON architectures, in which case this local enclosure is often colocated in a *digital subscriber line access multiplexer* (*DSLAM*)-type enclosure. Generally, for this architecture the local access portion of the plant is provided by a large fiber access cable, from which individual drops to houses from the access cable

are implemented by small cross-connect boxes, aerially mounted or buried. The local enclosure feeds multiple local access cables and forms the boundary of the local access and distribution segments of the metro-access network.

Note that our figures show, in honored carrier tradition, that every significant point of cross-connection in the network will use some form of a *patch panel* (*PP*). A patch panel is equipment that enables concentration of all physical interfaces of equipment and cable in a location (i.e., either fiber, copper, or coax connectors) to be centrally connected onto ports of the patch panel. Then, interconnection among equipment interfaces or interlocational transport cable can be accomplished by cross-connecting the ports of the patch panel, as needed for provisioning purposes. We note that most access patch panel ports are manually cross-connected. This is an important fact in studying network restoration since automation for rerouting connections in the network cannot be achieved at such choke points. By preconnecting transport fibers and ports of equipment into the patch panel when they are installed, the patch panel enables rapid reconfiguration of equipment and transport fibers without impacting the physical interfaces on the equipment or transport cables themselves. Furthermore, for provisioning and maintenance purposes, inventorying of equipment cable interconnection can be deduced by storing a *map* of the interconnection of the ports of the patch panels and work orders to add new or reconfigure equipment interconnection can be specified in terms of port IDs on the patch panels.

The GPON splitter offers usually a 16 : 1 or 32 : 1 *wavelength-division multiplexing* (*WDM*) splitting capability. Let us refer to the direction of transmission towards the customer as *downstream* and the reverse as *upstream*. The fiber on the upstream side of the splitter is then routed through the distribution portion of the network, sometimes even cross-connected through other local enclosures along the way, to the central office or hut where the fiber terminates on the OLT. One can think of today's typical OLT equipment as a PON device on the downstream side and an Ethernet switch on the upstream side. Thus, the middle of the OLT forms the conceptual boundary between the distribution network and metro-core network. Generally, the upstream side of the OLT equipment connects to a router or combo router/Ethernet switch that hangs off the metro-core network.

Notice in Fig. 11.4a or 11.4b that this upstream switch is *dual-connected* to other switches in the metro network. Therefore, this is, generally, the first encounter of network restoration capability in this metro-access segment. We discuss network availability later in an end-to-end context. Note that we display with a red dashed line the path of transmission between the customer premises and the *regional datacenter* (*RDC*), noting that this functional name varies per local carrier, such as *video hub office (VHO)*, etc. This is because a residential broadband service is almost always bundled with some form of entertainment or video content service; therefore, the purpose of the metro-access network in this example is to connect the customer to the servers and switches in the RDC, where the customer is provided with his/her various optional contracted services, such as broadcast video, video-on-demand, voice services, and Internet service. The customer can receive video content in the downstream direction (often via some form of multicast) or two-way phone service from circuit switches or two-way Internet service from the peering *Internet service provider(s)* (*ISPs*). Note that in many carriers this ISP is owned by the same holding company (corporation) as the access carrier.

Although optical fiber in the local access network is not highly deployed in general, when optical fiber is used exclusively to transport residential customer data and content, the architectures shown in Fig. 11.4a,b are typically the most pervasive form of access. At this stage we could go into a lengthy discussion of how the various PON fiber architectures have penetrated different parts of the world, but this would be beyond our present scope. We instead generalize that because of its high installation and maintenance cost, architecture Fig. 11.4a (single family unit) has had little penetration. In fact, as of the writing of this chapter many US carriers have trimmed back their planned investment in this architecture (for example [11.5]). In contrast, architecture Fig. 11.4b, the *multiple dwelling unit* (*MDU*) or, more colloquially, *apartment building*, has fared better because of its significantly lower cost per subscriber and ability to grow the subscriber base more rapidly. Furthermore, if an apartment building is prewired and has a standardized configuration, then addition of new subscribers can be very rapid and efficient. However, overall adoption of architecture Fig. 11.4b has not fared comparatively well in the USA compared to other countries, such as Korea and Japan [11.6], where a higher proportion of people live in apartment buildings and there are more nationalized approaches to promote the MDU PON fiber architecture.

Residential Hybrid COAX/FIBER Access

Hybrid coax/fiber access networks are the far more prevalent type for residential broadband services in the USA. This is mostly due to many decades of the amortized cost of cable coax installation and evolution. We illustrate this in the architecture shown in Fig. 11.4c via an oversimplified picture, where a cable modem is connected via inside wiring and then *drop cable*

to various sizes of coaxial cable in a local enclosure. Generally, the outgoing coax cable is of a different resistance/impedance and connects to the distribution network at a minihead-end. Note that there are many different types of coaxial architectures which we do not attempt to show here. However, today most adhere to the DOCSIS standards for transmitting content and encapsulating data packets [11.7]. Like the PON architectures, for residential broadband service, the purpose of the access and distribution networks is to connect the customer with the Regional Datacenter or to the *cable head-end* in the case of cable companies (also called *multiple system operators* (*MSOs*). (For brevity in Fig. 11.4, we show the head-end and RDC architectures sharing the same schematic but, in reality, they would be distinct locations across different types of carriers). Note in the figure that the minihead-end is placed at the boundary of the access/distribution network and that switching equipment connects via fiber to the metro-core network. At some near point the customer packets and signals encounter optical fiber and stay on a fiber network for the majority of their trip to their destination. We provide these detailed pictures of segments and their interconnections because the point where fiber begins and its type of architecture in that network segment will be a critical factor of why, where, and how network resiliency is provided. This is discussed in Sect. 11.2.4.

Wireless Access

The most widespread residence access architecture today is shown in Fig. 11.4d, which shows a cellular wireless network. The devices that connect over this network range over a wide range of devices, including phones, tablet computers, PCs, etc. There are also devices that offer broadband service by providing a modem in the customer premises with Wi-Fi or Ethernet connections to other devices. The modem then connects to a broadband carrier over the cellular network. This type of device is often called a *wireless hotspot* and today most data phones and tablets can mimic such a device, although the pricing for bandwidth on such devices varies depending on whether such a wireless broadband service is provided by the carrier.

The wireless modem/device then connects to the cellular network via a base station, with various vintages of cellular technology depending on location and wireless carrier. The base station is usually mounted on a cell tower antenna system, often shared with a plethora of base stations of different vintages and carriers. Typically, the base station consists of an antenna system mounted at the top of the tower and then cabled to an *intelligent access device* (*IAD*) located at the bottom of the tower. The IAD connects to an Ethernet switch, usually provided by a *metro Ethernet network* (*MEN*) carrier as some form of *network terminating equipment* (*NTE*). Figure 11.4c shows optical fiber or coax cable connected to the Ethernet switch. The Metro Ethernet carrier then provides a pair of unidirectional Ethernet private lines (often dubbed with various other Ethernet-related names, such as VLAN, VPN, E-line, etc.) between the base station NTE Ethernet switch and a paired NTE in the *mobile telephone switching office* (*MTSO*). Such an Ethernet private line from a base station to its MTSO location is typically called *cell-site backhaul* or *wireless backhaul*. These Ethernet services are provided as Enterprise Services by almost all incumbent local carriers and are usually provided/leased with various *quality of service* (*QoS*) and *class of service* (*CoS*) features, such as packet loss guarantees and data rate ranges, e.g., typically between 50-Mb/s and 1-Gb/s *committed information rate* (*CIR*). The feasible QoS features depend strongly on network resiliency and network availability. The IAD communicates with a *multiservice node* (*MSN*) device in the MTSO where it separates the voice and data services. Generally (in 4G, LTE architecture) it then forwards the data streams to a datacenter called the *evolved packet core* (*EPC*), where further service processing and routing occurs. Note that there are other methods and technologies to transport the Ethernet private line that we did not show, mostly line-of-sight (point-to-point) wireless microwave technologies. However, most of these wireless technologies are short distance and terminate on a central office or other distribution location where fiber is provided. For example, if the wireless carrier also has wireline serving central offices in that metro area, then they may transport the Ethernet private line from the cell site, which is not in the footprint of the wireless carrier, to a location that is in the footprint of the wireless carrier.

The wireless industry is also deploying variations to the wireless architecture described above, wherein the base station and antenna functionality described above is partitioned ever further. In this case (often designated within the evolving 5G technology) the base station equipment is centrally located rather than at each antenna location. In that case, the cell site locations become antenna-only locations and then their analog signals are backhauled (often called *fronthaul*) to the central base station via local access fiber. In such an architecture it is argued that the centralized base station can better load balance interference and channels among its many cell sites and thus significantly improve throughput over multiple cell sites compared to the (4G/LTE) architecture described above. However, we note that this requires a significant reconfiguration of the local access and cellular network plus richer deploy-

ment of fiber because of the very high speed required for fronthaul (usually 10 Gb/s or higher per antenna site). Various options of this idea include installation of more small cell sites that are centrally managed. This approach, while studied and promoted intensely requires significant capital investment by carriers. Thus, its potential deployment and extent is still under debate.

An interesting observation of architecture Fig. 11.4d is that the Ethernet backhaul can also be provided by architectures Fig. 11.4a,c, but in a business services (Enterprise) context. That is, some fiber and coax carriers offer enterprise versions of their residential broadband IP services using similar technology as shown in Architectures Fig. 11.4a,c. However, most large incumbent carriers who provide a fiber PON residential broadband service do not mix their enterprise Ethernet services (illustrated in Fig. 11.4e) with the transport of their residential PON services on the same fiber facilities downstream of the switch in the Metro CO/hut.

Another very important observation of architecture Fig. 11.4d is the network restoration ability in the local access segment of the network. For, while cellular technology is usually viewed as a poor second cousin to the previous architecture in terms of performance and bandwidth, it ironically has some advantages in terms of network resiliency. This will be discussed in Sect. 11.2.4.

Enterprise Access

Finally, architecture Fig. 11.4e shows a simplified picture of the enterprise Ethernet service. Such a service is most often of the *virtual private-line service* (*VPLS*) type, which connects Ethernet customers to their locations (or to different enterprise locations) in the same metro or in far metros. One of the advantages of a VPLS-type service is that the MEN carrier can make an enterprise customer's wide-area network appear to function as a simple LAN. Furthermore, the MEN carrier often provides an enterprise portal (basically a Web site) where the enterprise customer can configure many features of his/her wider area Ethernet network, such as firewalls, VLANs, encryption capabilities, committed information rates (CIRs), Ethernet private lines, private IP addressing, etc. Providing enhanced Ethernet services is one of the most hotly contested enterprise services among MEN carriers today.

A final comment about architecture Fig. 11.4e is that MSOs (cable companies) are also MEN carriers for enterprise services, hence the designation of either fiber or coax in the local access portion of the network in Fig. 11.4e. In fact, the MSOs are keen to compete with other incumbent carriers using enhanced versions of their DOCSIS cable plant. In particular, for the small to medium business (SMB) enterprise locations, they often experience an edge in customer premise self-provisioning of DOCSIS modems because most business location managers historically have coax cable preinstalled up risers, across floors/telecom closets, and into customer locations.

11.1.3 The Metro-Core Segment

Because of the evolution of technologies, regulatory climate, and economic/business environment, today's metro-core networks come in all vintages, types, and architectures. To that end, we have shown a prototypical (and very simplified) version in the middle of Fig. 11.3. In particular, we only show the metro Ethernet network (MEN) (introduced in the previous section), the (legacy) SONET, *dense wavelength-division multiplexing* (*DWDM*) layer, and fiber layers. We note that while vintages of layers whose nodes consist of legacy technologies, such as circuit switches, *digital cross-connect systems* (*DCSs*) and *frame relay* switches still exist, virtually all major carriers have plans to cap and retire them. For example, some of these are illustrated in Fig. 11.5, which is adapted from the presentation in [11.8]. While the present focus of most metro carriers is to transport packets over the metro Ethernet network layer (or a similar packet network), larger carriers also lease legacy *time-division multiplexing* (*TDM*) private lines to other carriers and large enterprises. The low-rate private-line market (DS3 and below) uses many of these legacy networks, shown in Fig. 11.5 but omitted from Fig. 11.3, as described above. In the future, high-rate private-line circuits will be provided as either a switched Ethernet *permanent virtual circuit* (*PVC*)-type service over the MEN layer or as a private-line high-rate circuit over the DWDM layer (but where the circuit usually encapsulates an Ethernet PVC circuit) using various forms of muxponders or *optical transport network* (*OTN*) switching cards on the DWDM equipment. (A *muxponder* is a combination of TDM multiplexer and optical transponder). Since the technologies are basically the same as in the intermetro segment, more detail on these technologies will be described in Sect. 11.1.4.

However, in Fig. 11.3 we still show the legacy SONET/SDH ring layer. We do this for the following three reasons: (1) large quantities of SONET rings are still common in most USA networks and in fact, are carrying many circuit switches and private-line circuits. Perhaps to the despair of some metro carrier executives, significant purchasing of SONET equipment continues; (2) SDH rings are still very common in Europe. This is mainly because many European countries are roughly the size of some large USA metros and the resiliency qualities of SDH rings have certain advantages; and (3)

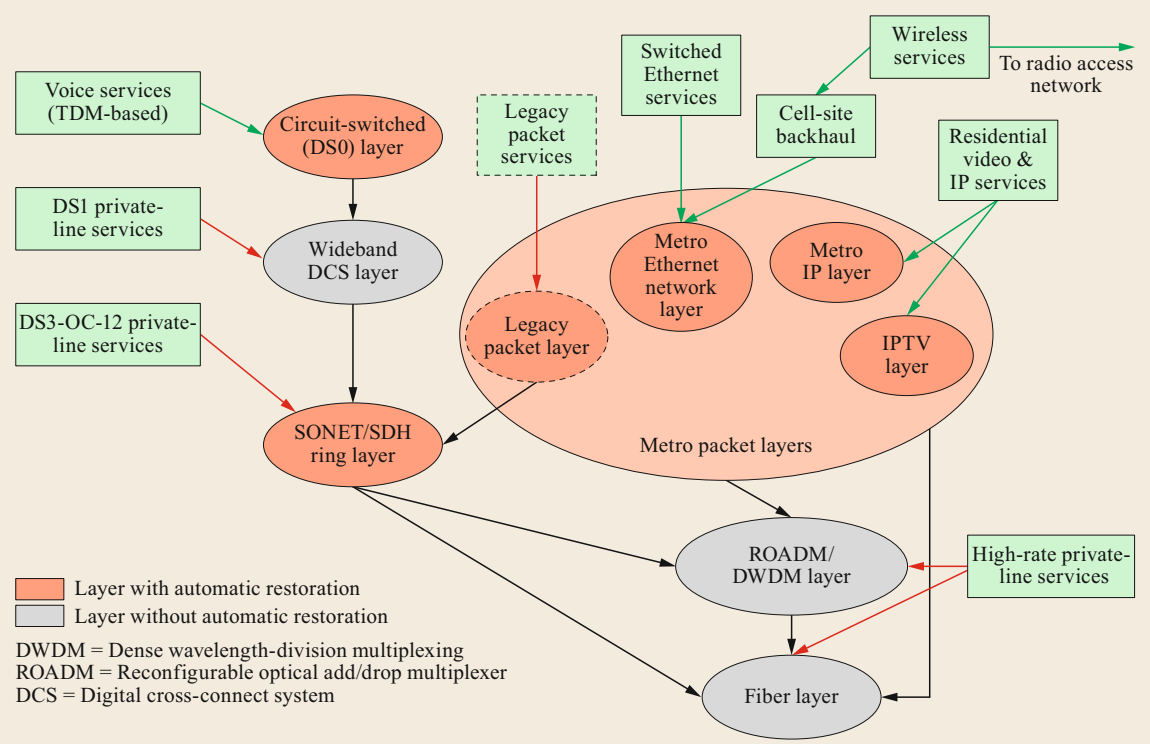


Fig. 11.5 Illustrating some legacy technology layers in the metro-core segment

as the optical layers of today's networks have evolved away from SONET/SDH, this has implied a significant strategy shift to where and how restoration capabilities exist. We will examine that shift in Sect. 11.3.

We make another important observation about legacy network layers. While most carriers make claims and issue press releases of advanced packet capability, old circuit switches are amazingly persistent in the incumbent carriers of the USA and Europe. In fact, one can state that the likelihood of a large proportion of one's voice call being carried over a circuit-switched network (64 Kb/s trunks) is high. Even calls made over cellular phones get routed at the MTSO onto circuit-switched networks. The reasons for this are mostly the legacy and prevalence of world-wide voice standards, signaling, local penetration, and rules for revenue-sharing among competing voice carriers.

Metro Ethernet Network (MEN) Layer

The metro Ethernet network (MEN) layer has become the workhorse of today's metro-core network. As noted above, while most voice calls and TDM private lines are carried over circuit-switched, DCS, and SONET networks, the vast growth of demand for bandwidth has arisen from residential broadband access and enterprise Ethernet services (such as Internet access, LAN services, Ethernet private line, VPLS, and cell-site backhaul). For example, the residential broadband services described in Sect. 11.1.2 all intersect the MEN from the metro-access segment and are carried as some form of enterprise Ethernet service, often by the same residential broadband services provider. The nodes in the MEN layer of Fig. 11.3 represent routers. This dichotomy in the use of IP routers and Ethernet switching comes from the fact that Ethernet interfaces have emerged as the economical interface of choice on virtually all switches. This was a significant driver for the evolution of the network and its resiliency methodology, as will be discussed in Sect. 11.3. In Fig. 11.4 as one goes downstream towards the edge of the metro-access network, one finds smaller Ethernet switches, i.e., with limited Layer-3 routing capabilities. Conversely, as one travels upstream towards the metro-core network, one encounters larger Ethernet switches with dual IP Layer-3 routing capability. In fact, most MEN networks are a hybrid, wherein they appear as Ethernet switches on the exterior (add/drop side of the switch), but behave as Layer-3 routers on the interior (network-side of the switch), mostly through various *Internet engineering task force* (*IETF*) VPLS approaches. The principal reason for this dichotomy is the carrier-grade resiliency and interoperability of Layer-3 switches. This feature

is also critical to our chapter goal and will be discussed later.

ROADM and Fiber Layers

While there are important differences among the technical characteristics of DWDM technologies between the metro-core and intermetro segments (e.g., amplifiers, coherency, need for regeneration, optical impairments, optical reach, etc.), from the standpoint of resiliency functionality, the DWDM and fiber layers of the metro-core segment are similar to that of the intermetro segment and as such will be discussed in more detail. However, there are differences in characteristics of outages and outage recovery between the segments and these also will be discussed later. Therefore, we conclude the section on the metro-core segment by noting that the links of the switches in the MEN today are mostly 10-Gb/s and 100-Gb/s Ethernet links. These links are routed as circuit demands over either direct fiber or the DWDM layer. The nodes of the DWDM layer are usually some version of *reconfigurable optical add-drop multiplexer* (*ROADM*).

As of this writing, the largest carriers are deploying (or planning to deploy) 400-Gb/s MEN links between the hub offices. Various flexi-grid DWDM spectrum approaches are being considered, if the client interfaces (in this case from the router links) can interoperate with the optical layer on the ROADMs [11.9]. Because of the diversity of traffic sizes at COs (or huts), it is still economical to route some packet links over direct fiber at the smaller demand nodes of the MEN layer. Thus, it is not uncommon for a MEN link to route over a mixture of direct fiber and ROADM links. This makes automatic and universal optical-layer restoration a challenge, as we will discuss in Sect. 11.3.3.

11.1.4 The Intermetro Segment

The intermetro-network segment is shown in the rightmost section of Fig. 11.3. However, note that Fig. 11.3 is a simplification that does not show the many legacy network layers that are still in existence today. For, similarly to the metro-core segment, there are various legacy layers, such as circuit-switched, DCS layer, SONET broadband switching, frame relay and other layers still in practice. Generally, these legacy layers mostly still carry TDM voice services and PVC or TDM/SONET/SDH private-line services. Figure 11.3 shows, most importantly, the *Internet protocol over multiprotocol label switching* (*IP/MPLS*) layer. It then shows the *video backbone layer* and the other (lower) network layers: OTN switching, DWDM, and fiber. We will describe the architectures of these layers. This will provide a basis for the resiliency discussion in the next major section. Before embarking we make the important comment that only a limited number of very large carriers in the world fully deploy all these network layers, let alone over a large national (or continental, such as Europe) footprint. Most ISPs deploy only a few (or none) of these other layers and, instead, lease various connections from other carriers who deploy lower layers. The reader should keep this in mind while we describe these layers. We will provide some more background on this limited aspect at the end of this section.

IP/MPLS Layer

As with the MEN of the metro-core segment, the IP/MPLS layer has become the modern workhorse for intermetro carriers. Thus, when Ethernet services need to connect across different metros, they can be transported over MPLS virtual links, often called MPLS *labeled switched paths* (*LSPs*) or *pseudowires*, for which their QoS and CoS features can be preserved. The use of MPLS as a protocol underlayer is critical, as it creates a Layer-3-free methodology that enables a mix of different types of traffic. This is an important feature to guarantee security, segregation, and the ability to provide different degrees of CoS/QoS to private enterprise traffic versus general Internet traffic. It also enables enterprise private LANs to use their own local, customized IP addresses that originate at their metro endpoint locations, yet not conflict with general Internet addressing in the intermetro network. Because IP routing and MPLS label switching are such integral parts of the resiliency of this layer, we will postpone more discussion about their functionality until Sect. 11.2.2.

Video Backbone Layer

Why do some carriers deploy a separate overlay network for video transport versus combining video traffic with other Internet and enterprise traffic in the IP/MPLS layer? The answer boils down to QoS and resiliency of the video service itself. Providers of residential broadband services (as shown in Fig. 11.4) must customize the design of their intermetro networks to achieve high network availability. This is because customer perceived quality of video service is very sensitive to outages, even short ones. And, most importantly, the intermetro network will distribute much of the same content to all customers (in multiple metros) and so disruptions in the intermetro network will impact many customers. It is generally considered in the industry that less than *five nines* (5 9s) availability will be viewed as poor performance by large quantities of video customers. Five nines availability means the service is up 0.99999 of the time, which implies no more than 5 min cumulative downtime per year. We delve into these issues more in later sections.

OTN Switching Layer and DWDM Layer

The OTN switching layer is shown below that of the IP/MPLS layer, meaning links of the IP/MPLS packet layer (router-to-router links) can be transported as connections (circuits) over the OTN layer. The OTN layer can also transport high-rate private-line traffic for those carriers who offer to lease circuits to other carriers and enterprise customers. However, many carriers prefer to route some or all their router links over the DWDM layer and hence in Fig. 11.4 we show an arrow from the IP/MPLS layer that routes directly onto the DWDM layer, thus skipping the OTN switching layer. As economics and technology have evolved, a common architecture (or *policy*) for intermetro carriers is to carry the smaller (less bandwidth) links or a particular *role* of router link on the OTN switching layer. Most design their intermetro IP/MPLS layer in a hierarchical method and consequently classify their routers and switches into various *roles*. For example, *edge* routers are homed onto *backbone* routers. The links between the edge and backbone routers are typical candidates to be carried by the OTN switching layer. We note that these internal role names differ widely across carriers, e.g., edge routers are sometimes called *access routers* (not to be confused with *metro-access segment*) and subcategories of backbone routers are called *aggregate* routers. See [11.2] for more detailed examples.

We provide some background on the OTN switch here, but more will be discussed in the section on network evolution. The OTN switch is an evolution of the *intelligent optical switch* (*IOS*). See [11.10] for a vendor announcement of an IOS product. The IOS is a SONET or SDH-based cross-connect system with automatic restoration capabilities. It was (and still is) widely used to carry TDM (pre-OTN) private-line traffic, where the basic switching unit is the SONET STS-1 or higher rate SDH STM-1 signal. As the OTN standards evolved, some of these same ideas and principals were used to develop the next-generation system. The OTN switch generally switches constituent channels of OTN signals at the ODU-1 rate (roughly 2.5 Gb/s) or, with standardized extensions, at the ODU-0 rate (1.25 Gb/s). The ODU-0 rate is especially adapted to transport and switch 1-Gb/s Ethernet private-line services, including those gigabit Ethernet links of a packet overlay network from the carrier itself.

We need to first make a technical distinction here between use of *OTN signals* and *OTN switching*. Virtually all DWDM equipment in the DWDM layer use the data protocols associated with the ITU OTN standards to channelize their high-speed digital signals. They also use the overhead data channels, alarm features, *forward error correction* (*FEC*), and other features associated with the OTN standard. Thus, one also can encapsulate, channelize, and transport a variety of different rate signals, such as 1-, 10-, 40-, or 100-Gb/s Ethernet, fiber channel, or even TDM signals over a very high rate wavelength (such as 100 Gb/s) by multiplexing these signals into appropriate ODU signals. If OTN switching is not deployed, the typical implementation of this capability in ROADMs is via muxponder cards on the ROADM. The muxponder ports form point-to-point, channelized signals that are demultiplexed at the far-end muxponder and vice versa in the opposite direction. Typically, to cross-connect a lower rate signal (say from ROADM-A to ROADM-B) the highest rate signal must be demultiplexed by a muxponder at A and then manually cross-connected from the add/drop port of muxponder A to the add/drop port of muxponder B via a fiber cross-connect device (e.g., fiber patch panel) and then remultiplexed by muxponder B. This demultiplexing function is common in a large intermetro optical network. For example, this function is needed if the carrier deploys ROADMs from different vendors and a circuit needs to cross different vendor ROADM network domains.

This is illustrated via an example in Fig. 11.6, adapted from [11.11], where a 10-Gb/s circuit is provisioned between ROADMs A–G. For example, this circuit might transport a higher layer link between two routers which generate the client signals at ROADMs A and G. The ROADMs in this example support 40-Gb/s channels/wavelengths. To provision our example 10-Gb/s circuit, we must first provision two 40-Gb/s *channelized* circuits. Furthermore, because of optical reach limitations in this example, the 40-Gb/s circuit must demultiplex at F and thus traverse two lightpaths in the second subnetwork. A *lightpath* is defined as a path of optically cross-connected DWDM channels, i.e., with no intermediate optical-electrical-optical (OEO) conversion. This need for signal regeneration requires interconnection between the ports of the two transponders at ROADM F. Once the two 40-Gb/s channelized circuits are brought into service, two 10-Gb/s circuits are provisioned (one is A–C and the other D–G). The end-to-end client signal is interconnected to the muxponders at A and G and the two subnetwork circuits are interconnected via the muxponder ports at C and D (assume ROADM C and D are in the same CO). Note that without OTN switching capability, this example uses a mixture of three different types of cross-connect technology: manual patch fibering (e.g., at node F), remote controlled optical cross-connect (e.g., at node B), and electrical time-division multiplexing (e.g., assigning the 10-Gb/s circuit to a channel of the channelized 40-Gb/s circuit at A).

In contrast, if the carrier deployed the OTN switching layer, then the high-speed links of the OTN layer

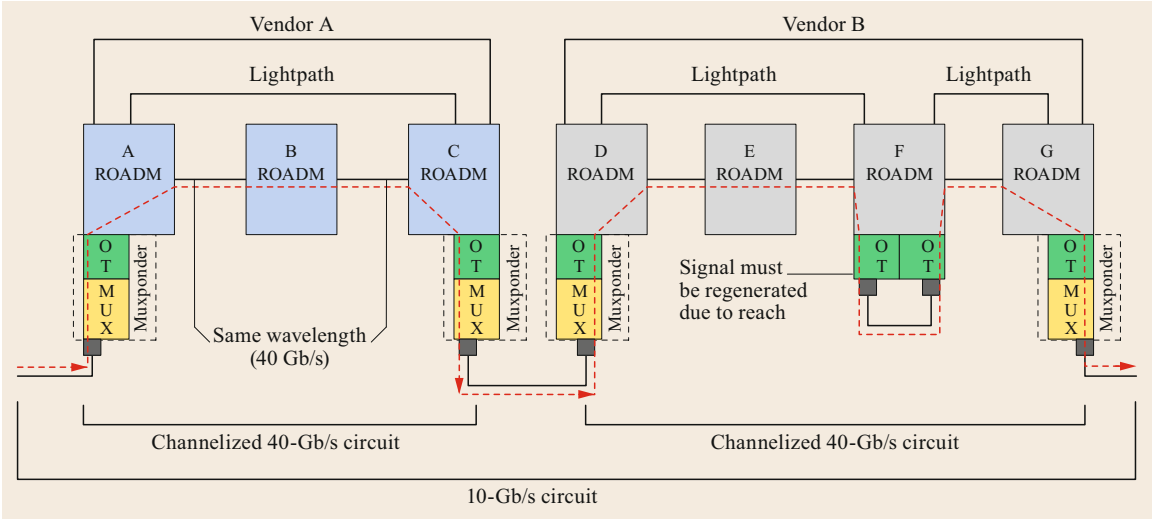


Fig. 11.6 Example path of 10-Gb/s connection (circuit) in the intermetro segment using ROADM muxponders (lightpath: path of optically (i.e., w/o electrical regeneration) cross-connected DWDM channels with same wavelength; OT: optical transponder; MUX: 10-Gb/s to 40-Gb/s multiplexer, 40-Gb/s circuit is *channelized* into 4×10-Gb/s channels; muxponder: integrated OT and MUX)

would be routed as high-speed connections (client signals) over the ROADM network. Figure 11.7 illustrates how this might look with an OTN switch at CO A, CO C/D, and CO G. In terms of the OTN network layer, this picture represents a topology of two ODU-3 (40 Gb/s) OTN links: one between OTN-A and OTN-C/D and another between OTN-C/D and OTN-G. The 10-Gb/s circuit in the example of Fig. 11.6 enters the OTN switch at A. In the direction of transmission left to right, it is encapsulated into the ODU-3 high-speed signal and then transports over the A–C lightpath. At ROADM-C, the ODU-3 signal add/drops into the OTN switch (via the OTs), and then the OTN switch automatically demultiplexes, cross-connects (via its switching fabric), and then remultiplexes the 10-Gb/s signal into the ODU-3 link between D and G (via two lightpaths).

The OTN switch has sophisticated automatic restoration features and so is very relevant to this chapter. We describe these in Sect. 11.2.2. One final observation for the intermetro segment is that in contrast to the metro-core segment, where links of the packet overlay networks (such as the MEN) often route over direct fiber, direct fiber routing is very uncommon today in the intermetro segment. Most large intermetro carriers deploy DWDM equipment in virtually all their COs. Thus, these carriers who provide the intermetro segment are well positioned to provide resiliency via some form of optical restoration method. However, we find (to the contrary) that this is not the case. This fact will be described and explained in the next few sections.

Leased Connectivity

As previously mentioned, only a few large ISPs will fully deploy all the lower network layers. In fact, very few (if any) ISPs own all the intercity fiber that they use and many lease all their fiber. Therefore, most ISPs lease some type of lower layer connectivity. The most common types of lower layer connection leasing are in order from top to bottom: Type-1) virtual private line (a virtual circuit); Type-2) lower layer private line, often called simply a *circuit*); and Type-3) *dark* fiber, often via a long-term lease called *indefeasible right of use* (*IRU*).

Although Type-1 and Type-2 look similar, they are quite different. For example, Type-1 is provided over the IP/MPLS layer of another carrier. Therefore, such a virtual circuit can indeed route through intermediate routers via MPLS label switching, which we describe in more detail later. In fact, a carrier with a more limited IP/MPLS-layer network might indeed lease Type-1 virtual circuits to connect its routers (or Ethernet switches) from the IP/MPLS layer of another (usually larger, more diverse) carrier. In contrast, the Type-2 circuit lease is generally not considered packet *virtual* and does not pass through any intermediate routers. That is, the circuit is totally provided via lower layer technologies, such as routed over a sequence of ROADMs, OTN switching layer, or direct fiber. While many intermetro carriers lease connectivity of Type-1 or Type-2, very few indeed extensively lease Type-3 dark fiber. An example of such a carrier in the USA is CenturyLink, who now expanded its leasing of intermetro-segment

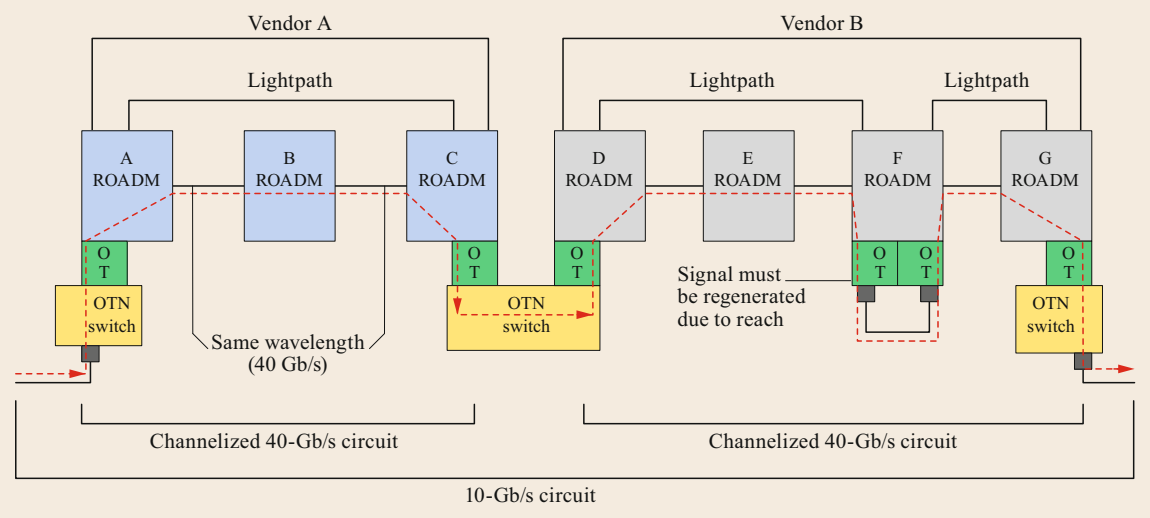


Fig. 11.7 Example path of 10-Gb/s connection (circuit) in the intermetro segment using OTN switch (lightpath: path of optically (i.e., w/o electrical regeneration) cross-connected DWDM channels with same wavelength; OT: optical transponder (of ROADM))

dark fiber through its acquisition of Level-3 Communications, Inc. [11.12]. Prior to the acquisition, many of the largest USA commercial ISPs, government entities, and *research and education networks* (*RENs*) leased intermetro-segment dark fiber from Level-3.

Why do most intermetro carriers who own fiber tend to not lease dark fiber (Type-3)? A generic (although not universal) answer is that they make substantially more revenue per unit of bandwidth by leasing the first two types of circuits, plus they believe that if they lease dark fiber that they might negatively impact or *cannibalize* their business of leasing connections of Type-1 and Type-2.

We list a few examples of types of carriers who deploy a subset of Fig. 11.3 (not exhaustive). For example, consider a carrier who:

- Does not deploy an OTN switching layer and instead relies on the muxponder capability of ROADMs and manual (or in a few rare cases automated) fiber patch panels as illustrated in Fig. 11.6.
- Deploys the OTN switch layer but uses it solely for grooming. That is, it uses the OTN switch to cross-connect circuits, such as shown in Fig. 11.7, but does not implement its advanced network restoration features.
- Does not offer residential broadband service and so does not deploy a video packet overlay layer.
- Offers residential broadband service but transports its video content over a shared IP/MPLS layer, i.e., does not deploy a separate video packet overlay layer.
- Offers residential broadband service and perhaps also intermetro private-line services but leases its intermetro video transport from another carrier. However, note that the leasing option can occur anywhere and everywhere – that is, every carrier might lease capacity or IP transport from another carrier in certain situations.

Enterprise Private Networks

Given our previous discussion of the various parts and combinations of layers that different carriers might deploy, an analogous discussion can apply to Enterprise private networks. These networks can be created by a company leasing various parts of layers in a similar fashion to a carrier who sells (leases) network capacity. However, ignoring those RENs and companies who lease a few scattered private-line circuits, the main application of such large private optical networks is to support interdatacenter communications. The difference in functionality is that some of the largest enterprise companies customize and narrow the functionality of their networks. For example, consider an Internet service enterprise (i.e., this could be a standalone company or a separate business unit of a large carrier) who deploys servers in datacenters that may offer search services, video/web content, cloud services, advertising, or social networking. Such an enterprise usually connects end-customers with their servers by employing a *content distribution network* (*CDN*) carrier. A CDN provider has multiple peering connections with many ISPs and metro-access-segment carriers that enable it to reach a large percentage of residential

customers. The traffic between datacenters of the Internet service enterprise is created by its own servers. This creates a unique advantage in that such an enterprise can mold and shape its traffic, CoS, and QoS in conjunction with the transport of its traffic among datacenters. Therefore, there are customized, forward-looking configurations and implementations that arise with such interdatacenter networks. But, as with the complex issue of resiliency in intradatacenter networks, we cannot cover all such private network configurations here. However, we will discuss a few specific examples in the last section on potential future extensions.

11.2 Network Resiliency in Today's Network

We now discuss where and how network resiliency is provided, but within the context of the architectural descriptions provided previously. As a precursor, we note that because of the disparate types of network segments, layers and architectures, plus the varied restoration techniques that we will describe, the reader will see that piecing together reliability standards is more art than science. But, first we give more clarification about network resiliency and how it is measured and evaluated.

11.2.1 Network Resiliency, Availability, QoS, CoS

By definition the engineered QoS of a service (often qualified by its specified CoS) must include an expectation of its network resiliency. Also, measuring network resiliency requires good traffic models, outage/failure detection methods, well-modeled restoration methods and both real-time and predictive outage-quantification methods. Thus, unfortunately, our characterization becomes somewhat theoretical or hypothetical and one might be surprised to find that most networks in practice are neither built nor maintained under such a strict analytical framework. There is no handbook of network practices that preaches *thou shall build your network thusly and thou shall be rewarded with desired network resiliency*.

Our experience indicates that in practice most networks are engineered *reactively* for resiliency: when network outages occur network administrators and operations personnel subsequently make network alterations, often resulting in rearchitecting or redesigning the network, its measurement processes and/or the operations practices. In other words, networks slowly evolve via experience, practice, and development of new technologies. The resulting network states balance the opposing forces of improving network detection and restoration techniques versus business economics and organizational realities. Therefore, we will proceed to analyze network resilience by first defining the types of outages that can occur and discussing what defenses have been provided in the networks previously presented. This might help the reader to recognize where network resiliency may fall short of desired levels, but where the business tradeoff is obvious. A high-level diagram of the sources of QoS impacts that can degrade expected network resiliency is shown in Fig. 11.8. At the highest level, we divide them into two major causes: (1) un-expected changes in traffic pattern and (2) network outages.

Potential Sources of QoS Degradations: Changes in Traffic Pattern

Unexpected changes in traffic can be caused by a large variety of causes and will vary with the type of traffic, the type of carrier, and depend on type of services, layer and network segment. For example, an intermetro ISP often provides (sells) packet transport or customer access to CDN providers. For example, CDNs are used to provide broadband video content to residential broadband customers of ISPs. CDN providers often have multiple points where they interface with such an ISP (called *peering points*). It is a practice of CDN suppliers to load-balance transport of their content across their peering points for different reasons, such as data storage backup, reconfiguring their own network, load balancing their servers across different datacenters, lowering network latency, rebalancing the load on their links at peering points, reacting to outages in their own network or servers, or balancing spikes in network traffic (just to name a few). While most of these changes to the traffic are instigated by the CDN, they appear to the ISP as randomly generated changes in traffic patterns. In other words, they are not human-instigated from the point of view of the ISP operations team. While these types of changes in traffic pattern and intensity are not easily modeled and predicted, in most cases they can be accommodated by adding additional network capacity and/or adding more dynamic routing capability at the appropriate layer (mostly packet layers). However, economic/business realities mentioned previously can generally constrain the methods of reaction by the ISP.

Thus, in most cases a large change in traffic pattern or intensity causes some loss of QoS and then remediation, both short and long term, occur. In packet layers, with newer *interior gateway protocol* (*IGP*) traffic

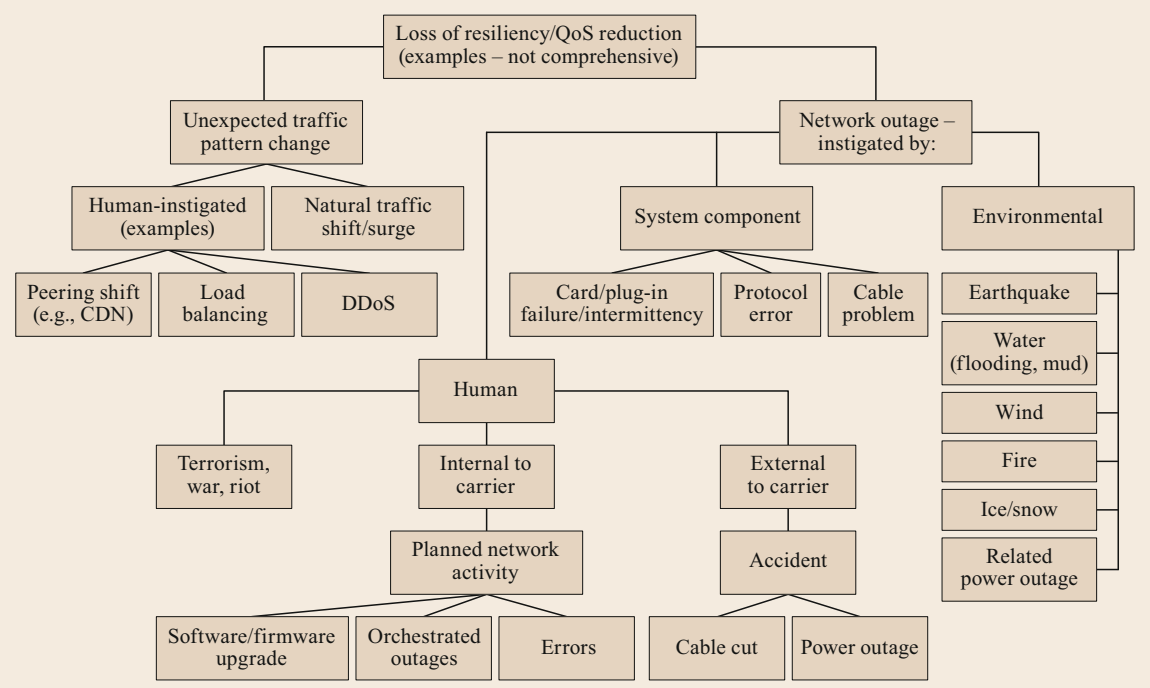


Fig. 11.8 Example graph of potential source of QoS degradations

routing methods, link congestion and capacity is considered in routing techniques, such as *resource reservation protocol – traffic engineering* (*RSVP-TE*) [11.13], *open shortest path first – traffic engineering* (*OSPF-TE*) [11.14], or *intermediate system to intermediate system – traffic engineering* (*IS-IS-TE*) [11.15]. With these methods, as traffic patterns change, so do the assigned virtual tunnels over which the traffic is routed. Furthermore, as newer *software-defined network* (*SDN*) capabilities emerge, intelligent network controllers or orchestrators can automatically react to changing traffic by creating more fundamental changes in underlying capacity at lower layers. We discuss examples in Sect. 11.3.6.

In virtually all carriers, the last resort for short-term traffic changes (or if no such auto-routing techniques are installed) is for the *network operation center* (*NOC*) to instigate changes in traffic routing, such as reweighting network links in packet routing methods or even shutting down various links or sources of IP traffic to prevent downstream congestion. This latter technique is often used to combat a particular human-instigated type of change in traffic pattern (but not caused by the carrier itself), namely, cyber-network attacks, such as *distributed denial of service* (*DDoS*). In such an attack, one does not want to accommodate the routing of more traffic but, instead, cut-off the source(s) without affecting legitimate traffic in the process.

Longer-term remediation generally includes performing a network planning/design process to anticipate the changes in the traffic matrix over the next planning period (usually 3–6 months) and then planning network capacity augmentation *jobs* (often called *capacity provisioning*) to add or reconfigure link capacity. The process of adding/reconfiguring network capacity often has a long lead time because of the need to install and configure equipment in COs/huts (where, because of limited installation crews, jobs for each CO/hut often must be combined and scheduled) and outside cable construction or rewiring jobs, often called *outside plant* jobs. There are some proposed techniques to speed up the capacity provisioning process, such as *as-needed* or *on-time* capacity provisioning, but these ideas usually require some degree of preinstallation of equipment, cards and prewiring (e.g., through a patch panel), including advanced technologies of equipment that integrate multiple layers in a given *box* [11.16]. However, despite this description, it is always a challenge in the network planning and design process to differentiate between the short-term changes in traffic pattern versus longer term (more permanent) changes.

Network Outages

Following along the right subtree of Fig. 11.8, another culprit that can reduce expected QoS below the ideal is network outages. Unfortunately, while most published

papers tend to define a somewhat simplified list of sources of outages (usually overabstracting to a *node or link failure*), the list of the real sources of such outages is large. The list can be divided into three basic types, as shown in Fig. 11.8: (1) human-instigated outages, (2) system-instigated outages, and (3) environmentally instigated outages.

Human-Instigated Outages

Human-instigated outages are very common and perhaps the most common, depending on network segment. Following down the subtree of human-instigated outages, we see three subtrees: (1) outages caused by terrorism, war, or civil unrest; (2) outages instigated by the carrier itself; and (3) outages caused by those external to the carrier (excluding outages of the first type).

Carrier-Instigated Outages and Outages due to Planned Network Activity

Internal, carrier-instigated outages cover a wide variety of causes. However, we just concentrate on one type, for which a lone subtree is labeled *planned network activity* in Fig. 11.8, which is related to planned maintenance, network configuration, and operations activity. In any carrier there is significant daily network activity. Every day network equipment somewhere is being tested, engineered, installed, and reconfigured and interlocation media (fiber, copper, or coax cables) somewhere is being installed, repaired, or modified. Usually the operations organization (or NOC) produces a daily log of planned network activity, sometimes called *planned cable intrusions* or analogous terminology. If possible, the NOC tries to prepare the network for these daily events by rerouting traffic around a link, network node, or equipment port on nodes that might be affected. Of course, this requires the ability to reroute traffic, which we will soon see is not universal. Because of inability to avoid an outage caused by such *planned network activity*, we get the further subtrees: (1) software, firmware upgrade, i.e., that are not fully *hitless* to QoS; (2) orchestrated outages, for example where the NOC schedules the activity at *off-hours* and hard-moves enterprise customer connections to alternate equipment; and (3) unanticipated errors.

Let us briefly discuss unanticipated errors caused by planned network activity. These include mis-cross-connecting circuits and equipment interconnections and inadvertently taking down a part of the network. These latter outages are particularly problematic since the service impacts can be large in the metro-core or intermetro-core segments. Furthermore, they are hard to detect, since a network element may not have failed and therefore the network detection notifications may be confusing or missing. Examples of this involve the fiber patch panel in COs/huts. Generally, installers prewire the ports of switches and DWDM equipment to the patch panel. Then, links (especially inter-router links) are installed by cross-connecting the correct ports on the patch panel. For example, in Fig. 11.6, if the ports of the muxponders at ROADMs C and D are mis-cross-connected, it will be difficult and lengthy to sort out what happened. Another impact of mis-cross-connection is that a link with live traffic may be inadvertently disconnected. For example, consider all the optical *directions* (i.e., *links* in the DWDM layer) associated with the ROADM-A in Fig. 11.6. These DWDM multiplexed links are enabled by connecting patch fibers from the ROADM MUX and DEMUX cards in each direction to the fiber patch panel and then connecting those patch-panel ports to patch-panel ports that connect the fibers from the cable vault of the CO. Figure 11.6 shows one direction/link between ROADM-A and ROADM-B, but many other directions (links) can emanate from ROADM-A to other locations (not shown). The ports associated with all the directions for ROADM-A will likely be in proximity to one another on the patch panel. Thus, if operations or provisioning personnel happen to inadvertently interfere with the patch cord to one of these working ports, an entire ROADM direction may experience an outage. For example, the entire link between ROADM-A and ROADM-B may go down. This could potentially affect many 100s of gigabits worth (or even terabits worth) of traffic in metro-core or intermetro networks.

Environmentally Instigated Outages

Looking now at the subtree labeled *environmental* outages, we note that while the large events generally have low probability, they have the worst impact. One reason is that while we illustrate the metro-access, metro-core, and intermetro segments in a separated fashion, the infrastructure media containing intermetro cables always overlap those of metro-core cables (e.g., conduit and aerial runs). Thus, such events also can affect large areas of metro-access networks in addition to the metro-core locations. There are many examples of large environmental outages, some of which are shown in the subtree of Fig. 11.8. We note that we formally categorize fire as an environmental outage, even though sometimes fires are human instigated.

Flooding is by far the most pervasive of these environmental events. Examples include hurricanes in the Southeast portion of the USA and monsoon rains in Brazil. Mud is a secondary effect of flooding from rain. Mud slides and floods are quite common in many parts of the world. Earthquakes are not as common as flooding, but the impacts can be much greater, wherein a huge number of simultaneous CO and metro-access

locations are taken out of service. One of the secondary (and the most common) service-affecting impacts of all these events is loss of power. Many COs have backup power generators, but as one travels downstream from the intermetro network, through the metro-core network and into the metro-access network, one finds gradually less sites provisioned with backup power. Generally, as one might anticipate, there are few (if any) cost-effective methods to maintain network resiliency against the larger of these events. Fortunately, such environmental events are less frequent than the other sources in the tree. We discuss more in the section on how resiliency is provided. We also refer the reader to a sequence of papers on this topic in IEEE Communications Magazine [11.17, 18].

System Component Outages

Finally, we discuss the subtree of outages labeled *system component*. The outages down this subtree represent failures of network components of the equipment represented by the nodes of each of the layers, such as shelf controllers, interface cards, plug-ins (e.g., optical small form-factor pluggable (SFP), 10 gigabit SFP (XFP)), *wavelength selective switch* (*WSS*), splitters, multiplexers, etc. These outages represent inherent failures of these components themselves, rather than outages not contained in this subtree that cause them to malfunction, e.g., cable cuts, power outages, etc. Equipment suppliers often express their estimation of the frequency of such outages per each major component, such as the *failures in time* (*FIT*) rate, which is defined to be the expected number of failures in one billion hours of operation (the reciprocal times one billion is the meantime hours between such failures). The meantime to repair such component failures is determined by the maintenance policy of each carrier and can also vary greatly across network segment and layer.

Most publications only examine this subtree plus the (human-caused subtree of) fiber cuts. They perhaps reason that all the other sources of outages ultimately result in outages of system components anyway, so why not lump all the outage-tree events together and estimate a convolution under this one system-component subtree. So, why do we show all these other sources of outage and QoS disruption in Fig. 11.8? The answer lies in the more formalized state space of potential outages and the associated probabilities associated with each state, abstractly represented by a distribution with a mean time between outages and outage repair time. That is, outages outside of the subtree labeled *system component* in the tree of Fig. 11.8 can cause multiple outages of the components identified within this subtree. For example, a particular potential flood, cable cut, or an earthquake are all outages that target a set of links and nodes in specific network segments and layers of Fig. 11.3. However, the probabilistic characteristics of the members of the system component outage event set (basically a type of *shared risk group*) are basically defined by the underlying outage, rather than these individual system components themselves. It would be impossible to properly model an outage event which affects a large set of links and nodes of the network layers if we did not know the basic cause of outage. While we admit that modeling all the potential events in Fig. 11.8 is intractable and, thus, some approximation is needed, having a knowledge of these relationships and how to model an outage state space properly can give insight into where and how to approximate the state space while bounding probabilistic estimation error.

Performance and Availability Assessment

Given our discussion above about modeling the (abstract) outage state space, a *rigorous* performance and availability assessment of the network that estimates its ability for a service to achieve its target QoS (as a function of each CoS) is complex: the network has to be designed, engineered, and operated so that this QoS is maintained under the bombardment of *potential* traffic changes and outages summarized in Fig. 11.8. Thus, in a theoretical sense, to engineer such a network one must have a traffic model, the probabilities of outages, including their distributions, expected interarrival and repair times, the corresponding restoration or rerouting mechanisms, and metrics that reflect the resulting QoS.

Because of the overwhelming complexity of the previous statements, designing a network to meet such a service QoS is more of an art than a science and, hence, this is the reason for our earlier comments about why most carriers do this in a *reactive* rather than a *proactive* way (despite what they may state publicly or publish from their Research organizations). Lacking all this intractable modeling sophistication, carrier experiences over many years or decades provide them with the opportunity to collect statistics on some of these types of outages (not including the *once in a century* outages). The first such event causes a major disruption, but then the carrier can be ready for the next one. This readiness includes a business decision about establishing a target network downtime for the next event. That is, given the frequency and extent of a potential outage in Fig. 11.8, plus its impact on expected QoS, is it worth the cost (if even possible) to make the network sufficiently resilient – or should the carrier tolerate the expected drop in QoS? A good example of this type of decision is given by USA Hurricanes Sandy and Harvey [11.19].

If we are to formalize an assessment of performance and availability, we need to briefly discuss the metrics

for expressing QoS. Network availability is much easier to evaluate for private-line services, such as leasing a 10-Gb/s Ethernet circuit service between two locations. If the circuit maintains its *bit error rate* (*BER*) below a threshold (a rigidly defined metric), then it is *available*. If it goes above for a prescribed amount of time, it is *unavailable*. Such a metric provides a simple 0 or 1 binary state. However, virtually all such private-line circuits are leased to form a link of an overlay packet network, so the 0/1 state can be deceptive. Typical metrics of packet networks are packet throughput, rate of packet loss (as a function of time) and packet round trip time or packet delay. However, there is benefit to using the simple 0/1 unavailability/availability model to unify concepts of QoS across all the different layers we have discussed. An example of such a simplification is quantifying network congestion via a utilization ratio threshold on a packet link. If utilization is below the threshold for a give proportion of time, then QoS targets of packet loss and delay can be assumed to be met. The simplification is that if the traffic congestion on any link is *above* the threshold, then that link is unavailable. Translating this to packet services, if the path of a packet service (of a given CoS) routes over a link that exceeds this threshold, then the service between the source and destination switches is defined as unavailable. In addition, if any of the lower layer links over which a packet link routes is unavailable, then the service is also defined as unavailable. If any of the components that comprise the lower layer link are unavailable, then that lower layer link and hence the packet link and hence the service is unavailable and so on. Of course, we then also must introduce rerouting/restoration methods in these network layers to offset the unavailability of that service. We will discuss these methods in the next section.

Given our informal discussion of the 0/1 availability model above, how then do we express whether an entire network along with its transported services meets the collections of target QoS? A network-wide availability metric can be expressed across the entire network by assessing the traffic matrix of services, their respective classes of service (CoS), and their intensity (volume) over each point-to-point. Two ways to express this are *average availability* and *worst-case availability*. For example, if any point-to-point service is unavailable, then the worst-case performance of the overall network is *unavailable*. As an example of the importance of CoS, suppose we wish to express the network availability objective for a *best effort* service as three nines and a premium service as five nines. In this context, worst-case availability is useful for broadband carriers to provide service to their metro head-ends (or RDCs or VHOs) from their national head-end. Every carrier would prefer to achieve the magical *five nines* of availability to reach every head-end because individual customer availability will further drop as data crosses the metro-core and metro-access network. After we present network resiliency methods, we further discuss these concepts of availability for residential broadband and enterprise services in Sect. 11.2.4.

Finally, simply determining the resulting (theoretical) network availability of a given (fixed) network is itself an intractable problem. If fact, this is an extremely complex network problem to describe, let alone solve, especially in multilayered networks. However, a simplified description can be found in [11.1, Sect. 2.2] and a more detailed multilayered example can be found in [11.20]. The significantly harder problem is how to cost-optimize the design/sizing/layering/routing of a network to achieve a target QoS. Perhaps now the reader may get a feel for why most network managers perform *reactive* architecture, design, and network management rather than proactive.

We next discuss the main resiliency capabilities of the architectures we described earlier. We again examine Fig. 11.3 and reveal the purpose of the colors in the figure. A network layer that is colored signifies where automatic restoration can be provided today. However, this is qualified by whether a carrier deploys that layer and enables its restoration features. We discuss this network segment by network segment, but in reverse order of our earlier descriptions.

11.2.2 Resiliency in the Intermetro-Core Segment

Figure 11.3 shows restoration capability in three layers in today's core intermetro segment: the OTN switching layer, the video backbone layer, and the IP/MPLS layer. However, as described previously only a few large ISPs will deploy all these layers in a comprehensive way. A key observation is that in all these cases no significant, automatic restoration capability occurs at the DWDM layer. This might seem surprising to some readers given the title of this book and chapter. For this reason, we provide the later section that discusses the evolution of resiliency and these architectures.

Restoration Methods in the IP/MPLS Layer

In general, the most common restoration capabilities in this layer are IGP reconvergence and *MPLS fast reroute* (*FRR*) [11.21]. We mentioned earlier some of the more capacity-sensitive IGP routing mechanisms, such as OSPF-TE, ISIS-TE, and RSVP-TE, within the context of responding to changes in traffic, rather than network outages specifically. However, these same IGP mechanisms can be used to respond to outages. In the

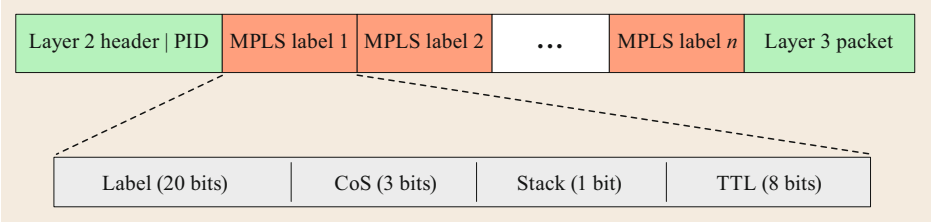


Fig. 11.9 Typical protocol data unit for MPLS (after [11.2])

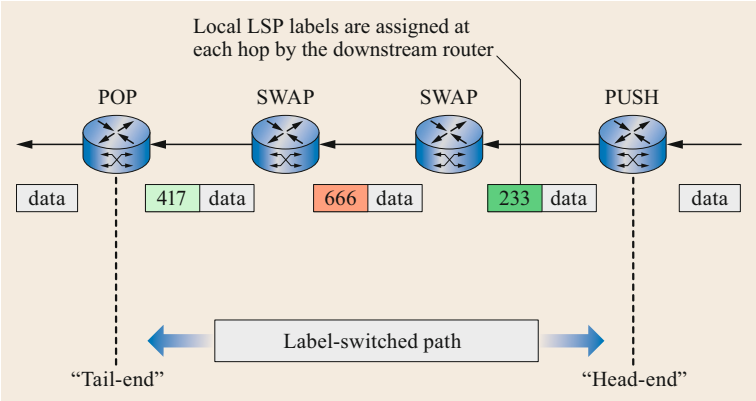


Fig. 11.10 Example of MPLS label-switched path (LSP) (after [11.2])

case of a network outage, wherein a packet link goes down or experiences reduced throughput, all these IGP protocols reflect this change via broadcast of *link state advertisements* (*LSAs*) and then recompute their paths between routers based on remaining paths and capacities. However, IGP reconvergence is slow compared to legacy SONET rings, which were originally engineered to restore within 50 to 100 ms, with the longer times for more complex outages, such as node failures (to be described later). This was one of the motivations to develop MPLS fast reroute. We also make a technical note that the most common way the IGP topology view sees a link capacity reduction (rather than the entire link going down) is with link bundling techniques. Each router vendor has its own method of bundling router-to-router links into a larger bandwidth bundle. The IGP protocols then only see the bundle and not each individual link. Also, the Ethernet protocols allow multiple links to be bundled into a larger bundle called a *link aggregation group* (*LAG*) and all switches allow an IP address to be associated with the LAG bundle instead of each individual link member. Each of these types of link bundles has its own rule about how packets are load balanced among its members and how individual failing links are handled and reported. We refer the reader to [11.2] and [11.22] for more detail about link bundling in packet networks.

Figure 11.9 shows the most common method of how MPLS encapsulates IP packets in an MPLS header (Figs. 11.9 and 11.10 are adapted from [11.2]). It consists of one or more MPLS labels, known as a label stack. The first 20 bits are the actual numerical label. There are three bits for inband signaling of class of service, followed by an end-of-stack bit and a time-to-live field, which serves the same function as an IP packet time-to-live field (i.e., to prevent infinite cycling of packets and to enable traceroute-type commands). Figure 11.10 shows how MPLS is used to define a type of virtual link in MPLS networks, called a *label-switched path* (*LSP*). To create an LSP, a path of links is precalculated by the routers, for example using a protocol such as RSVP, or via a centralized controller using more advanced SDN operations (discussed later). At the origin of that path (the *head*) a label is pushed onto the queue shown in Fig. 11.9. Note that in the world of packet routing, all connections are uni-directional. The label is generally chosen locally by the router and used to identify the outgoing port(s) of the desired link in the MPLS forwarding table (not shown). Intermediate routers along the path have their MPLS forwarding table similarly defined to switch the incoming packets with the given incoming label to the outgoing ports of the next link in the path, upon which a new local outgoing label is *swapped* with the incoming label on the MPLS stack of the packet. Finally, at the last router (the *tail*) of the LSP, the incoming label is removed (*popped*) off the queue and the packet *virtually* exits the LSP.

MPLS-based restoration methods are heavily used by intermetro- and metro-core-segment carriers to achieve network resiliency against many of the outages

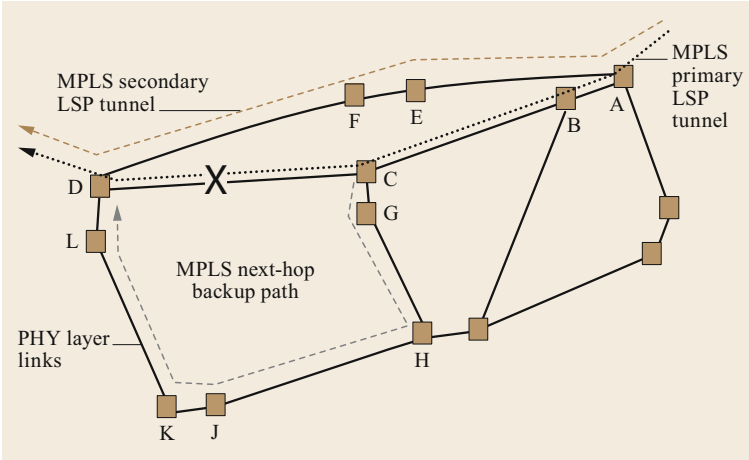


Fig. 11.11 MPLS fast reroute and IGP reconvergence (after [11.23])

identified in Fig. 11.8. Today, a hybrid combination of MPLS FRR and IGP reconvergence is commonly used to cover a wider variety of traffic changes and outages shown in Fig. 11.8. To illustrate, consider a primary LSP from router A to router D, as shown by the dotted path A–B–C–D in Fig. 11.11. This path is computed by a protocol, such as RSVP-TE under the *no–outage* network state, where there are no outstanding *link down* LSAs (that is, any previous ones have been cleared). Additionally, with the *next-hop* version of FRR, for each router link, a backup LSP has been precalculated and the forwarding tables of all the routers along the path of the LSP have been calculated and loaded with these label/port(s) switching rules. Note that this is a *link* in the IP/MPLS layer, not a link in the ROADM or fiber layer. Such a path is shown for the link C-to-D (C→D) in Fig. 11.11. If link C→D has an outage and a corresponding alarm is detected at the router at C or if router C has been notified by an exterior outage notification method, such as *bidirectional forwarding detection (BFD)* [11.24], then router C instigates the LSP FRR bypass mechanism that activates the push/swap/pop process discussed above. This process is instigated simply by pushing the predefined label of the bypass LSP onto the stack of any packet that is destined for link C→D. Since this is all done locally by router C, the bypass operation can execute very rapidly. If fact, most of the restoration time is outage *aging* and other IP timer mechanisms to prevent oscillation, false alarms, and other potentially nasty consequences.

The FRR next-hop link restoration mechanism can be fast, but it has downsides. A major downside is occurrence of a wider outage than just a single router link. In fact, an entire router could go down or be taken down for a maintenance event or other issue, i.e., the node labeled *software/firmware upgrade* in Fig. 11.8. Many of the outage events in Fig. 11.8 would produce multiple link outages in the IP/MPLS layer and clearly many of them are not statistical *rare events*. For example, if intermediate router C goes down in Fig. 11.11, then all the links coincident on it go down. The FRR link bypass method will not work and all packets trying to get from A to D will be lost. We note that there is also an MPLS FRR next-next-hop mechanism defined in the standards, wherein a path around router C can be predetermined. However, this is a potentially complex mechanism with its own additional downsides and we found through experience that it is generally not deployed.

A better method was found to cover the shortcomings of the MPLS FRR next-hop mechanism, namely by augmenting the FRR rerouting process with IGP reconvergence. The network resilience strategy behind this idea is as follows: A single link outage has a higher probability of occurrence than a multiple link outage. Therefore, a significant portion of probabilistic downtime (network unavailability) can be remediated with MPLS FRR next-hope link rerouting. For those lower probability (but higher impact) outages, such as downed routers and large-scale outage events, after a time-out period to complete network reconvergence, all the primary LSP paths are recomputed. By the virtual magic of packet/MPLS routing techniques, new paths can be implemented with virtually no packet loss by simply recalculating and reloading the MPLS forwarding tables and then pushing/swapping/popping labels of incoming packets. So, once the IGP reconvergence process has been completed and the MPLS forwarding tables reloaded, the entire network can converge to a new end-to-end routing scheme that compensates for the particular outage(s). In Fig. 11.11, the case of A→D traffic when the entire intermediate router C goes down is shown by the MPLS secondary LSP tunnel that now avoids router C. Of course, in this case traffic that originates or terminates at router C is lost, but this is

unavoidable for this high impact outage. Generally, the IGP reconvergence process takes from 1–60 s in large networks, so the packet loss would be in that range. This range is wide because each carrier sets its IP timers (for which there are many) to provide confidence in the process and avoid interference problems.

There is yet another downside of FRR next-hop link restoration, namely inefficient use of spare packet-link capacity. In fact, in large, highly connected networks, in many cases MPLS FRR will retrace its route through the same intermediate routers. To understand this in Fig. 11.11, consider traffic travelling from H→D along the path H–G–C–D. When link C–D fails, FRR will reroute it over path H–G–C–G–H–J–K–L–D. Furthermore, enough link capacity must exist on the links of the bypass path to accommodate the (now) extra packet traffic due to the bypass of link C→D. Clearly this bypass method is inefficient. Therefore, once the IGP reconvergence process is complete and capacity-sensitive routing methods are employed (as identified earlier), the revised end-to-end MPLS primary LSPs will be more capacity efficient. This can be combined with a CoS/QoS routing methodology so that the high-priority traffic will enjoy sufficient capacity on the bypass next-hop route, but the lower priority (e.g., *best effort*) traffic may find some congestion during the most severe outages. However, once IGP reconvergence is accomplished, the lower priority traffic will enjoy less congestion, as more efficient routes are established.

Clearly the above discussion leads us into one of the most important processes to economically provide QoS, namely network capacity design and planning. Abstractly stated, in this process one must engineer/optimize enough network capacity to maintain the objective QoS (as a function of CoS) under the plethora of traffic surges and network outages shown in Fig. 11.8, including how and when network restoration methods are applied at each potential layer. In fact, as previously noted, simply determining the resulting (theoretical) network availability of a given network capacity can be an intractable problem. This network design and routing problem in multilayered networks is too complex to be described here. However, a simplified description can be found in [11.1, Sect. 2.2].

Restoration Methods in the OTN Switching Layer

We briefly describe how mesh-network restoration is implemented in the OTN switch layer. Again, recall that not every carrier deploys the OTN switch and enables it for mesh-network restoration. A *Mesh-network* is one in which there are no architectural constraints about deploying rigidly structured topologies, such as hub-clusters, rings and chains. Let us first observe that since restoration techniques (such as we described previously) are universally deployed in the IP/MPLS layer today, there is essentially no ROADM-layer restoration process deployed. Thus, besides the legacy SONET/SDH rings, the OTN-layer restoration process is one of the few remaining (so-called) *optical-layer* restoration processes deployed. Figure 11.12 (adapted from [11.23]) gives a schematic of how the OTN switch can be deployed to handle network outages, i.e., respond to downed links of the OTN switch layer. Note the tight coupling between the network planning/design process and the restoration operation of the OTN switch

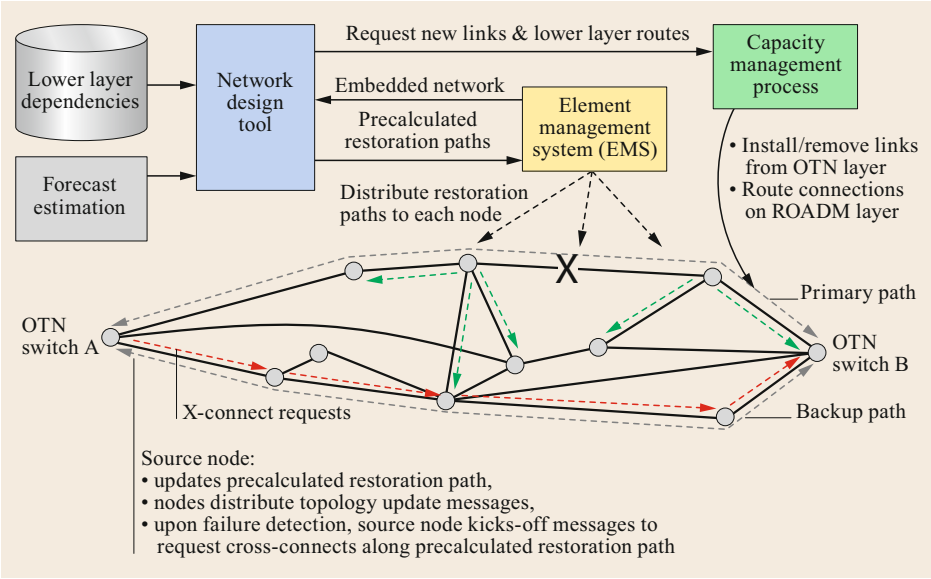


Fig. 11.12 OTN switching layer restoration process (after [11.23])

layer. This is highly motivated by our previous discussion on expected service QoS against potential network outages and how the network is designed and capacitated to meet that expectation. For the OTN layer, one can simplify by thinking of its connections as demands for private-line services which have a premium level CoS.

However, maintaining QoS gets very complex when the OTN switching layer also transports (as its demands) links of a higher packet layer (think back to the multilayer example of Fig. 11.2). The links of the packet layer that route over the same OTN link form an SRLG. Furthermore, the OTN link routes as a demand (circuit) over the DWDM layer and a DWDM link in turn routes over the links (spans) of the fiber layer. Thus, each of these links at lower layers forms an SRLG of packet-layer links. Describing such a complex network design process is beyond our scope. However, we will state that even the most complex network planning systems for this task (some even developed by the author) use narrowed assumptions about which parts and aspects of the outage tree given in Fig. 11.8 are modeled. Figure 11.12 shows such a network design tool. This tool models the parts of the QoS design process we have described previously to generate a future network design of the OTN Layer. We use the term *future* because there is usually a delay (lead time) in deploying links in this layer via a capacity management process, which in turn is usually part of a complicated capacity provisioning process that involves: equipment ordering, provisioning systems, network databases of record, issuance of work orders, management of installation, deploying testing crews, etc.

To provide network resiliency against link outages at the OTN switching layer when provisioning a network connection (circuit) over this layer, the network design tool calculates a primary path plus an end-to-end diverse backup path. That is, it calculates a secondary path whose links do not route over any of the lower layer links of the primary path. Stated alternatively, the SRLGs containing the links along the primary path are distinct from those of the backup path. If the carrier chooses to deploy a vendor *element management system* (*EMS*), the two paths are downloaded to the EMS, otherwise downloaded directly to the source OTN switch by the network design tool. The source node (OTN switch) of each provisioned connection (circuit) provisions the connection by sending cross-connect command messages in the OTN overhead data communications channels along the links of the primary route. This is a complex handshaking protocol whose description we omit here. The source node also stores the path of each of its provisioned connections along with its backup path. If a link outage occurs, link-down alarm information is flooded to all source nodes, as depicted by the blue dashed arrows in Fig. 11.12. Upon receipt of the link-down alarm at the source node (OTN-A), for those connections that route over the downed link, the source node begins a restoration process that reroutes those connections to their backup paths. The signaling to reprovision a connection is similar to the original path set-up, as shown by the red dashed arrows in Fig. 11.12, except a *bridge and roll* type of switch mechanism is often used to complete the connection with no disruption. For a descriptive example of a bridge-and-roll process see [11.25].

Since the backup path is diversely routed compared to the primary path, one of the simplifications of the network design process is that the probability of an outage of multiple, simultaneous lower layer links is low and, therefore, the probability that both the primary and backup path are simultaneously down is also low. Of course, glancing at the complexity of our network outage tree in Fig. 11.8, the probability of a multiple outage may be small, but not rare. For example, a common type of dual outage is when a link (or node) is down for planned network activity, such as capacity augmentation, rearrangements, software upgrade, etc. This process might take a long time and thus during this period a separate link outage might occur.

To accommodate this multiple link outage possibility, the OTN restoration methodology has a second phase rerouting step, where the source node dynamically computes new routes for those connections that remain down, i.e., cannot be restored after the initial rerouting process. There is typically a time-out between the first phase (using the precalculated backup path) and the second phase. The first-phase OTN switch rerouting process operates relatively quickly (but not universally as quickly as legacy SONET rings), usually in terms of at most a few seconds, depending on size of the network. The QoS motivation here is that this approach provides fast network restoration for a subset of network outages in the tree of Fig. 11.8 with highest probability of occurrence (or shortest mean time between outages), namely, single OTN card, ROADM link, or fiber span outages. For those outages beyond this subset, the restoration time will be longer or in some cases the connection will not be restored. However, even though circuits rerouted in the second phase take longer and the outage can affect larger groups of circuits, if the probabilities of the occurrence are sufficiently low, then the expected QoS objective can be achieved. As we mentioned earlier, since the metrics of such a complex tradeoff are hard to assess, this tradeoff is usually determined experientially by tracking statistics on circuit downtimes over time caused by actual network outages.

The timing and phasing of the OTN restoration process interaction can be fine-tuned by each carrier. Also, the carrier can choose to alternatively use the vendor-provided EMS or vendor-provided *network management system (NMS)* to provide the functionality of the network design tool. When a link is brought back into service (repaired) a *clear* alarm is distributed and the source OTN switch can move the connection back to its primary route, a process which is sometimes called *revertive*. However, to leave the rerouted connections on their secondary paths is generally not recommended else the network would be left in a un-designed state; if another outage occurs in such a state, the network might not meet its expected availability target. To address this problem and because the network changes over time from outages and changing traffic and network capacity, the OTN layer network also provides a *rebalancing* feature that reroutes already provisioned circuits to more efficient paths, enabled via a combination of network design tool, EMS, NMS.

11.2.3 Resiliency in the Metro–Core Segment

The metro core in Fig. 11.3 shows the metro Ethernet network (MEN), legacy SONET/SDH rings, DWDM and fiber layers. Although SONET/SDH is a legacy layer and most carriers are in the process of capping and/or retiring it, we provide a brief description to give perspective on its ability to restore as an optical-layer restoration method. However, we point out that this has been and will likely continue to be a prolonged retirement process in most legacy carriers. We will discuss both its evolution and de-evolution in the next section.

Figure 11.13 (from [11.1]) shows the two most popular architectures for the SONET/SDH ring. For simplicity, we describe the SONET version here, noting that the SDH version is very similar. See also [11.26]. The 2-fiber *unidirectional path-switched ring* (*UPSR*) (Fig. 11.13a) is configured over two unidirectional fiber rings, one transmitting in a counterclockwise direction over the nodes, the *add-drop multiplexers* (*ADMs*) of the ring (shown as the *outer ring* in Fig. 11.13), and the other transmitting in the clockwise direction (shown as the *inner ring* in Fig. 11.13). A SONET STS-*n* signal enters and leaves the ring via add/drop ports of the ADM at the two ends of the connection (let us call the nodes A-end and Z-end, respectively). For a ring in a nonfailed state, it transports the two uni-directional components of the SONET signal by A transmitting to Z and Z transmitting to A over the channels of the ring in the same counterclockwise direction on the outer ring. A STS-*n* connection transmits over *n* consecutive channels (or time slots), starting at a channel numbered $1 + kn$ (where k is an integer ≥ 0). It simultaneously sends duplicate signals from A to Z and Z to A in the clockwise direction on the inner ring (shown with blue dotted lines in Fig. 11.13). In a nonfailure state, a port selector switch at the receiver chooses the normal ser-

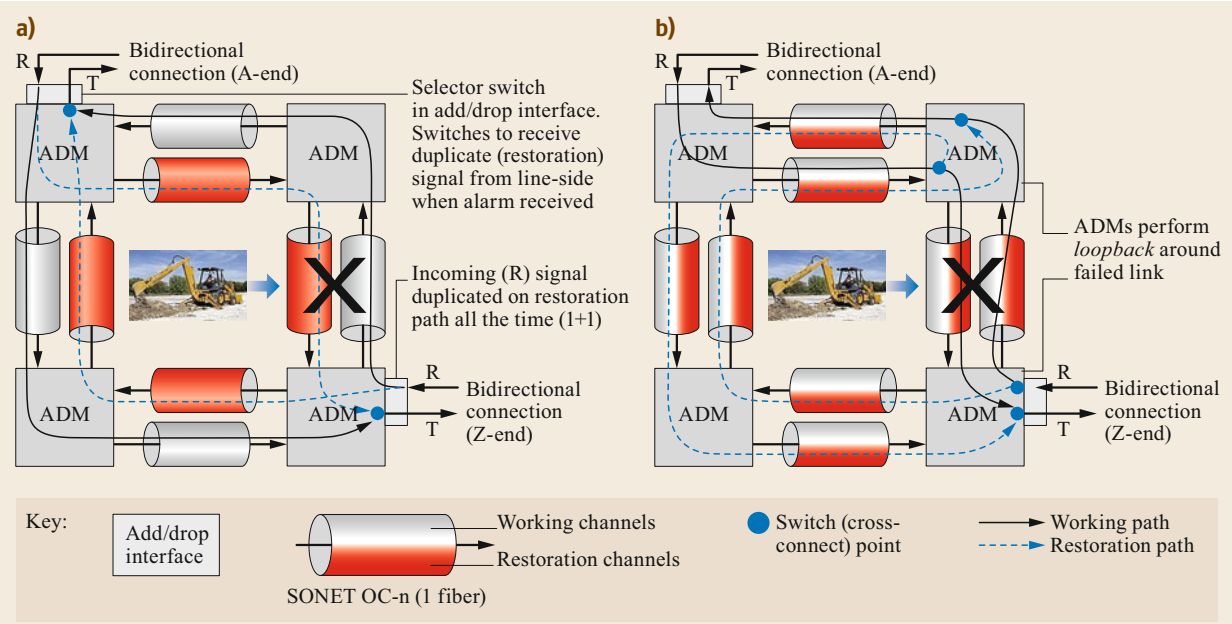


Fig. 11.13a,b SONET/SDH ring restoration (after [11.1]): **(a)** 2-fiber unidirectional path-switched ring (UPSR), **(b)** 2-fiber bidirectional line-switched ring (BLSR)

vice signal from one direction (counterclockwise). If this signal fails, i.e., one of the end ADMs receives an alarm, then the port selector switches to the alternate (duplicated) signal shown in blue. It reverts to the original signal after receipt of a clear signal or under other control messages.

In the 2-fiber *bidirectional line-switched ring* (*BLSR*) (Fig. 11.13b), the first half of the channels (time slots) of the high-speed optical signal of each link (i.e., ADM to ADM link) in each direction are used for transmission when the ring is in the nonfailed state. The second half of the channels is reserved for restoration shown as shaded in red. When a single link failure occurs, a loop back is executed at the add/drop port of the nodes surrounding the failed ring segment and the failed segment is patched with the restoration channels over the links in the opposite direction of the failed segment around the ring. In the case of node failures and multiple failures, complicated procedures using squelch tables and other mechanisms are standardized to prevent mis-cross-connection for connections whose ends are in the failed segment. Connections must be assigned to the same channel numbers on each link of the ring over which they route. A BLSR has the advantage over a USPR that the same channels (time slots) on different links can be assigned to different add/drop signals whose routes do not overlap links of the ring.

There was also defined a 4-fiber *BLSR* (not pictured), plus some other minor versions that were never deployed and we omit. In the 4-fiber version, one gets a full channel system on a pair of fibers plus a fully redundant system on another (backup) pair of fibers. This results in redundancy in the transponders in each direction. To picture the contrast, in the 2-fiber system of Fig. 11.13 there is a transponder in the (pictorially) left/west direction and then another transponder in the right/east direction. (However, note that when one looks at the physical equipment there is no sense of direction.) For example, the 2-fiber OC-192 system gets only 96 working (service) channels, whereas the 4-fiber system enjoys the full 196 working channels. Therefore, if one of the transponders fails (in one direction) on the 2-fiber system, then loop back must still occur for all channels. That is, it performs similarly as a fiber outage. In the 4-fiber version there is an extra set of (backup) transponders in each direction. Therefore, a transponder outage can be treated differently than a fiber outage. For example, in the case of a transponder outage, the system can execute a protection switch to the backup transponder and avoid a loop back around the other direction of the ring. This gives the entire ring higher availability since it can recover from dual failures where both a transponder outage and fiber cut occur. Other major advantages are that maintenance can be performed on the electronics on a single link without putting the ring into a loop back plus the restoration channels on the backup system can be used for other purposes. However, a huge disadvantage with the 4-fiber version is that it literally doubles the transponder costs compared to the 2-fiber version. As a result the 4-fiber BLSR was not widely deployed.

For those SONET subnetworks with more than four nodes (as shown in Fig. 11.13) and where the demand (signals) form a more a dense traffic matrix *mesh* (that is, the traffic/connections between most of the point-to-points is nonzero), the BLSR configuration typically uses optical capacity more efficiently than the UPSR. Since most connections (demands) in the feeder route of a metro-access network home on the same serving CO (such as in Fig. 11.4e), these demands do not form a dense traffic matrix network and thus the UPSR was mostly used to connect locations in the metro-access segment to the metro-core nodes, whereas the BLSRs were used to interconnect the COs of the metro core.

We note that most references generally label SONET/SDH rings, the legacy IOS switch, and the OTN switch with restoration enabled to be categorized as *optical layer restoration*. In reality this is a stretch of terminology since these restoration architectures make no fundamental changes to the optical components, e.g., transponders, lasers, amplifiers, wavelengths, arrayed waveguide grating (AWG) etc., but instead modify the payloads within the optical signals by modifying the signals after they convert to electrical digital signals at each node. While we described the OTN switch within the context of the intermetro segment, we note that some carriers also deploy a limited form of OTN switching, as well as its precursor legacy IOS switch, in the metro-core segment. However, it is much less used in the metro core compared to the intermetro core. Furthermore, if deployed in the metro core, then it tends to be used mostly for grooming with the restoration feature disabled. Therefore, we do not focus further on the OTN switching layer in this section.

That leaves the brunt of restoration capability to the MEN, which restores very similarly as described in Sect. 11.2.2 on the IP/MPLS layer. As noted earlier, the advantage of this architecture in most metro-core-segment carriers is that the MEN is a VPLS-type implementation, which is very amenable to enterprise customers to interface their Ethernet LAN environments.

11.2.4 Resiliency in the Metro-Access Segment

Let us examine the first three sample residential access architectures given in Fig. 11.4. Because of cost

and difficulty, in most of the world the metro-access-network portion has no restoration capability. That is, if an outage occurs along the path from the customer premises to the feeder network demarcation point, there is no installed methodology to reroute the connection. Many of the last mile infrastructure media in architectures Fig. 11.4a–c are aerial in vast parts of the United States.

Resiliency of Residential Wireline Access Architectures

This makes it even more susceptible to the potential outages on the environmental subtree of Fig. 11.8. Just one of these type events, such as a wind, ice, snow storm or a power outage coupled with the fact that it is unlikely such an outage will be repaired in less than 4 h (240 min), increases the yearly downtime almost halfway to three nines (500 min per year). This longer repair time is because in this part of the metro-access network a *truck roll* (maintenance personnel physically traveling to the site) is typically required to repair the outage and a large portion of this repair visit involves physically locating the source of the outage. In fact, a power outage from a downed or damaged utility line caused by wind, ice, or snow will usually far exceed four hours because the telecommunication cables usually ride on the lower part of poles and are often damaged. Repairs are complex and jobs have to be queued and are often dependent on power companies. Also, even if there is no damage to the network, power losses to the residences usually cause simultaneous outage of broadband service because the majority of residences do not have any form of power backup and residential broadband carriers almost universally do not provide battery backup for the residential gateways and set-top boxes for Internet services, although a few provide battery backup for wireline phone service. All these outages on the environmental and human/external-to-carrier part of the tree in Fig. 11.8 are in addition to the normal equipment outages that might occur to the PON and MSO equipment in the metro-access part of the network, such as failure of a PON plug-in on an OLT. However, many of these also require truck rolls to repair. Thus, this quick analysis shows that it is very unlikely that customers of these architectures will receive better than three nines of availability (and usually worse) over the long term.

We now illustrate why we left the metro-access segment to the end of the resiliency discussion. Consider a point-to-point packet flow, $P(A, Z)$, from a customer premises source in a metro-access network, A, to a destination in another (different) metro-access network, Z. For example, this could be a customer in metro access network A being served by a web site or server that is in metro-access network Z. Assume this flow is transported over the IP/MPLS layer of metro-core network $A1$ on the A-end and the IP/MPLS layer of metro-core network $Z1$ on the Z-end. Also assume it is transported between the IP/MPLS layers of metro $A1$ and metro $Z1$ via the IP/MPLS layer of intermetro-core network B. Because of aggregation of different IP flows, the multiplexed transport flow in network B could indeed contain hundreds, thousands, or even tens of thousands of individual point-to-point or multipoint connections of the same path. If we assume that the worst case availability of node-to-node (router-to-router) packet flows in networks $A1$, $Z1$, and B is at least five nines (that is, no more than five minutes of downtime per year for each of these segments) and then suppose that metro-access network A has at least three nines of availability (no more than 500 min of downtime per year), then, what can we conclude about the end-to-end availability of the flow $P(A, Z)$?

The answer is that when looking at availability of paths across multiple segments, one does no better than the weakest segment; because, assuming statistical independence, the end-to-end availability is the product of the three availabilities, which is less than the minimum of the three availabilities. Therefore, in this example the end-to-end availability of connection $P(A, Z)$ is no better than three nines and perhaps worse. Given our previous discussion about metro-access network availability for architectures Fig. 11.4a–c (that is, network availability for that segment is on the order of three nines or worse), why then provide such high availability in the metro-core and intermetro-core networks? Unfortunately, alluding back to our earlier comment about *more art than science*, we find that there is no simple, definitive answer.

However, one motivation is that carriers are more cautious about the potential magnitude of an outage, e.g., the number of customers affected times the length of outage. For, as one travels downstream, towards an end customer in Fig. 11.4, potential outages along the path generally affect fewer customers. Another motivation is that networks in the metro-core and intermetro segments provide transport to multiple types of customers. Thus, the residential customer in the access architecture shown in Fig. 11.4a may share the same metro-core or intermetro-core network as an enterprise customer in architecture Fig. 11.4e (Fig. 11.4e to be described in more detail later). Because of potential diverse fiber or coaxial routing in the metro-access network, the enterprise customer in Fig. 11.4e could enjoy much better network availability, hopefully approaching the *golden* standard of five nines. However, we admit that in practice such a stringent level of availability on an end-to-end basis is very hard to achieve. At

the minimum, to get close to this golden standard for the enterprise customer requires the very high availability of both the metro core and intermetro core that we described previously. Furthermore, a large carrier may lease transport across its metro-core and/or intermetro-core network to other access carriers, including its own enterprise business units, and so needs to provide universally high availability for those segment networks.

Resiliency of Enterprise Wireline Access Architectures

Let us now examine the architecture in Fig. 11.4e, which serves enterprise customers. We will discuss the architecture in Fig. 11.4d afterwards. The main difference between the architecture in Fig. 11.4e compared to residential architectures is that residential access in the USA and much of the developed world is dominated by coaxial cable. In contrast, this characteristic is less true in the world of enterprise services, where fiber is more prevalent, but not dominant because of the large quantity of small-to-medium businesses (SMBs), such as hotels, motels, small stores, etc. Our main observation is that the previous analysis of service availability for residential broadband Internet customers (architectures in Fig. 11.4a–c) is unacceptable for enterprise services, especially for larger enterprise services in large urban buildings.

Carriers handled this situation in the past with SONET/SDH path-switched rings, such as the stalwart UPSR. We discuss its downward evolution in the next section. The availability of the SONET architecture is much closer to the five nines availability that is typically characteristic of the intermetro-segment network. One reason is, whereas residential broadband service is often provided aerially, architecture Fig. 11.4e is often buried fiber or coax and it is quite common to provide an alternate, fiber-diverse path, shown by the light-colored dotted line. Thus, this dual-path access architecture is left vulnerable to switch or multiplexer outages at the building common space (usually in the basement or other designated closet area on the lower floors) and at the CO or hut location, but not to the myriad of other outages in Fig. 11.8, except for maintenance events, larger scope environmental outages, and other outages already discussed in the core metro and intermetro segments. Note that, technically speaking, carriers who lease Ethernet private lines for cell-site backhaul generally regard these services as enterprise services (not residential). However, we discuss cell-site backhaul in the discussion of the architecture in Fig. 11.4d instead of here in the context of architecture in Fig. 11.4e because the resulting service availability (network resilience) is engineered quite differently for end-to-end cellular services.

We mentioned path-switched SONET/SDN rings are in the (very long) process of being capped and retired so, then, what technology will be used to provide the dual-path switching in the metro-access network for enterprise-services? Many different architectures can be used. One example is that the carrier can define an Ethernet VPN and then define two packet layer-3 or layer-2 virtual links over which the VPN is routed. These links originate and terminate between the network terminating device at the enterprise common space (not shown in detail) and its network terminating device at the destination, e.g., in the first Metro CO or at another enterprise location on the far end. One virtual link routes over the primary path and the other link over the secondary path. Then, an end-to-end detection mechanism, such as BFD, can detect whether one of the links (paths) has failed. A simple mechanism, such as OSPF in the network terminating equipment, can be notified and perform network reconfiguration to use the secondary path, assuming it also remains up during the outage. In this case this simple overlay network consists of two nodes (i.e., Ethernet ports) and two links, so it operates very quickly.

Other Ethernet packet-like ring restoration methods have been developed, such as the ITU-T recommendation *Ethernet ring protection switching* (*ERPS*). For example, see [11.27] and [11.28]. Another example is IEEE 802.17 *resilient packet ring* (*RPR*) [11.29]. However, it is not clear which of these (if any) will see widespread adoption among access carriers across the world. In addition, their adoption will certainly not achieve the extent of SONET and SDH rings in the past.

Resiliency of Cellular Access Architectures

We now examine the architecture in Fig. 11.4d. The cellular network architecture and cell-site backhaul was discussed in Sect. 11.1.2. Although metro-access carriers may offer some form of restorable Ethernet PVCs under expensive enterprise service access options, few carriers provide fiber-diverse paths between a base station and its first metro CO. Also, in general the QoS of cellular service is worse than wireline service because of more susceptibility to weather events, traffic surges, limited spectrum, and interference from other cell sites. Wireless QoS is an extremely complex topic that involves more factors than wireline, such as power issues, distance from a cell site, line of site issues, geography, protocols, etc. Cell-site backhaul experiences similar outage impacts as described in the analysis of availability for architectures in Fig. 11.4a–c and then in addition the wireless portion of cellular service is affected by the wireless QoS impacts mentioned. Therefore, at first glance it appears that cellular service has

much worse network resiliency than the wireline residential broadband Internet architectures.

However, there are two differentiating and mitigating factors that may blunt some of this conclusion. First there are different QoS metrics between wireline and wireless. Because of the limited spectrum, substantial sharing (spectrum reuse), and different received power signatures of a cell service, the provided bandwidth (throughput in either direction) of the cell service does not enjoy the consistency of wireline throughput. Thus, it depends on how we define *availability* of cell service. Within a serving area, it is rare that one is totally disconnected from a cell site. Conversely, its QoS can be significantly reduced because of all the factors described above. Thus, like packet services in Sect. 11.2.1, if we define the availability QoS threshold sufficiently low (i.e., if QoS is below this threshold, then the service is unavailable), this architecture can almost be as resilient as wireline broadband service. Alternatively, if we define this threshold high, then resiliency is much lower because of reduced throughput. However, there is no universal definition of what such a QoS threshold for cellular service should be.

The second factor involves overlapping coverage by multiple base stations. Surprisingly, the second factor can indeed raise the resiliency of architecture Fig. 11.4d ABOVE that of architectures in Fig. 11.4a–c. For, the cell service will not enjoy availability to the SAME cell tower better than the architectures in Fig. 11.4a–c because cell-site backhaul experiences similar outage impacts. So, from this restricted viewpoint this classifies the service with a similar availability, around three nines, assuming that the threshold QoS defined above is less than that of its wireline counterpart. However, almost all cellular devices are within range of multiple cell towers. Since most cell towers (in the USA) are miles apart, they are likely to have different first gateway COs and thus diverse paths over the metro-access network, including possibly some part of the path routing over line-of-site microwave radio. Note that this first CO is not the MTSO, but rather the first point where the backhaul path intersects the metro-core network. This fact provides a very robust mechanism for resiliency. For, if an outage causes a cell site to become disconnected from the network, it is very likely that a cellular end device can and will switch to an alternate cell site, which is unaffected by the outage. Ironically, while this overlapping of cell sites is one of the most difficult problems in cellular engineering and management, namely *signal interference-to-noise ratio* (*SINR*), it also provides a robust methodology against network outages. If switching an end device to one of the surviving cell sites does not lower the throughput below the QoS threshold, then the network is resilient against the outage for those end devices. One can see an additional complexity of cellular QoS in that different end devices will experience different availability; uniform coverage definitions are difficult. Note that the addition of more small cell sites, as suggested by 5G, may further enhance the network availability, assuming the fiber paths to the neighboring small cells are diverse.

So can we, in general, claim that the network availability of a cell service as defined by the architecture in Fig. 11.4d is superior to the residential wireline broadband service architectures of Fig. 11.4a–c? On the basis of level of sophistication and implementation of cellular technology in today's cellular networks the answer is *no*. This is because engineering the QoS for cell service is significantly more complex, requires significantly more investment than that of wireline services, and has required decades of emerging wireless technologies, plus considerable network deployment and investment to get cellular services to today's level of widespread use and expectation of QoS. For example, if we set the threshold of QoS availability as that which enables uninterrupted, high-quality video service on data phones and personal devices, a rate typically in Mb/s, within a carrier serving area (assuming no service-affecting network outages, no bad-weather states, and ignoring video provider server difficulties), then we find that the level of availability differs widely across geography and different carriers for the reasons outlined previously. Another real-life example that can force the availability of cell service to be less than wireline, is the confusion of end devices to choose between Wi-Fi access points and cellular base stations. Many of today's end devices tend to latch onto Wi-Fi access points that have little or no throughput and/or get confused between choosing cellular or Wi-Fi communication links. This is because there is no reliable TCP-like layer that sits below TCP (or other streaming protocols) that can evaluate the availability/QoS between the links and seamlessly switch to the more reliable of the two. As an end device moves around, this problem occurs more frequently. This problem often causes downloads, such as steaming video content, to hang or drop. This is a good example of what we meant when we said that the resiliency of today's wireless networks even with the robustness to choose among surviving base stations is still short of wireline QoS. However, the capability to choose among different underlying wireless transport is still rapidly evolving and therefore, theoretically, as these difficulties are surmounted it will be possible for cell service availability to exceed that of the residential wireline broadband services described above (as we have defined availability in terms of a packet service QoS threshold). An example of protocol work to introduce such a layer is given by *multipath TCP*

(*MPTCP*) [11.30]. This protocol adds a layer below TCP that maintains multiple paths (lower layer links) and chooses among them. For example, it could maintain a background keep-alive connection over a cellular network virtual link as well as over a Wi-Fi connection.

One other outage issue for cellular service in this context merits discussion. The major environmental events listed in Fig. 11.8 can bring cell service availability below that of wireline broadband service. In particular, those events that cause power outages at the cell sites are most problematic, whereas this impact is usually less to the metro-access portions of wireline residential broadband Internet services (architectures Fig. 11.4a–c). For example, the PON architectures of Fig. 11.4a or Fig. 11.4b either see power outages at the apartment site (in which case all customers have lost service at that site anyway, regardless of architecture) or at the OLT site or minihead-end in a CO or hut. The latter generally have backup power. However, such is not the universal case for cell sites. As mentioned earlier, earthquakes and hurricanes have forced many carriers to think about equipping their cell sites with automatic back-up generators – and, of course, this capability in turn requires costly regular maintenance and certification of generators and safe fuel supplies [11.19]. Because the world has evolved so predominantly into wireless service some of the latest approaches to mitigate severe environmental disasters use rapidly deployable equipment and methodologies. Some of these are called a *movable and deployable resource unit* (*MDRU*). An example is given by the devastating 2011 Japan Tōhoku earthquake and resulting tsunami that disrupted communication facilities, as well as severely crippled nuclear power plants [11.31]. Other examples can be found in [11.17]. Even if these events only occur once per decade, if we factor in a multiday outage across an entire coverage area (that is, where a wireless end device is within range of multiple affected towers) into the yearly downtime calculation it raises the downtime significantly. For example, a five-day outage is 7200 min of downtime. Divided by 10 years, this is 720 min year^{-1}, which forces the availability to below three nines per year.

11.3 Evolution of Resiliency in Optical Networks

Looking at the present day view we provided previously, the inquisitive reader has already asked the question, *would we architect the same end-to-end architecture if we start from scratch with today's available technologies?* The answer is usually: *No*. This is because the architectures we presented arose from a slow evolution of competing and emerging forces, such as new services, new technologies, carrier corporate business models, emerging economics, and market realities. To prove this point, we can look at the types of network architectures that are quickly installed in emerging countries. They often look quite different to what we presented above [11.32].

11.3.1 Early Optical Transport Systems

We will now explore how and why the metro-core and intermetro-core networks evolved, but still focusing within our overall goal of network resiliency. This will explain what may seem at first glance an odd collection of architectures for network resiliency in Figs. 11.3 and 11.4. The earliest optical transport systems used nonstandard optical encryption schemes. The capacity of these early systems ranged roughly from 45 to 135 Mb/s. They were built to transport DS1 (T1), $n \times$ DS1, and DS3 signals in the USA. There were other intermediate rates defined, such as DS1C, DS2, but these rates/formats never played a large role. Plesiochronous rates, such as 2 Mb/s (E1) and 34 Mb/s (E3), were more common in Europe. These DS*n* or E*n* signals were architected mostly to carry voice trunks in early optical transport systems.

These early optical systems were generally point-to-point systems deployed in a two-way coupling, meaning that going in one direction all the constituent (client) signals were multiplexed into channels at the near-end terminal, then demultiplexed into drop ports at the far-end terminal and then vice-versa in the opposite direction. These early systems offered various forms of link protection/restoration, wherein a parallel pair of fibers could be connected to the equipment bay and high-speed switching would switch to the alternate fiber pair if an outage was detected. Switching was provided either in a 1:1 fashion (rerouting the channels of the high-speed system using only one set of high-speed cards) or 1+1 fashion (also called *path protection*), where the high-speed signal is split over a primary and backup/redundant path and the healthy signal is selected at the end terminal. This path selection could also be done on the ports of the receiving card of a channel independently of other channels carried by the optical transport system. There was also investigation into various 1:N restoration methods, where a single protection system covers multiple working systems. However,

1:N protection was not generally deployed for fiber or transponder outages and instead evolved mostly to protect shelf plug-ins and cards. That 1:N card protection strategy is still employed today.

11.3.2 SONET/SDH Optical Transport Systems

As optical transport systems became more prevalent, the deployment of many point-to-point systems, each with its own with line/fiber protection, became unwieldy and expensive. To overcome this problem, various add/drop features were added to intermediate terminals in these pre-SONET/SDH systems. However, these early optical transport systems were very limited in their add/drop capability and the variation in vendor product capabilities was wide.

To help understand this evolution of optical networks and their resiliency capabilities, it is helpful to study the evolution of the USA intermetro network, which was dominated in the early years of optical network development by the Bell System and AT&T. The AT&T Bell System was divested in 1984 following a USA federal court order called a *consent decree* into eight corporations: AT&T and the seven *regional Bell operating companies* (*RBOCs*). In addition, the RBOCs formed their central R&D arm called Bellcore in that same year. However, the intermetro (long distance) network remained with AT&T after divestiture and so we see a consistent evolution of the intermetro architecture through the 1980s and beyond. AT&T had settled on a pre-SONET 1.7-Gb/s optical system, called *FT series G*, for transporting DS3s through the late 1980s and middle 1990s [11.33]. The copper-based DS3 (or T3) signal was its target CO intraoffice cross-connect signal (i.e., client/constituent add/drop signal to the optical terminals). They used a Broadband DCS (tradename DACS-III) to cross-connect the DS3s in a CO, analogous in concept to that of the OTN switch in Fig. 11.7, although more technically primitive. This intermetro optical transport network of the era had some add/drop capability and employed forms of link/fiber protection for its high-priority private-line traffic. However, to reduce the cost of expensive 1:1 or 1+1 fiber system protection, they leveraged the cross-connect ability of the Broadband DCS to implement a centralized DS3 rerouting scheme called FASTAR [11.34]. Because FASTAR used a centralized controller to send cross-connect commands to the Broadband DCSs, this restoration scheme was relatively slow for large outages, i.e., slow compared to later SONET ring restoration speeds and IP-layer virtual rerouting. This was because at the time, their core network mostly transported voice trunk groups: DS0 trunks multiplexed into DS1s and then DS1s multiplexed into DS3s. The voice trunk network was very highly connected and so had its own form of dynamic (alternate) routing and therefore the speed of DS3 restoration was not as critical for this service. For more detail on AT&T's dynamic routing in circuit-switched (mostly voice service) networks see [11.35].

The reader may have already started to ask the question: *Why are the topics of FASTAR and pre-SONET fiber optic systems important to today?* The answer is that while AT&T had settled on (what it thought at the time was) its long-term transport and resiliency architecture, the evolution to today's networks started leapfrogging rapidly from that point in time. Also, the metro carriers did not go with the same architecture and were busy exploring alternative technologies, literally a technology generation ahead. As mentioned previously, the point-to-point pre-SONET/SDH fiber optic systems used in the metro-core network were becoming unwieldy. Note that during this period (1984 to the middle 1990s), voice trunks were still the predominant traffic/demand to be transported in the metro-core networks, excluding the MSO cable companies, who during that time mostly distributed rebroadcasts of over-the-air television channels. Bellcore was tasked to find a better solution to transport the odd collection of TDM signals (including Europe, Asia, and North America) and to standardize (even possibly synchronize) the signaling formats in the emerging optical transport systems. Thus, SONET was developed. The first accomplishment of SONET (later followed by the SDH standard in the European-based ITU-T), was to define the next generation of constituent digital rates and channels inside optical signals by defining the basic SONET synchronous transport signal (STS)-1 (SDH STM-1) internal rate. Bellcore decided to make a synchronous digital encapsulation, wherein as one increased the optical line-rate multiplexing capability, the bit rate would remain proportional. However, they also incorporated stuffing bits to accommodate the various asynchronous subrate signals. This standardization also enabled easy add/drop capability and the first flexible ADM. An important side-light of this standardization activity was the standardization of overhead bits and signaling bits. This enabled one to rapidly pass alarm information as *signatures* within the SONET overhead along the route of ADMs in a SONET system. Signature means a repeating bit or symbol code rather than an explicit protocol message. Once the add/drop standards were developed, the next step was standardization of a restoration/protection scheme and, in particular, one that was more efficient than the overlapping, point-to-point pre-SONET systems. By *efficient*, we mean both more automatic and controllable from a network management standpoint, as well as more capacity efficient.

At this point, a well-architected approach to optical-layer restoration could be adopted. The most deployed products were the SONET/SDH UPSR and BLSR, which we described in Sect. 11.2.3. However, we will find later that this new-found euphoria would soon run its course. Once the standard was developed and vendors developed products, the metro-core Telephony carriers in USA, Japan, Europe, etc., began to install SONET and SDH systems en masse. Because of the size and relative proximity of the countries in Europe, SDH rings could also be used for intermetro-core transport and restoration. In fact, as the Internet began to rapidly grow in the late 1990s, additional standards for encapsulating packet links (data transport) inside of TDM signals (e.g., DS1, DS3, E1, E3, STS-n and STM-n) enabled the carriers to reuse the same processes they had already developed for years that provisioned these TDM signals that transported circuit-switched trunks. For example, a DS3 signal that carried almost 45 Mb/s of data packets was called a DS3 *clear* signal. This enabled rapid growth of a retail and wholesale leasing market for data-carrying private lines (*circuits*). The term *wholesale* applied when a carrier leased such signals to another carrier and *retail* otherwise. In addition, these circuits could be sold as reliable (resilient) circuits, because the portion of a circuit in a metro-core network could be transported over SONET/SDH rings.

But what about network resiliency in the intermetro-core network during this rising period of SONET/SDH products? Use of the SDH rings for the intermetro-core segment in Europe, Japan, and other geographically smaller countries solved some of that question. But in the USA, as we described previously, the dominant intermetro-core network carrier (at that time) used pre-SONET optical transport systems and a slow DS3 rerouting scheme for most of its traffic. What occurred next was a tumultuous and short-lived love affair with SONET rings. AT&T attempted to cover its vast USA intermetro-core segment with a series of overlapping OC-192 BLSR 4-fiber rings. Note that to function as expected (from a QoS standpoint), the ADM-to-ADM links of a BLSR ring must be fiber-diverse of one another, else multiple link outages have a much higher likelihood of occurrence. Multiple link outages can cause a set of nodes of the ring to become disconnected from another set, in which case the ring has much less resiliency.

This was a daunting task to route and develop a series of massive rings across the country. It was quickly discovered by researchers that there were far better approaches for the USA intermetro core. One of the earliest articles on this can be found in [11.36]. These early studies showed that SONET rings were, from an economic standpoint, a bad choice to provide restoration at the optical layer in the intermetro-core segment. The optimization principal behind this discovery was that with a well-connected network (e.g., every major switch location has a topological degree ≥ 3), rings do not take advantage of maximal sharing of capacity over nonsimultaneous, potential fiber-span outages. Furthermore, one of the major shortfalls of SONET rings in intermetro networks (which is true also in metro-core networks) is that while they appear attractive when the entire network can be covered with a few large giant rings, when the network size and increasing demand dictates that multiple rings begin to overlap and overlay, the SONET/SDH network-wide architecture becomes unwieldy and expensive (plus a BLSR is limited to 16 ADMs). To understand this problem of multiple ring deployments, consider when a circuit needs to cross two rings. It needs to demultiplex and be connected through a fiber patch panel at the CO where the ADMs from the two rings are colocated, analogous to the cross-connect at locations C and D in Fig. 11.6. As the bandwidth requirements of these networks grew, plus the fact that commercial SONET/SDN rings never effectively exceeded the 10-Gb/s line rate, it became necessary to deploy many *intersecting* or *stacked* rings. *Stacked* rings mean that two or more independent rings are deployed whose ADMs are located in the same set of COs. However, there is nothing particularly special about two stacked rings. They are basically just a special case of two rings whose ADMs intersect at one or more COs. For, if a connection needs to be transported over the two rings, then it must add/drop at an ADM of each ring in the same CO and be cross-connected via a patch panel between the ports of the two ADMs. Extrapolating this problem over time, as the size of the Internet and private packet networks grew, deployment and management of SONET rings eventually became unwieldy, even in USA metro-core networks [11.1, Fig. 6]. Some metro networks grew to literally thousands of SONET rings, making the network virtually impossible to visualize. In fact, as of the writing of this chapter, many networks still have this situation. Furthermore, the management of the inter-ADM circuits via patch panels became a challenge unto itself.

We note that this intersecting ring problem could potentially have been ameliorated with deployment of a broadband DCS that internally integrates/connects the two ADMs. In fact, some vendors did provide potential solutions for these hybrid architectures by producing *ring cards* that could be installed on a broadband DCS. But this leads to the important question: if a DCS needs to be used to manage the massively overlapping ring problem, then why not use a better overall mesh restoration scheme, as potentially provided by the DCS? These issues plus the restoration optimization analysis work

of [11.36] led to the advancement and deployment of the intelligent optical switch (IOS) that we briefly described earlier. However, some breakthroughs in optical and digital multiplexing technology were needed to enable this evolution. To understand these breakthroughs, we digress a bit.

In telecommunications technology, the concept of multiplexing signals was invented to take advantage of economy of scale. For, using the same media (copper, radio-frequency-over-air, fiber, coax, etc.), one can transport more individual signals by multiplexing them via different technology schemes, such as analog carrier or digital time-slotting. However, there is a downside in that to provide the ability to add or drop individual signals which comprise the channels of the multiplexed signal, one must figure out an efficient way to *demultiplex* the signal. As we described previously, prior to SONET/SDH, one effectively had to demultiplex the entire signal and, as a result, this economic tradeoff created an extremely difficult network design optimization problem, called the *multiplex bundling problem*. For example, see [11.37], which was the first paper published in the newly formed Telecommunications Section of the Operations Research Journal. SONET time synchronous technology was developed to more efficiently multiplex and add/drop individual channels. However, SONET optical transport systems were only offered in very simple architectures, such as linear chains and rings. The IOS technology offered a breakthrough in that it found efficient ways to demultiplex the SONET high-speed signal into its constituent STS-1 channels at the incoming port on every IOS switch and then reassemble the through signals with any newly added or dropped signals on the outgoing ports of the switch. This required very high-speed digital processing so that delay and jitter would not accumulate and exceed a given tolerance. Thus, IOS technology overcame the historically difficult, combinatorial multiplex bundling problem, as well as enabled the deployment of more practical mesh-network deployments. In addition, another benefit was that it also enabled the development of rapid restoration methods based on more capacity-efficient mesh-network topologies. This answers the question we posed above.

The IOS is basically a progenitor of the OTN switch and thus the restoration method in the IOS is very similar to that of the OTN switch (described in Sect. 11.2.2), except the IOS uses SONET STS-1 (51.84 Mb/s) or SDH STM-1 (155.52 Mb/s) channelized building blocks versus ODU-0 (1.24 Gb/s) building blocks, as in the OTN switch (however, note that the internal switching architecture of the OTN switch may use even smaller data chunks than what we listed). One final remark in this section is relevant to our later discussion on the difficulty of implementing ideas for network resiliency in commercial carrier networks. The IOS paved a path for mesh-network restoration in optical networks. Yet, even though some of the first vendor models were demonstrated at the Optical Fiber Communications Conference (OFC) in 1999 by Lightera, and given the attractively simple-sounding idea of storing single fiber-diverse restoration paths for each provisioned connection, it took almost a decade of implementation and a staff of highly qualified analysts, optimization experts, and planners to deploy it to its full capability and maintain it (Lightera was later acquired by Ciena who renamed the switch *Core Director* [11.38]). In fact, as of this writing, it is still deployed and providing network resiliency in the intermetro-network segment [11.23].

11.3.3 The Decline of the Path-Switched Ring and the Emergence of IP-layer Restoration and Ethernet

Concurrent to the widespread deployment of SONET and SDH rings in the mid-1990s, packet networks were rapidly deployed. Carriers started by routing their router-to-router links over SONET/SDH transport systems or over packet private-line services, such as frame relay-based PVCs. As described in detail in Sect. 11.2.2 one of the fundamental capabilities of Layer-3 IP control protocols is their ability to recover and operate from massive network outages. This distributed survivability feature was one of the original goals of the government-sponsored ARPANET, the progenitor of today's private-based Internet. Thus, from their origins packet networks based on IP protocols inherently contained restoration capabilities. During this period an early technical competition sprung forth about which network layer would be the ultimate provider of network resiliency. We remark that a lower network layer, ATM, made a brief appearance during this evolution and despite its strong endorsement by many significant parties (including the US federal government, Bellcore, ITU, some carriers), it soon thereafter made an ignominious departure and thus we skip its discussion.

During this period, UPSR rings became the darling of resiliency to enterprise locations in metro-access networks, assuming diverse fiber paths for the ring could be provided, which is usually the case in dense urban environments. However, as one can see, we do not even show a UPSR ring in Fig. 11.4e. With the emergence of ultralow-cost gigabit Ethernet interfaces and switches, the SONET/SDH UPSR fell out of favor and is being replaced by the myriad of packet and Ethernet-based diverse routing technologies. We again highlight the discussion of Sect. 11.2.4: to approach five nines

availability for premium Enterprise services, it is critical to have a media-diverse routing capability in the *last mile* of the metro-access network. However, it is important to note that while the UPSR was the legacy gold standard for Enterprise access, we cannot yet say there is a *gold standard* replacement for the evolution to Ethernet-based restorable routing or even if a gold standard will emerge.

One additional observation about the emergence of Ethernet-based interfaces is that as already mentioned, a historical wholesale market developed for leasing DS1s (T1s) for trunks for voice service and then later the DS1 was enhanced and standardized to carry packet data traffic (the un-channelized *DS1 clear*). This enabled the T1 to become the workhorse of the private-line carrier industry. Jumping forward a few decades, one can now think of the gigabit Ethernet (*GigE*) signal as its modern day analog. For, almost all enterprise services today interface with carriers at $n \times$ gigabit Ethernet signal/rates, but usually contract for a smaller CIR than the maximum gigabit rate of the interface. Cell-site backhaul is an excellent example of this evolution. Cell sites used to interface with carriers via the $n \times$ DS1 interface. In today's networks, carriers mostly upgraded this interface to gigabit Ethernet, as shown in Fig. 11.4e, but police the traffic to the CIR appropriate for the traffic on each cell site. Thus, a wholesale market for leasing private-line GigE signals has arisen.

11.3.4 DARPA CORONET Study

Perhaps the most comprehensive study of the best combination of technologies to provide resiliency in an intermetro-core network was undertaken during the DARPA CORONET program [11.39]. The program began under the leadership of *Saleh* and *Simmons* and was motivated in part by their earlier history with intermetro carriers, optical start-up companies, and their original ideas about network design, as expressed in their paper [11.40]. Also see related work in the later book [11.41]. CORONET was a long-running DARPA program: a total of seven years and three phases, which by DARPA stringent contract renewal and review standards would have to be called quite successful. CORONET assembled teams of experts in network design and optimization from commercial carriers, academia, and the telecommunications optical and packet vendor community. We describe some of its findings here.

CORONET Goals

Some of the principal goals of CORONET were to explore the architecture, operation, and commercial feasibility of a packet-over-optical intermetro-segment network with the following features (requirements): (1) a dynamic optical layer, wherein circuits (either high-rate private line or links of the IP layer) can be rapidly provisioned under a highly distributed control plane, including *bandwidth-on-demand* (*BoD*); (2) high network availability, provided under a range of CoS for connection services at each network layer, where each class is targeted to survive a different severity of network outage; and (3) rapid connection/circuit provisioning. The severity of network outage was defined in terms of increasing combinations of fiber and/or total CO outage [11.42, 43]. We discuss later why they chose this particular outage model.

Because our primary interest in this chapter is network resiliency, we skip the provisioning and bandwidth-on-demand aspects of CORONET and instead briefly describe the types of solutions that CORONET recommended to handle its different range of modeled potential outages. Along with these recommended architectures, network capacity and performance studies, DARPA CORONET developed a nonproprietary, hypothetical (yet quite realistic) optical network topology of the USA network, shown in Fig. 11.14. Because of the participation in the project by analysts and network designers associated with large carriers, this hypothetical optical network has topological characteristics that are similar to those of the largest USA intermetro carri-

Fig. 11.14 DARPA CORONET hypothetical intermetro network

ers. These include key cities, fiber spans, DWDM links, node diversity, etc.

CORONET also derived a hypothetical global optical topology. We do not show this expanded global topology because it was derived mostly to test the global limits and constraints of the restoration algorithms, latency, and provisioning times, rather than to represent a realistic commercial, carrier-based network.

Summary of CORONET Results

CORONET defined many aggregate classes of demand/service. These included packet services and highest-rate private-line services, called *wavelength services*. Links between routers in the IP/MPLS layer that are provisioned to provide packet services were basically treated as a *slow* demand for high-rate private-line services, whereas other private-line services were classified as either *slow* or fast/dynamic (bandwidth-on-demand). The first Phase of CORONET did not model lower rate private-line services (below wavelength rate) and tacitly assumed these would be provided by the packet layer. However, it was found as the project progressed and through carrier feedback that providing private-line demand only via the IP/MPLS layer was not commercially practical and as a result an OTN-layer was added in later CORONET Phases. Thus, when this change is incorporated into later CORONET project progress, the good evolutionary news is that the intermetro-segment picture of Fig. 11.3 is fundamentally consistent with CORONET findings. [11.42], [11.43], and [11.44] provide many other important findings and details, such as the two-phase restoration in the IP/MPLS layers, which is similar to what we described in Sect. 11.2.2, how to design restoration in such networks to achieve class-of-service objectives, such as best-effort and premium services, how to mitigate router outages in a capacity-efficient manner compared to dual-homing, how to provide bandwidth-on-demand, and results on provisioning times.

We will not delve into these details, but instead briefly mention the results of the CORONET study of optical-layer restoration and higher layer restoration. CORONET proposed highest-rate private-line services defined by different survivable classes, namely, single (optical) link outage, simultaneous dual link outage, CO outage (restoration of through traffic only), simultaneous link and CO outage, and simultaneous triple outage. However, to be realistic, as one steps through the classes-of-service that must survive the more stringent outages, the magnitude of the corresponding traffic class decreases accordingly. For example, to expect large amounts of wavelength traffic to survive the worst-case type of outage, namely a simultaneous triple outage, is unrealistic because it would require a very highly connected network with significant spare capacity; therefore, this highly survivable traffic class was specified to be a small fraction of the total, e.g., no more than 5%. Furthermore, there were time limits placed upon how long it took to restore a circuit. CORONET project requirements required very rapid restoration times, e.g., approximately 50 ms or less for single link outages within the continental USA.

Before proceeding to further discuss studies/results on optical and IP/MPLS restoration methods in the early CORONET phases, we want to ground them within the commercial context of this chapter, namely that we are aware of no carrier who sells wavelength-rate private-line services as defined by the CORONET survivability classes; consequently, one should think of the CORONET demand/traffic classes and their network requirements as illustrative and the corresponding results as helpful to explain how and why network architectures evolved to their present state. Extending beyond commercial carriers, DARPA is a USA Department of Defense (DoD) entity and, as such, its affiliated federal agencies had interest in the results of the study of these conceptual outage traffic classes to help answer generic questions of USA network survivability and how such outage-stringent private-line services could potentially be offered within a commercially feasible approach. Furthermore, the USA federal government participates in and/or supports many RENs. Constructing such advanced services and corresponding network capabilities on a limited basis is more feasible than in large commercial networks. Also, Phase III of the CORONET project evolved to study such private-line capabilities using dynamic connection provisioning (bandwidth on demand), but within a commercial business context. We discuss more of this aspect when we discuss private networks starting in Sect. 11.3.6, *CORONET BoD*.

If we now refer to Fig. 11.3 (i.e., the colored network layers) and back to the discussion of resiliency in today's intermetro-segment networks in Sect. 11.2.2, we mentioned that there are no commercially deployed networks with optical layer restoration other than the OTN switching layer, legacy SONET/SDH, or legacy IOS layer. However, we also mentioned that these are not *purely optical* restoration methods, wherein optical lightpaths themselves are reconfigured. To explore why this situation transpired, it is useful to compare against a hypothetical, yet practical optical-layer restoration method. To this end, CORONET researchers decided to recommend and study a wavelength connection restoration strategy that had the most practical chance of

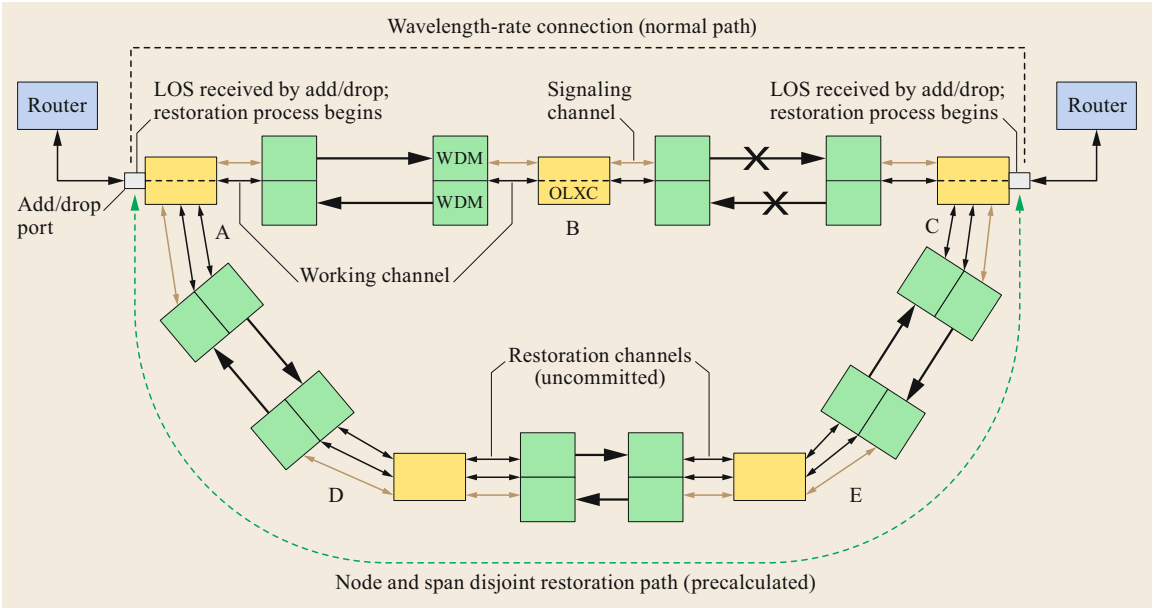


Fig. 11.15 Failure detected. ROLEX restoration: Outage detected at both ends (after [11.45])

implementation. They converged on a patented optical restoration method, colloquially called *rapid optical layer end-to-end X-connection* (*ROLEX*) [11.45, 46]. This restoration method has similar restoration characteristics to that of the OTN switching layer described previously, in that a secondary path that is fiber-diverse of the service path is precomputed and stored in the originating (and possibly terminating) node(s), namely, ROADM(s) or optical cross-connect(s), of each provisioned circuit.

This method is triggered by detecting a loss of signal or control-plane notification of an individual wavelength-rate connection (circuit) at the origin and/or terminating nodes of that connection. See Fig. 11.15, borrowed from [11.45]. In the two-ended version of ROLEX, a message protocol is initiated over the digital overhead channels of the circuit from both ends and along the precomputed restoration path. Note that such data overhead channels are clearly available in OTN standards. This protocol attempts to reserve the spare wavelengths (channels) in each node-to-node link along the restoration path, including algorithms for contention control as multiple circuits vie for unused wavelengths (channels) (Fig. 11.16). These messages meet somewhere in the middle of the path. Note that if the source side message travels faster, it could even meet at the opposite side of the first link (Fig. 11.17). The cumulative restoration time is minimized by wavelength (channel) cross-connection being enacted as the protocol progresses, instead of waiting until the process reaches the end node. However, if the signaling process is unsuccessful at any point along the path, the process is halted and a *back-off* process is initiated that sends disconnection messages along the path to clear selected channels.

Of course, at a minimum the network must be designed with sufficient spare wavelengths and transponders for the process to restore all high-rate private-line services against any single fiber link (span) outage. Fiber spans are shown in blue in the CORONET hypothetical network of Fig. 11.14. If this is so, then the probability of occurrence of circuit back-off is low. As in the OTN switching layer, when more complex outages occur, wherein both the primary and secondary paths fail or there is a high degree of contention for spare channels, then a second (and slower) dynamic path selection process is instigated to try and find alternate paths for those connections that could not be rerouted on the first attempt. We note that CORONET analysts did study methods to also precalculate and store multiple alternate paths to handle the more complex outages. However, given the inherent difficulties in implementing optical-layer restoration, we feel that these intense precalculation methods are not practical. Further note that CORONET Phase-1 requirements provided relaxed restoration time requirements for multiple simultaneous outages.

Much of the CORONET study investigated how to optimize the capacity of the network to provide the required network availability of the wavelength-

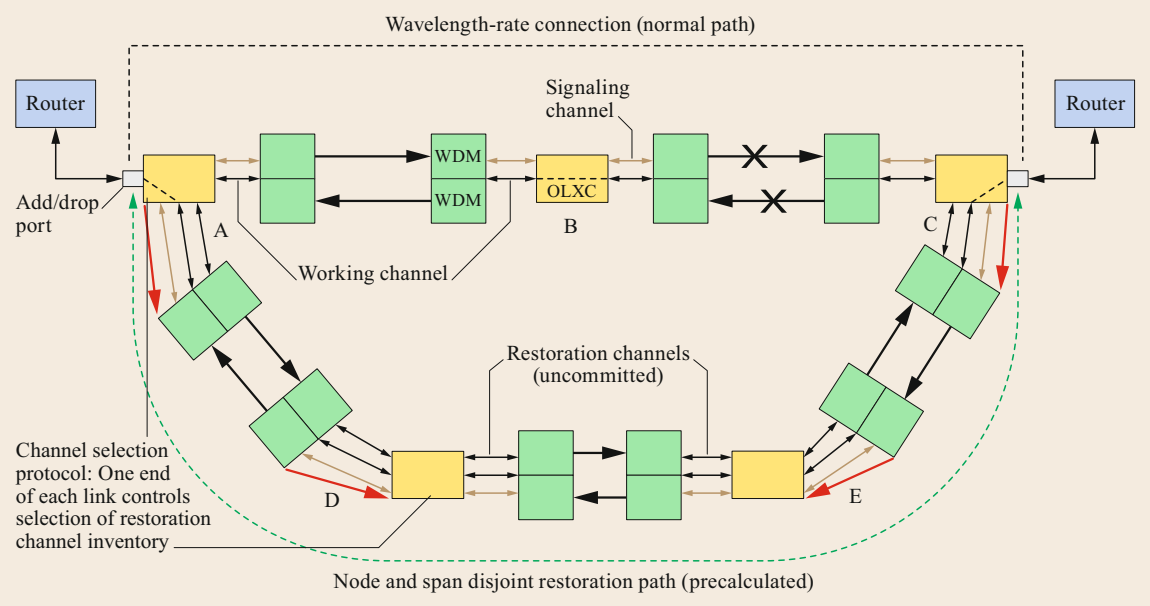


Fig. 11.16 Restoration process begins at both ends. ROLEX restoration: Signaling process instigated at both ends (after [11.45])

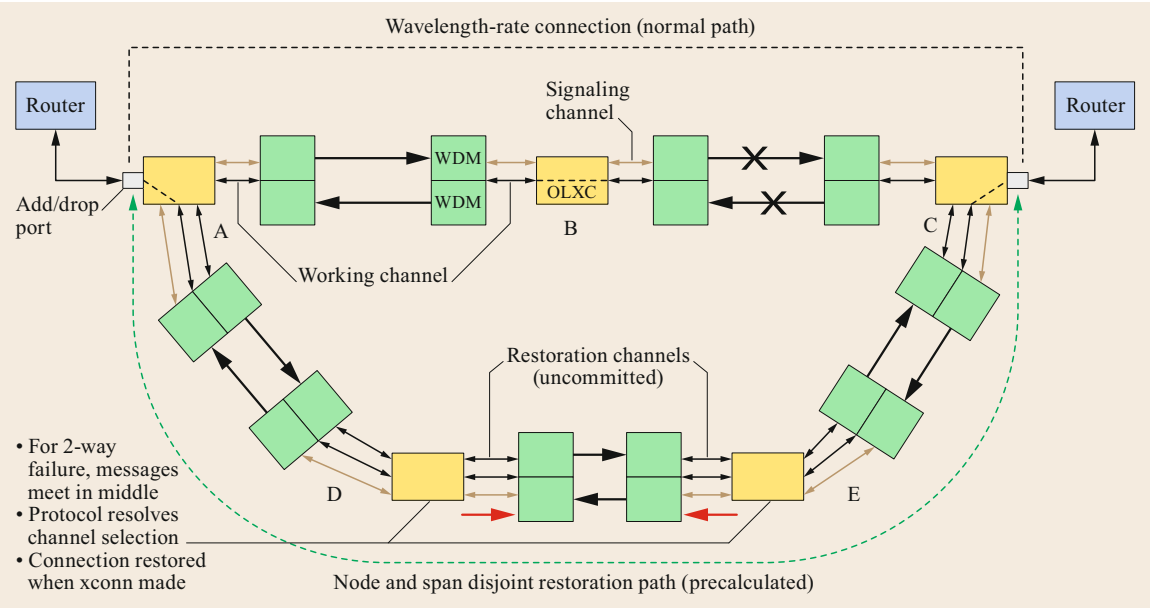


Fig. 11.17 Messages meet at middle node. ROLEX restoration: Signaling messages meet in the middle (after [11.45])

rate private-line services against these types of outages. Early phases of CORONET investigated simulation of the ROLEX protocol on the CORONET network with different sets of demands with varying traffic intensities. Phase-1 CORONET only evaluated the capability of the ROLEX control-plane protocol to converge properly and evaluate convergence times on reasonable controller hardware platforms at each optical node. Surprisingly, Phase I CORONET found that the control-plane portion of the optical-layer restoration method generally achieved the restoration time requirements within the hypothetical USA network. See [11.43, Sect. V.B.].

11.3.5 The Emergence and Dominance of IP-layer Restoration

One of the milestones in the evolution of resiliency in the metro-core and intermetro segments, is the migration away from lower layer resiliency of the past (mostly SONET/SDH-based architectures) to the situation of today where IP/MPLS is the principal layer where resiliency is provided. In this section, we will examine some of the motivations and realities that molded this evolution. But before we list these reasons, let us review a few key observations in the history of metro-core and intermetro evolution:

- At the time of their development and subsequent massive deployment around the world, SONET and SDH transport systems emerged as a standardized optical technology that solved most of the complex multiplex bundling problems as well as established a robust market for low-cost leasing of private-line circuits that could transport either voice trunks or links of emerging packet networks, the so-called *clear* un-channelized encapsulation protocols, such as clear DS1, $n \times$ DS1, DS3, $n \times$ DS3, E1, E3. In addition, SONET/SDH rings provided a resilient lower layer transport technology with very rapid restoration speeds (generally within 50–100 ms) and resulting high network availability.
- As data traffic increased, literally many orders of magnitude over decades, the SONET/SDH-based IOS switch technology provided a more capacity-efficient and provisioning-efficient alternative than SONET/SDH rings, although restoration speeds were slower.
- The OTN switch was a next-generation evolution of the IOS switch that provided significantly higher signal-rate capability, was compatible with the emerging OTN standards, provided the ability to seamlessly encapsulate multiple packet-based signal formats, and provided superior restoration capability than its predecessor IOS switch.
- As DWDM technology became common and wavelengths achieved higher rates, the DARPA CORONET project suggested a reasonable technology approach to provide optical-layer restoration.

Then, given all these seemingly attractive points, why does the intermetro segment of Fig. 11.3 show restoration capability at the IP/MPLS layer, but no restoration capability in the optical (DWDM) layer, plus a large arrow from the IP/MPLS Layer that skips the OTN layer? Most of the answers to this question lay buried in the internal decision making, network economics, and operations of large carriers; hence, there is no simple and quotable outside reference that we can offer to answer this question. However, we compiled a list of factors that we have derived from years of analyzing, architecting, and optimizing these networks.

Factor 1: DWDM-layer Restoration Was not Fully Proven as Viable in CORONET

While the CORONET Phase-1 ROLEX methodology appears to be a viable and attractive optical layer restoration alternative, one of the tasks of CORONET Phase-2 and Phase-3 was to suggest a path to commercialization of the ideas from Phase 1. However, as in all research projects, along the way new realities and directions emerged. In particular, while Phase-1 simulations demonstrated that the ROLEX control plane can achieve very fast restoration speed, it turned out harder to prove when the project attempted to introduce actual hardware devices whose optical cross-connect (provisioning) capability could match the speeds of the control plane. Essentially, a CORONET field trial explored the characteristics of existing ROADM technology connection (lightpath) provisioning that needed to be advanced to achieve these very rapid restoration times. These speed roadblocks included amplifier wavelength power balancing processes, laser tuning, and methods to filter out optical impairments in the transponders. Unfortunately, the results of this analysis were unpublished, but the findings were found to be consistent among most leading DWDM equipment vendors at the time. Another significant roadblock to advancing these characteristics was that from the point of view of carriers, the DWDM lightpath provisioning process was never developed to be rapid because it was part of a planning and provisioning process, the so-called *capacity provisioning process*, which historically was (and still is) generally interwoven with *operations support systems* (*OSS*) and human provisioning processes – but certainly never envisioned to occur on the order of a few seconds or less. To modify equipment and provisioning/control systems to solve these problems and perform subsecond provisioning along a path of ROADMs or optical switches requires significant equipment development, as well as streamlining of the OSS environment. Plus, an accompanying and convincing business case would be needed to motivate carriers to consequently push their DWDM vendors to enable such rapid features. This business model would have to show (1) business necessity for such rapid restoration/provisioning; (2) convince carriers that it could (even technically) be accomplished; and (3) be accomplished at small additional cost. After establishing these findings, CORONET Phase III determined to study an alternate path to optical-layer provisioning and resiliency.

Factor 2: Centralized Control Re-Emerges

The original CORONET requirement was for a restoration mechanism under distributed control, as evidenced in Figs. 11.15–11.17. However, before the CORONET project was initiated in 2006, distributed network control was *in vogue* in the Telecom industry, especially in government; but, during the time of the CORONET project (2006–2014), distributed control fell out of favor and yielded to SDN-based *centralized* control ideas. Given the above points and the lack of a strong business case for rapid, distributed wavelength-rate restoration, CORONET Phase-3 changed course and pursued a more fruitful approach to use the emerging SDN capabilities to craft a more attractive business case for bandwidth-on-demand of wavelength-rate services, as well as subrate (below wavelength rate) services. This is discussed again in Sect. 11.3.6.

Factor 3: It is More Efficient and Economical to Engineer Resilience in the IP/MPLS Layer than in Lower Layers

Studies at the time showed that it was more efficient and economical to engineer resilience in the IP/MPLS layer than in lower layers. This is a complex topic that requires the analytical techniques we described previously concerning cost-optimal network design, CoS/QoS modeling, and characteristics of the outage probability space complexities. Ultimately these became business decisions inside carriers. But without delving into mathematical network optimization and business cases, we list some the reasons:

1. To achieve high network availability, the reliability experience with router hardware, protocols, and required maintenance procedures (such as software upgrades) indicated that routers would have nontrivial downtime. Therefore, IP backbones are often designed with sufficient spare capacity to restore the network (or at least premium services) from the potential outage of an entire router, whether due to hardware/software failure, maintenance activity, or external event. To better understand this, note that no amount of the restoration capability in the lower layers (OTN or DWDM) can protect against a higher layer node outage (e.g., downed router). Therefore, as multiple studies showed, the installation of adequate packet-layer link capacity to restore against router outages also provides significant protection against nonsimultaneous outages of other types (e.g., router component equipment, optical-layer failures). Router/node coverage does not usually protect 100% against (say) single fiber outages, but optimal placement of a small amount of additional capacity does the trick. Placing DWDM layer capacity for this little bit of extra capacity was not cost-effective compared to placing it in the IP/MPLS layer.
2. IP/MPLS-layer links have discrete capacities defined by their line rate. Usually only a few critical links are at maximum QoS-threshold utilization (where *threshold* is defined as before to mean the level of maximum utilization that provides a desired packet QoS for the given classes of service); therefore, there is usually a large amount of such *marginal* capacity available in most links, even after potential traffic is rerouted over them to accommodate the worst-case outage. This effect, when combined with the above extra capacity for router outages, made the case against DWDM-layer restoration even stronger.
3. Finally, one of the strongest reasons is that, as mentioned, packet services can be prioritized into classes (CoS), each with its own QoS target. In fact, one of the key factors in determining the theoretical QoS of the premium class is how that traffic is treated during potential network outages or unexpected traffic surges. As an example, a common business practice in carriers is to allocate no extra IP/MPLS capacity for restoration of best-effort traffic. However, it is important to note that does this does NOT mean best-effort traffic gets clobbered at every outage. To the contrary, the lower class traffic usually only suffers when hit by one of the potential *worst-case* outages and even then some amount of best-effort traffic will likely not suffer. Without delving into probabilities and maximal potential link traffic flows, we can state this mathematical phenomenon as follows: to meet the desired network availability, the outages with the worst expected loss (probability times magnitude) generally dictate the main amounts of extra capacity needed for restoration. Stated (over) simply, the network is designed for the worst case. Thus, if an outage of less magnitude occurs, there is usually enough capacity for most traffic to be rerouted successfully. And when best effort and premium traffic are being rerouted over a common router link, priority queuing allocates the capacity first to the premium traffic. Thus, the effect is that the best-effort traffic either gets through without QoS degradation or some of it might experience delay or some loss during the worst outages. None of the methodology and nuances that we just described about the optimum handling of premium versus best-effort traffic during outages can be equivalenced by providing DWDM-layer restoration. This is because DWDM-layer restoration is brute force: it (coarsely) restores entire router links at full wavelength capacity by rerouting them over alternate DWDM links.

In summary, initial packet-layer networks were routed over SONET rings and then later some carriers routed them over IOS networks as multitudes of overlapping, overlaid SONET rings became impractical. But, as Internet bandwidth soared through orders-of-magnitude growth over the past two decades, the design of IP/MPLS-layer networks for reconfiguration started to dominate resiliency planning. MPLS fast reroute and capacity-sensitive Layer-2, Layer-3 routing were later installed to further improve this resiliency. Thus, for the reasons cited above, more IP/MPLS-layer links were provisioned directly onto the DWDM layer at the wavelength-rate and with no lower layer restoration provided.

Optical-Layer Restoration Ideas in the Literature

There is a plethora of published ideas to provide optical-layer (or integrated IP-over-optical-layer) resiliency, yet very few are implemented in commercial networks. To give the reader some practical feel for why this is so, we discuss a once popular idea in the Academia, *p-cycles*, pioneered in a paper by *Grover* and *Stamatelakis* [11.47]. Our analysis here is aided by the DARPA CORONET study, which analyzed it in Phase-1. P-cycles attempted to address a limitation of SONET/SDH rings. While they perform rapid restoration, they are closed topologies and cannot share restoration capacity among one another. This impact is significant when one considers a metro-core network with hundreds or even thousands of small ring subnetworks. Thus, the high-level appeal of a p-cycle is clear: to achieve the speed of rings, yet mitigate the restoration capacity penalty.

While we will not describe p-cycles in detail, the p-cycle idea removes the rigid hardware identity of an ADM-based ring (such as a BLSR). The p-cycle idea is to precalculate a set of diverse paths and reserved restoration channels that route over a shared set of fiber links. A p-cycle identifies a set of reserved, restoration channels that are fully cross-connected in one big, closed loop. When a link of the cycle (consisting of many channels, including both in-service/working and restoration) detects an outage, for each (working) connection that routes over that link, the p-cycle node breaks (i.e., disconnects or un-cross-connects) its bidirectional cross-connects at the two end nodes of the link, then breaks the cross-connects of a restoration channel (on the cycle) at the same end nodes, and then cross-connects the channels of the working connection into those of the restoration loop. This implements a *loop-back* around the other direction of the p-cycle. Another feature of the idea is that a more *mesh-like* restoration strategy occurs for connections that route over a *straddling span* that is not on a p-cycle itself. This connection can choose to loop-back in either direction of the p-cycle. In fact, two straddling connections can be restored on the same numbered channels of the p-cycle by routing in opposite directions along the p-cycle.

The inaugural paper on p-cycles launched hundreds of follow-up papers in a extensive reference tree; consequently, many manifestations of this idea emerged in the literature [11.48]. Yet, the p-cycle idea was never implemented in commercial, production equipment. To help understand why, we temporarily regress to again examine the SONET/SDH BLSR. Analogously, even though the BLSR is appealingly simple at a high level, it took years to overcome operational difficulties. Some of these include: accurately detecting transitory failure states, maintaining a consistent topology view among all ADMs, correctly rerouting connections under multiple link outages, and reliably handling and reporting revertive/nonrevertive states.

In a BLSR, each ADM must keep careful track of the surviving topology. For, when two noncoincident links fail the ring becomes disconnected and no restoration should occur. However, if two coincident links fail (e.g., an ADM failure due to a power outage or other serious outage in the Tree in Fig. 11.8), loop-back is still possible for a limited set of connections. Because each ADM operates somewhat independently, no restoration should take place until the alarms have settled, otherwise mis-cross-connection and contention can occur. For example, consider the BLSR of Fig. 11.13b. Label the nodes (ADMs) starting from top left and moving around the ring clockwise, as A, B, C, D. Consider two 1-hop (1 link in length) connections: connection #1 routes over A–B and connection #2 routes over A–D. Since they do not overlap, these two connections can both be assigned to channels with the same number, $j \leq (n/2)$, where n is the total channels of the high-speed link (e.g., $n = 192$ for an OC-192 SONET ring). However, if node A fails, nodes B and D should not start restoration loop-back cross-connection procedures, else the messages will contend for the same restoration channel, $(n/2)+j$. Thus, the other nodes must kill all the connections that originate or terminate at node A, else two separate connections could either be mis-cross-connected or experience a channel-contention deadlock. To avoid this possibility, the SONET and SDH standards built in a messy feature called *squelch tables*.

The reason we described the complexities of a node outage in a BLSR is to illustrate how difficult it is to implement restoration ideas in a commercial network, no matter how seemingly simple, yet maintain the desired high level of resiliency. For, to achieve a five nines state of BLSR availability, one must carefully analyze,

handle, and test such corner cases. Turning back to the p-cycle idea, one problem is that a SONET/SDH BLSR ring does not precross-connect restoration channels. So, the idea of a fully cross-connected restoration loop (ready to be broken into and recross-connected) was never proven to be feasible in an actual SONET/SDH environment. Because of the much more complex mesh network topology of the p-cycle, potential impacts to connections is far riskier and needs significant error handling to establish its resiliency in a real network. Like SONET/SDH rings, a significant operational problem can occur when multiple links fail. To avoid mis-cross-connection and channel contention, the p-cycle idea requires that every possible state of multiple link outage be precalculated, and the desired path and restoration channel usage stored. Furthermore, if a network outage condition occurs that is not represented by one of the precalculated network outage states, bad network conditions or contention could arise. In addition, this requires that a complex optimization problem be run every time a connection is provisioned in the network. See [11.49] and [11.50] for some of the CORONET study optimization results.

To further complicate this situation, we note that links do not necessarily go down in unison. Fibers do not always break cleanly and complex outage states can occur if they go through *staging*, where one link goes down and then another soon follows upon accumulating bit errors. Also, a link might be taken down intentionally for network maintenance. During this maintenance window, if another link outage occurs, the network must transition from the original outage state to a new one (not just jump to a precalculated multiple link outage state). In addition, how to detect when an alarm has *cleared* and how to reset the network afterwards is a very complex problem, even with the BLSR. Thus, when putting all these issues together under a wide set of realistic network conditions and recovery procedures, the CORONET project determined that the p-cycle method is logistically too difficult and risky (or perhaps even infeasible) to implement and it concluded not to pursue p-cycles beyond Phase-1.

To emphasize the difficult problem with precalculating detailed cross-connection responses to potential complex outage states, let us look back at mesh-restoration with the OTN switch. The main (again, alluringly simple) idea is based on storing a fiber-diverse backup path for all working connections. However, if the backup path fails for a set of given connections due to some of the multiple outage states we discussed above, once the link-state messages have settled the nodes revert to a slower dynamic process to calculate and cross-connect new paths on the surviving network for each downed connection. By falling back on a dynamic process for the un-restored connections, the method does not have to rely on precalculated paths that can foul up under complex outages. Even this took almost a decade of lab and in-network evaluation and modification to fully implement and provide confidence that it could reliably handle all network anomalies and corner cases.

11.3.6 Emerging Network Resiliency Architecture and Techniques

Over the many years we have seen the philosophy of network control vacillate between distributed versus centralized control. For example, at the beginning of the Internet (ARPANET), the US government was mostly interested in a network protocol that was rapidly survivable; consequently, distributed control was emphasized. Also, early restoration architectures under centralized control eventually proved too slow, such as [11.34]. Thus, we see the evolution of features of IETF protocols, such as various Interior Gateway Protocols (IGP) and associated protocols. Controls in an IP packet-layer network are mostly high-level *controls*, such as link weights/costs, timers, flooding mechanisms, and distributed topology generation. In contrast, Telecom companies in the past emphasized mostly centralized control via OSSs, and process-oriented procedures, such as circuit provisioning. However, many carriers began to show more interest in distributed mechanisms with philosophies such as *the network is the database* or quasidistributed (i.e., when compared to OSSs) within vendor control structures, such as element management systems, a construct which also arose in ITU-T *telecommunications management network* (*TMN*) concepts [11.51]. However, the tide has recently swung again in favor of centralized network management concepts. In this section, we explore emerging ideas for network resiliency while still mostly concentrating on actual network implementations.

CORONET BoD

The original CORONET Phase 1 requirement was for a restoration mechanism under distributed control, an example of which is shown in Fig. 11.15. However, as previously mentioned, as the project evolved, SDN concepts began to arise and, thus, centralized control had come into vogue again in the Telecom industry. Given the above points and the lack of a strong business case for rapid, distributed wavelength-rate restoration, CORONET Phase-3 changed course and pursued a more commercially fruitful approach to use the emerging SDN capabilities to craft a more attractive business case for bandwidth-on-demand of wavelength-rate services. CORONET Phase-3 demonstrated lab

prototypes with the ability to reroute optical layer connections upon demand using more modern control interfaces, SDN controllers, and associated *network orchestration capability* [11.52]. This also provided a more provable carrier business model for a more dynamically controlled optical layer, which was focused at private networks connecting large datacenters. As we noted previously, much of the connectivity for intermetro-segment private networks is leased, even by the large Internet enterprises. These customers tend to lease dark fiber between their massive core datacenters but then tend to connect their more numerous, peripheral datacenters and customer locations with a mixture of dark fiber and leased lower layer circuits.

The CORONET team found that the most pressing use-case for such a service was datacenter backup at another (distant) datacenter. For, today's large datacenters are so massive that private network overlays with circuits at rates less than the wavelength rate (e.g., less than 100, 400 Gb/s) cannot practically support datacenter backup. The CORONET BoD architecture and experiments enable circuits to be setup as needed at full-wavelength rate for regular backup activities. This created a more focused (in terms of features and capability) and tractable form of wavelength provisioning, rather than the very broad dynamic DWDM layer proposed in CORONET Phase 1. This service is attractive to many of the private-network Enterprise customers compared to more primitive existing procedures for backing up datacenters, such as using portable RAID (redundant array of independent discs) discs that are physically transported to another city. In addition, for some of the major environmental outages described in Fig. 11.8, bandwidth on demand (BoD) was found to be a useful resiliency mechanism to backup datacenters for replication in anticipation of a major environmental event [11.53].

SDN Integration of IP-over-Optical-Layers in Large Carriers

Many carriers are studying how to incorporate SDN ideas into their network management processes. An example can be found in [11.54–56] where an SDN controller is being built for an IP/MPLS network that will eventually work interactively with a more rapid bandwidth-on-demand provisioning process for the DWDM-layer (ROADM) network. The IP/MPLS layer generally follows the *two-phase* resiliency architecture we outlined for intermetro-segment networks in Sect. 11.2.2, wherein MPLS fast reroute (FRR) is used to achieve rapid restoration for their highest priority packet flows following a network outage. Then, after a network settling time, a second phase implements a form of IGP/MPLS reconfiguration where capacity-sensitive traffic engineering is used to reroute the MPLS tunnels to better handle the best-effort traffic and achieve more efficient routing. Network capacity is designed in the network planning process to meet the QoS of each CoS under various outage states. The SDN controller could possibly provide the following added advantages to this process:

- By monitoring the traffic changes more efficiently and providing the opportunity to optimize the routing of the MPLS LSP tunnels, the SDN controller can also be used to better adapt the MPLS tunnels to changing traffic, namely, the *unexpected traffic pattern change* subtree of Fig. 11.8, but also to a lesser degree take advantage of various time-of-day load differences.
- By incorporating the actual optical-layer routing (SRLG) information for each IP/MPLS-layer link, the SDN controller can also compute and download the FRR paths more rapidly and more efficiently than the crude methods current industry router vendors provide to auto-compute them in a distributed manner. For example, vendor methods usually include the ability to set via *command line interface* (*CLI*) or other network management protocol for each packet link a list of *link bundling* IDs. Then the FRR paths are computed in a distributed fashion by the routers to avoid certain paths using the link bundling information. Different vendor products offer other methods to control the FRR paths in a distributed manner, as well.
- As pointed out earlier, most of the extra/spare capacity in carrier IP/Layer router-router links is placed to handle worst-case restoration during a network outage or unexpected traffic pattern change. This amount of extra capacity turns out to be quite high, sometimes as high as 100% more than the peak-hour capacity required to meet QoS when the network is in a nonoutage state. This can be reduced by better capacitating and routing the resulting MPLS tunnels after an outage to more accurately achieve the QoS for each CoS in the second phase after an outage. For example, most of the savings can be achieved by carefully optimizing the best-effort traffic. Without the SDN controller and its associated traffic measurement modules, these tunnels are defined based on average traffic measurements and simplified built-in IETF routing and capacitating algorithms.
- By integrating the IP/MPLS tunnel definition process with ROADM layer provisioning, the SDN controller can utilize the large pool of spare wavelength capacity in the DWDM layer to adapt the underlying capacity of its router-router links. For

example, most intermetro ROAD-to-ROADM links have available 80–100 spare wavelengths upon initial provisioning. These wavelengths/channels are slowly put into service over time, thus enabling on average a large store of spare wavelengths. By reusing the DWDM-layer transponders on an existing lightpath or using preinstalled spare transponders, a rapid lightpath provisioning (or rerouting) process can be implemented using more advanced, or *open* network management interfaces to the ROADMS, such as using a REST interface with an information model based on a Yang protocol.

As an example of the last bullet about integrated IP-over-optical-layer coordination, if a router-router link between routers at CO A and CO D routes through ROADM links A–B, B–C, C–D (where A, B, C, D are COs where ROADMs are installed) and suppose an outage occurs along the fiber path of the B–C link (thus the ROADM B–C link goes down), then by using the same router interfaces and ROADM transponders on the ends of the existing lightpath (A and D), the SDN controller can optically reroute the lightpath on an alternate path of ROADMs between A and D. For example, it can bypass link B–C by reprovisioning the lightpath through alternate ROADM nodes E and F, with path = (A–B, B–E, E–F, F–C, C–D). Assuming no intermediate transponders are required (a nontrivial assumption), then the SDN controller can achieve this simply by disconnecting the A–D existing lightpath at A and D and then reconnecting through the alternate path with basic optical cross-connect commands to each ROADM, which are implemented in the ROADMs by laser tuning at the endpoints, WSS optical cross-connect processes and power balancing (the latter being the more challenging of the tasks).

This of course assumes that spare wavelengths (with the same wavelength/channel number) exist on each link of the alternate route which, as noted above, is quite likely given on average the large number of spare wavelengths per ROADM link. However, it is important to note that from the previous discussion about DARPA CORONET studies, this reprovisioning will not likely occur in subsecond time due to power balancing, laser tuning, and other optical signal quality necessities. However, considering that two-phase IP/MPLS resiliency is applied, for the example outage FRR has already occurred and it is now in the second phase of network reconfiguration. Since the network was planned for fast restoration of the premium CoS traffic, the lower priority traffic *may* be impacted in a worst-case type of outage. In such a case, the QoS expectation for that traffic class assumes it will take longer to restore it to its full throughput. Therefore, a slower reprovisioning process for the affected lightpaths may be tolerable. If so, such an integrated multilayer restoration capability provides the greatest opportunity to save total IP/MPLS-layer capacity.

However, with greater advantage often comes greater complexity. A first step of integrated IP-over-optical-layer network management is to target rapid initial lightpath provisioning for new IP/MPLS-layer links and to avoid OSS cost and complexity. The next step, to enable coordinated interlayer network restoration in a functioning commercial network, is a technically difficult application of SDN and still needs practical development before it reaches maturity in commercial carriers.

Large Enterprise Private Networks

As we noted, there are many large enterprises who own very large datacenters, usually along with a larger number of smaller, *outlying* (or *remote* or *peripheral*) datacenters that home on the larger datacenters. Some of the largest of these enterprises are the Internet-based corporations, for example, those who sell CDN, search capability, video content, web content, enterprise-cloud services, or sell advertising associated with these services. Some of these outlying datacenters can be datacenters owned by (or provided for) their enterprise customers (enterprise customer of the large enterprises). For example, if a large enterprise offers *Internet-as-a-service* (*IaaS*) or *computing-as-a-service* (*CaaS*), or *software-defined wide-area network* (*SD-WAN*) from its portfolio of cloud services, then they may connect to smaller datacenters owned by (or provided for) their cloud customers. The large enterprises tend to deploy large private networks whose major purpose is to transport traffic among their datacenters, called an *inter-datacenter network*. Furthermore, they tend to operate their datacenters themselves, e.g., within a network-supporting business unit.

Given an outlying datacenter in one of the enterprise access networks, there is a slim metro-core component to connect the datacenter to an intermetro network. However, we note that as we discussed in Sect. 11.2.4, to achieve close to the preferred five nines of availability of the private network, the datacenter has to be diversely connected to the intermetro point of presence and in fact is often *dual-homed* into two different points-of-presence (here, a *point-of-presence* is a boundary node between the metro-core- and intermetro-network segments). However, ignoring this metro-core component, most of the interesting aspects of resiliency are in the intermetro segment. Furthermore, almost all such Internet-based corporations lease dark fiber from a large intermetro-segment carrier and install their own lower layer transport equipment, although such equipment is

usually installed, provisioned and maintained by a third party, including sometimes the intermetro carrier itself. However, these corporations also deploy leased, high-bandwidth circuits over a portion of their network, especially to connect the outlying datacenters to their larger, core datacenters. In fact, one can think of the organization within such a large enterprise that runs its interdatacenter network as a type of internally owned carrier. Thus, many of the concepts we discussed earlier will apply to this situation.

Because these large Enterprises usually control and manage both the traffic (typically from their servers), as well as the interdatacenter network, they can customize more forward-looking networking capabilities than a commercial carrier, who must sell connectivity to many different customers. This creates a unique advantage in that such an enterprise can mold and shape its traffic, CoS, and QoS in conjunction with the transport of that traffic. We note, however, that as intriguing as these independent capabilities sound, most of them do not get standardized into the general commercial carrier world.

There is quite a diversity of implementations in the published literature of large enterprise private network implementations. We mention one here to illustrate this point. Google is an example of one such large Internet-based enterprise with truly massive core datacenters scattered around the world (at last count, six in the US alone). Google has adapted some ideas from the *open networking foundation* (*ONF*) [11.57]. In particular, OpenFlow is a centralized method to microcontrol the routing of IP traffic versus the IETF distributed IGP that is implemented in most commercial carriers. Google has furthermore developed its own SDN controller to load-balance its traffic across the large DWDM links that connect its core datacenter gateway switches (routers), as well as to provide alternative routing in case of outages. To illustrate how Google can customize both its traffic and network using traffic engineering ideas for CoS, at a high level it roughly classifies its traffic as *delay-sensitive* and *delay-insensitive*, the latter being an extremely large amount of traffic. The delay-insensitive traffic, such as information gathered from background Internet search-bots, can be queued in case of network congestion, such as would occur from a network outage. This keeps the network utilization very high and cost efficient. In contrast, commercial carriers cannot implement this type of tight CoS to control customer traffic in coordination with generic network control of network routing and performance. Furthermore, the Google SDN controller can execute extremely rapid and time-sensitive load balancing across links. These capabilities enable very rapid restoration with minimal packet loss, perhaps even better than MPLS Fast Reroute, but with more efficient implementation. See the following sources for more detail [11.58–61].

There are major differences among how these large Internet-based Enterprises manage both their large, core datacenters and their interdatacenter private networks. Above is just one example. See [11.62] for another example. Therefore, it is doubtful that any of these customized implementations for large Enterprise networks will carry over in a general way to large commercial carriers.

RENs and ON Labs

In this chapter, we have focused mostly on implementation of resiliency in commercial carriers. In contrast, Research and Education Networks are networks where more experimentation and forward-looking ideas can be trialed and evaluated in a network setting. RENs are typically fully or partially subsidized by various government entities, for example by direct subsidization or study grants. They usually also involve heavy collaboration with (or sponsorship by) academic institutions. However, RENs can serve multiple functions, such as a testbed for experiments in advanced networking, as well as small carriers for working, day-to-day traffic among universities, government agencies, and some private research groups. This latter function provides a more realistic environment to experiment and evaluate with network resiliency ideas than a controlled lab. But, because of their wide research emphasis, covering the myriad of restoration and resiliency ideas across hundreds of world-wide RENs is beyond our scope.

ESnet is an example of a larger REN, a network mostly funded by the US Department of Energy. ESnet has installed an intermetro network with many 100-Gb/s wavelength-rate links that spans from coast-to-coast. Part of the task of this REN is not only to provide a sandbox to try out new networking ideas, but also to provide very high rate connections that allow bandwidth-hungry science experiments to exchange and process voluminous data that could either be technically impossible or unaffordable via a purely commercial carrier. Examples are high-speed collider experiments, neural networks for brains, and assembling and synthesizing distant astronomical views by coalescing data from many different telescopes around the world. Furthermore, RENs from different countries often link via high-speed links (including undersea). For example, ESnet partners (peers) with GEANT (pan-European backbone R&E network), NORDUnet (Nordic countries REN), CANARIE (Canada), and LHCONE (an overlay REN in Europe, North America,

and Southeast Asia). A key point is that ESnet has an SDN testbed that implements OpenFlow [11.63]. OpenFlow is very popular in Academia because it enables detailed control of packet routing. As we have seen, routing is integral to network resiliency. Furthermore, as we illustrated in Sect. 11.3.6, *SDN Integration of IP-over-Optical-Layers in Large Carriers*, the ability to implement integrated control over multiple network layers can be beneficial to network restoration and potentially enable more efficient use of packet-layer and DWDM-layer capacity. In fact, as we pointed out numerous times in this chapter, literally every resiliency technique basically involves some attempt to reroute traffic/connections around bad things.

We finally note that along with centralization becoming more *in vogue* in the academic world and in some private networks, the idea of using open networking software has also been heavily advertised, especially in the area of *open platforms*. In fact, use of open software, software infrastructure, and more flexible computing environments is also being vigorously pursued in large carriers. See [11.64] for an example of open networking in ROADMs. That said, we have seen many networking ideas come and (mostly) go over many decades in the Telecom industry (see more about this in the last section). To what effect these new ideas stand the test of time in commercial networks is for future historical assessment.

11.4 Summary and Future Directions

We help to summarize our chapter via a brief question-and-answer format.

11.4.1 Summary

We first attempted to answer the principal question *how are today's terrestrial carrier-based networks architected?* We then provided a high-level view of the three major network segments, along with the technology layers within each segment. The next major question was: *how is resiliency provided in each layer today?* To answer that question, we went into a detailed analysis of the types of QoS degradations that tend to confound networks and their services, including a class of degradations called *outages*. With those degradations established, we then described the various types of remediation provided at each segment and layer. At that point, someone unfamiliar with the evolution of the carriers, business climate, and telecommunication technology, would likely be mystified by the *end state* of the seemingly odd combination of network technologies, network sectoring, and restoration/resiliency methodologies. Therefore, to help clarify this we asked the next question: *what factors led to the current evolution of network layers, resiliency, and resulting network performance and availability?* Potential answers would be even more befuddling if the reader sought answers from the publication world in these related fields. For, very few of the published ideas (many of which are derived in theoretical/abstract network settings) have been implemented in commercial, carrier networks. To understand why this is so, we examined a few restoration methodologies that have been implemented or attempted to be implemented and provided insight into how difficult it is to successfully implement these ideas in a large, commercial carrier. Furthermore, one cannot explain this evolution via telecommunications technologies, protocols, and restoration ideas alone. In that regard, we delved into the business side of commercial networks, especially the metro-access segment, where business, government, and regulatory considerations and constraints dominate the networking architectures that have evolved.

To further help clarify the above answers, we introduced another important question: *how do the current evolved network architectures, restoration methodologies, and network management practices impact network performance and availability on an* end-to-end *service basis*. We answered this by carefully examining how overall network availability is impacted by the network layers, segments and services (e.g., business versus residential) and how *the weak link* is so critical to understanding the business side of this long evolution.

11.4.2 Key Enablers for Future Work

The Telecommunications industry and academia has many efforts underway for improving network reliability and resiliency in optical and related networks. In fact, we could devote an entire section/chapter to discuss all these different directions. In the spirit of this chapter, we will instead funnel this down to a few key, practical directions, which are dominated by work in advanced network control and management. However, we caution that the network management field has in the past hailed how it was going to achieve improved network performance and cost via advanced network management capabilities, only to become awash with modeling complexity, layers of protocols, and bureaucracy. For example, ITU-T TMN (Sect. 11.3.6) CMISE/CMIP

is an excellent example of such a failure of expectation [11.65]. To illustrate this over-expectation, we quote from [11.66]:

> Many benefits are expected from standards-based integrated network management. By making all internal network management functions available through standardized interfaces, service providers will achieve more rapid deployment of new services and maximum use of automated functions. Powerful new service management capabilities will be offered to customers by the interconnection of service provider and user management infrastructures. Vendors of network elements will offer specialized systems known as *element managers* which will integrate readily into a service provider's larger management hierarchy. Groups of service providers will enter into 'business level agreements' and deploy resource sharing arrangements which will be administered automatically through interoperable interfaces.

In all fairness, element management systems did emerge over this period, but failed to achieve the standardization or advantages claimed by CMISE/CMIP modeling. In fact, some claim that the advent of element management systems stalled the evolution of advanced network management modeling that is needed today. There are other examples, such as CORBA (common object request broker architecture) and distributed network management, that also did not achieve the stated goals or where the industry changed directions. Furthermore, element management systems and standardized network/object modeling for packet-layer switches did not manifest over this period. For example, many systems still use SNMP (simple network management protocol) MIBs (management information bases) for alarm reporting with limited configuration management capabilities, which are far from achieving the objectives of the above quote. Still today most hard configurations (*configs*) for packet-layer switches are set via vendor-proprietary CLI (Sect. 11.3.6).

With this dubious network management background in mind and to make future directions more meaningful, we will concentrate on the historically persistent problem areas in network management that must be improved before the plethora of future efforts can make substantive progress in commercial carrier networks.

11.4.3 Streaming Analytics, Cognitive Systems, Machine Learning

One of the major areas of pursuit is faster and more organized reporting of network metrics, followed by faster and more powerful network state analysis, and then followed by more rapid and automatic network remediation. In this regard, when referring to new protocols, network updates and network management techniques, *Najam Ahmad*, VP of Network Engineering at Facebook said that it used to take them a year to test and then deploy new network capabilities for their datacenter networks and interdatacenter networks [11.67]. The business logic was that this was needed to prevent something from going wrong because it is so hard to detect, analyze, and then remedy network outages and dysfunctionalities. Furthermore, even when one does so much extensive pretesting and evaluation against potential outages and problems, is it possible to predict all the possible problems in a fully operational network? Clearly, this previous approach imposed a large cost to do such long evaluations, plus it delayed important network reliability and/or new features. However, Ahmad claimed recently that now they move to deploy much more rapidly because Facebook is moving more towards rapid outage detection and root cause analysis. And by *rapid*, he emphasized the ability of the network management software to automatically detect and react to such outages and then make automatic network corrections and restoration. This is in contrast to past practice where personnel in the NOC manually collect data, perform root cause analysis and then override the network settings to remediate a service-disrupting event. This new approach gives the NOC and developers time to instead concentrate on making needed network, software, and product updates.

Of course, this strategy sounds great in theory, but it is not so easy to accomplish. One of the perennial impediments has been the classical alarm reporting and NOC methodologies of filtering through voluminous and mysterious syslogs and alarm reporting to filter down to critical information, let alone figuring out how to remedy it. As one can see from the details of this chapter, given the layering complexity of networks, figuring out the source of a problem from narrowly focused equipment logs makes this task difficult, if not impossible. Older alarm-reporting protocols such as switch SNMP traps, ITU OTN termination point models (or even SONET/SDH-based) alarm reporting have proven inadequate to accomplish the above goal. We instead need a more sophisticated network modeling capability and advanced alarm reporting protocol. Work on *streaming analytics and cognitive systems* [11.68] has the goal of modernizing the reporting of network state in complex networks. For example, some in this field are pursuing network models and protocols such as OpenConfig to better represent the different network states in modern network equipment, along with more straightforward and flexible object-oriented data structures and interfaces that can better represent the relationships of network objects. Furthermore, this ap-

proach supports better focused network metrics to accomplish rapid root cause analysis.

To enhance this approach, machine learning and AI combine with improved data structures for network metrics reporting to more rapidly pin down network/service disruptions and then automatically recommend remediation. Stated more simply, the intent of this new approach is to put more investment and *muscle* into analysis and remediation and less into costly preventative measures and overbuilding of the network. We have already described in the previous section how advanced network management and control systems, based on approaches such as SDN, can be used to aid in some of the remedial phase after the enabling of advanced analytics. A more recent reference on how to possibly use machine learning to better perform root cause analysis can be found in [11.69].

11.4.4 Network Discovery

One of the key failures of the earlier approaches was abstract (generic, nonspecific) network modeling, protocols, and data models; but, they missed the inside knowledge of the real complex interactive structures of networks, restoration methods, outage characteristics, and traffic, such as we presented earlier. As one can discern after reading this chapter, it is difficult to accomplish automatic root cause analysis and automatic remediation in such networks without having a good, detailed model of their layered structure. For, even the most sophisticated AI or machine-learning algorithms will find it impossible to discern the needed detail and inter-relationships about these network constructions, unless these specific relationships are correctly and properly modeled. The missing ingredient today is automatic discovery of these complex, layered network relationships and restoration mechanisms, as illustrated in Fig. 11.3. A principal impediment in large commercial carriers has been that the lower layer relationships must be discerned from operations support systems and network capacity provisioning data. A further complicating factor is that much of this data is replete with errors and omissions. For example, the pointers between higher layer links and lower layer connections are often incomplete or missing. See [11.11] for a discussion of some of these issues. We note that although this reference is a bit dated compared to the present chapter, unfortunately much of this situation persists in large carriers (perhaps less in the private networks of some large Internet service enterprises). We conclude that having spent a large part of our career developing network discovery algorithms, the field of network discovery (including discovery of interlayer relationships) needs to significantly evolve before the future directions described above can evolve fully.

11.4.5 Predictive Network Availability and Performance

Finally, we briefly mention the historical problem area of predictive network availability and performance. As we have stated, assessing the availability and performance of a network under the onslaught of potential QoS degradations is more art than science. In almost all practical cases, networks are fine-tuned to meet network performance objectives via experience, i.e., reactive network planning and management. The largest impediments to doing this more predictively are the areas of automatic network discovery and better alarm reporting and isolation, as described above, plus the ability to better model the effects of network restoration. As previously discussed in Sect. 11.2.1, an example of how to put all these complex pieces together to perform predictive network availability is found in [11.20]. We add that this methodology (applied to a totally different study network than presented in [11.20]) was used to accurately predict network availability for extending the national backbone IPTV network of a commercial carrier to a new area of the United States. The results were used to fine-tune the network topology (both IP and optical layers) and their restoration methods to achieve the predicted network availability. To reenforce our previous points about the need to advance the fields of network discovery and streaming analytics, the network topology data for this study was gathered manually from provisioning information and historical statistics. For example, mean time to repair (MTTR) and mean time between failures (MTBF) were gathered from carrier outage equipment historical data and vendor FIT rates (Sect. 11.2.1). That is, since automatic network discovery and streaming analytics were not available and given our comments about the difficulty of acquiring lower network-layer data and correlations, this aspect of the study required intense modeling work. However, given that such advanced capabilities might eventually exist, the reader will find future predictive approaches in [11.69].

Finally, harkening back to our initial comments from the chapter introduction, the results of the above study remain unpublished. To help the reader understand why carriers keep these sorts of studies proprietary, consider if they were published: some people (not necessarily friendly to the carrier) might claim the study results are a guarantee of QoS for the given service. Given the potential differences that we described between theoretical QoS predictions versus actual experience, this might lead to law suits, negative publicity, and even frivolous claims of IPR (intellectual property rights) infringement, which are quite prevalent in any large carrier.

References

11.1 R. Doverspike, P. Magill: Chap. 13: commercial optical networks, overlay networks and services. In: *Optical Fiber Telecommunications V B* (Elsevier, Amsterdam 2008)

11.2 R. Doverspike, K.K. Ramakrishnan, C. Chase: Structural overview of commercial long distance IP networks. In: *Guide to Reliable Internet Services and Applications*, ed. by C. Kalmanek, S. Misra, R. Yang (Springer, Berlin, Heidelberg 2010)

11.3 R. Doverspike, J. Reich: Automated planning and provisioning for carrier metro networks. In: *Conf. Design of Reliable Communications Networks (DRCN), Kansas City*, Vol. 2015 (2015), keynote talk

11.4 International Telecommunication Union: ITU-T G.984.1: Gigabit-capable Passive Optical Networks (GPON): General characteristics, http://www.itu.int/rec/T-REC-G.984.1/en (2009)

11.5 New York Times: Google curbs expansion of fiber optic network, cutting jobs, https://www.nytimes.com/2016/10/26/technology/google-curbs-expansion-of-fiber-optic-network-cutting-jobs.html (2016)

11.6 S. Hanatani: Overview of global FTTH market and state-of-the-art technologies. In: *18th Opto-Electronics and Communications Conference held jointly with 2013 International Conference on Photonics in Switching, Opt. Soc. Am.* (2013), paper WP4_1

11.7 Cable Labs: DOCSIS® 3.1 Technology, https://www.cablelabs.com/technologies/docsis-3-1 (2019)

11.8 R. Doverspike: Tutorial. The practice of network modeling and design in commercial telecommunications networkp. In: *GERAD Optimization Days* (2013)

11.9 F. Cugini, N. Sambo, F. Paolucci, F. Fresi, P. Castoldi: Adaptation and monitoring for elastic alien wavelengths. In: *21st Eur. Conf. Netw. Opt. Commun.* (2016) pp. 82–87

11.10 Businesswiere.com: Ciena announces CoreDirector FS for evolving networks to ethernet and OTN, http://www.businesswire.com/news/home/20090928005154/en/Ciena-Announces-CoreDirector-FS-Evolving-Networks-Ethernet (2009)

11.11 R. Doverspike, J. Yates: Optical network management and control, Proc. IEEE **100**(5), 1092–1104 (2012)

11.12 CenturyLink: CenturyLink completes acquisition of Level 3, http://news.centurylink.com/2017-11-01-CenturyLink-completes-acquisition-of-Level-3 (2017)

11.13 Internet Engineering Task Force: IETF RFC 3209: RSVP-TE: Extensions to RSVP for LSP Tunnels, https://tools.ietf.org/html/rfc3209 (2001)

11.14 Internet Engineering Task Force: IETF RFC 3630: Traffic Engineering (TE) extensions to OSPF Version 2, https://tools.ietf.org/html/rfc3630 (2003)

11.15 Internet Engineering Task Force: IETF Extensions to RFC 5305: IS-IS Traffic Engineering (TE) metric extensions, https://tools.ietf.org/html/draft-ietf-isis-te-metric-extensions-11 (2016)

11.16 J. Reich, S. Woodward, A.N. Patel, B.G. Bathula, W. Zhang, A. Chiu: Leveraging SDN to streamline metro network operations, IEEE Commun. Mag **54**(10), 109–115 (2016)

11.17 M. Nogueira, P. Cholda, D. Medhi, R. Doverspike: Disaster resilience in communication networks: part 1, IEEE Commun. Mag. **52**(10), 44–45 (2014)

11.18 M. Nogueira, P. Cholda, D. Medhi, R. Doverspike: Disaster resilience in communication networks: part 2, IEEE Commun. Mag. **52**(12), 56–57 (2014)

11.19 Bloomberg News: Harvey knockout of cell service revives talk of backup power, https://www.bloomberg.com/news/articles/2017-08-30/harvey-s-knockout-of-cell-service-revives-calls-for-backup-power (2017)

11.20 K. Oikonomou, R. Sinha, R. Doverspike: Multi-layer network performance and reliability analysis, Int. J. Interdiscipl. Telecommun. Netw. **1**(3), 1–29 (2009)

11.21 Internet Engineering Task Force: IETF RFC 4090: Fast reroute extensions to RSVP-TE for LSP Tunnel, http://www.ietf.org/rfc/rfc4090.txt (2005)

11.22 J. Yates, Z. Ge: Chap. 12: Network management: fault management, performance management, and planned maintenance. In: *Guide to Reliable Internet Services and Applications* (Springer, Cham 2010)

11.23 R. Doverspike, B. Cortez: Restoration in Carrier Networks. In: *Proc. 7th Int. Workshop, Design of Reliable Communication Networks, Washington* (2009)

11.24 Internet Engineering Task Force: IETF RFC 5880: Bidirectional Forwarding Detection (BFD), https://tools.ietf.org/html/rfc5880 (2010)

11.25 X. Zhang, M. Birk, A. Chiu, R. Doverspike, M.D. Feuer, P. Magill, E. Mavrogiorgis, J. Pastor, S. Woodward, J. Yates: Bridge-and-roll demonstration in GRIPhoN (globally reconfigurable intelligent photonic network). In: *Opt. Netw. Commun. Conf. OFC/NFOEC, San Diego* (2010)

11.26 Wikipedia: Synchronous optical networking, https://en.wikipedia.org/wiki/Synchronous_optical_networking (2017)

11.27 International Telecommunication Union: ITU-T G.8032/Y.1344 Amendment 1: Ethernet ring protection switching amendment 1, https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.8032-201611-I!Amd1!PDF-E&type=items (2016)

11.28 Wikipedia: Ethernet ring protection switching, https://en.wikipedia.org/wiki/Ethernet_Ring_Protection_Switching (2017)

11.29 Institute of Electrical and Electronics Engineers: IEEE 802.17: Resilient Packet Ring Working Group, http://www.ieee802.org/17/ (2011)

11.30 Internet Engineering Task Force: IETF RFC 6824: TCP Extensions for multipath operation with multiple addresses, https://tools.ietf.org/html/rfc6824 (2013)

11.31 S. Toshikazu: Resilient network architecture based on Movable and Deployable Resource Unit (MDRU). In: *ITU Workshop on Disaster Relief Systems, Network Resiliency and Recovery, Phuket*, https://www.itu.int/en/ITU-T/Workshops-and-Seminars/

drnrr/201305/Documents/S3P3_Toshikazu_Sakano.pdf (2013)
11.32 M. Ruffini, I. Monga, J. Wey: Connected OFCITY challenge 2018: lighting up the emerging world. In: *Opt. Netw. Commun. Conf. 2018, San Diego* (2018)
11.33 AT&T: AT&T Network continuity overview, https://www.corp.att.com/ndr/pdf/cpi_5181.pdf (2015)
11.34 C.-W. Chao, H. Eslambolchi, P. Dollard, L. Nguyen, J. Weythman: FASTAR – A robust system for fast DS-3 restoration. In: *Proc. GLOBECOM'91, Phoenix* (1991) pp. 1396–1400
11.35 G. Ash: *Dynamic Routing in Telecommunications Networks* (McGraw-Hill, New York 1997)
11.36 R. Doverspike, S. Phillips, J. Westbrook: Future transport network architectures, IEEE Commun.Mag. **37**(8), 96–101 (1999)
11.37 R. Doverspike: Algorithms for multiplex bundling in a telecommunications network, Oper. Res. **39**(6), 925–944 (1991)
11.38 EE Times: Ciena spends nearly $1 billion for Omnia, Lightera, https://www.eetimes.com/document.asp?doc_id=1138897 (1999)
11.39 DARPA Project BAA 06-29: Proposer Information Pamphlet (PIP) for Defense Advanced Research Projects Agency (DARPA) Strategic Technology Office (STO) Dynamic Multi-Terabit Core Optical Networks: Architecture, Protocols, Control and Management (CORONET), http://www.monarchna.com/CORONET-InfoPamphlet-DARPA.pdf (2006)
11.40 A. Saleh, J.M. Simmons: Evolution toward the next-generation core optical network, J. Lightwave Technol. **24**, 3303 (2006)
11.41 J. Simmons: *Optical Network Design and Planning*, 2nd edn. (Springer, Cham 2014)
11.42 A. Chiu, G. Choudhury, G. Clapp, R. Doverspike, J.W. Gannett, J.G. Klincewicz, G. Li, R.A. Skoog, J. Strand, A. Von Lehmen, D. Xu: Network design and architectures for highly dynamic next-generation IP-over-optical long distance networks, J. Lightwave Technol. **27**(12), 1878–1890 (2009)
11.43 A. Chiu, G. Choudhury, G. Clapp, R. Doverspike, M. Feuer, J.W. Gannett, J. Jackel, G.T. Kim, J.G. Klincewicz, T.J. Kwon, G. Li, P. Magill, J.M. Simmons, R.A. Skoog, J. Strand, A. Von Lehmen, B.J. Wilson, S.L. Woodward, D. Xu: Architectures and protocols for capacity-efficient, highly-dynamic and highly-resilient core networks, J. Opt. Commun. Netw. **4**(1), 1–14 (2012)
11.44 A. Von Lehmen, R. Doverspike, G. Clapp, D. Freimuth, J. Gannett, A. Kolarov, H. Kobrinski, C. Makaya, D. Mavrogiorgis, J. Pastor, M. Rauch, K. Ramakrishnan, R. Skoog, B. Wilson, S. Woodward: CORONET: Testbeds, demonstration, and lessons learned (invited), J. Opt. Commun. Netw. **7**, A447–A458 (2015)
11.45 R. Doverspike, G. Sahin, J. Strand, R. Tkach: Fast restoration in a mesh network of optical cross-connects. In: *Proc. OFC 1999, San Diego* (1999)
11.46 R. Doverspike, G. Sahin, J. Strand, R. Tkach: Methods and systems for fast restoration in a mesh network of optical cross connects, US Patents 6970417 and 7088676 (2005–2006)
11.47 W. Grover, D. Stamatelakis: Cycle-oriented distributed preconfiguration: ring-like speed with mesh-like capacity for self-planning network restoration. In: *Int. Conf. Commun.* (1998) pp. 537–543
11.48 G. Shen, G. Grover: Extending the p-cycle concept to path-segment protection for span and node failure recovery, IEEE J. Sel. Areas Commun. **21**(8), 1306–1319 (2003)
11.49 M. Eiger, H. Luss, D. Shallcross: Network restoration under a single link or node failure using preconfigured virtual cycles, Telecommun. Syst. **46**(1), 17–30 (2011)
11.50 M. Eiger, H. Luss, D. Shallcross: Network restoration under dual failures using path-protecting preconfigured cycles, Telecommun. Syst. **49**(3), 271–286 (2012)
11.51 International Telecommunication Union: ITU-T M.3010: Principles for a telecommunications management network, https://www.itu.int/rec/T-REC-M.3010 (2015)
11.52 R. Doverspike, G. Clapp, P. Douyon, D.M. Freimuth, K. Gullapalli, B. Han, J. Hartley, A. Mahimkar, E. Mavrogiorgis, J. O'Connor, J. Pastor, K.K. Ramakrishnan, M.E. Rauch, M. Stadler, A. Von Lehmen, B. Wilson, S.L. Woodward: Using SDN technology to enable cost-effective bandwidth-on-demand for cloud services, J. Opt. Commun. Netw. **7**(2), A326–A334 (2015)
11.53 Fiercetelecom.com: AT&T, ACS, IBM test elastic cloud-to-cloud networking concept, https://www.fiercetelecom.com/telecom/at-t-acs-ibm-test-elastic-cloud-to-cloud-networking-concept (2014)
11.54 M. Birk, G. Choudhury, B. Cortez, A. Goddard, N. Padi, A. Raghuram, K. Tse, S. Tse, A. Wallace, K. Xi: SDN use cases for service provider networks: Evolving to an SDN-enabled ISP backbone: Key technologies and applications, IEEE Commun. Mag. **54**(10), 129–135 (2016)
11.55 G. Choudhury: Software Defined Networks to greatly improve the efficiency and flexibility of Packet IP and Optical Networks. In: *Proc. Int. Conf. Comput. Netw. Commun. ICNC 2017, Santa Clara* (2017)
11.56 R. Savoor, K. Meier-Hellstern: Chap. 16: Network measurements. In: *Building the Network of the Future: Getting Smarter, Faster, and More Flexible with a Software Centric Approach (100 Cases)* (CRC, Boca Raton 2017)
11.57 Open Networking Foundation (ONF): https://www.opennetworking.org/
11.58 Google: Inter-Datacenter WAN with centralized TE using SDN and OpenFlow, https://www.opennetworking.org/wp-content/uploads/2013/02/cs-googlesdn.pdf (2012)
11.59 B. Koley: Software defined networking at scale. LightReading Big Telecom Event (BTE), Chicago, https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/42948.pdf (2014)

11.60 C. Lam: Optical network technologies for datacenter networks. In: *Opt. Netw. Commun. Conf. 2010, San Diego* (2010)

11.61 X. Zhao, V. Vusirikala, B. Koley, V. Kamalov, T. Hofmeister: The prospect of inter-data-center optical networks, IEEE Commun. Mag. **51**(9), 32–38 (2013)

11.62 A. Greenberg: SDN for the cloud, SIGCOMM, London, http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/keynote.pdf (2015)

11.63 ESNet: 100G SDN Testbed, https://www.es.net/network-r-and-d/experimental-network-testbeds/100g-sdn-testbed/ (2019)

11.64 Open ROADM MSA: Open ROADM MSA, http://openroadm.org (2016)

11.65 U. Black: *Network Management Standards SNMP, CMIP, TMN, MIBs, and Object Libraries* (McGraw-Hill, New York 1995)

11.66 J.K. Shrewsbury: An introduction to TMN, J. Netw. Syst. Manag. **3**(1), 13–38 (1995)

11.67 N. Ahmad: Network Operators. In: *Proc. OFC 2018, San Diego* (2018), summit keynote

11.68 L. Paraschis: Symposium on network management evolution to streaming analytics and cognitive systems. In: *Opt. Netw. Commun. Conf. 2018, San Diego* (2018)

11.69 T. Panayiotou: Leveraging statistical machine learning to address failure localization in optical networks, J. Opt. Commun. Netw. **10**(3), 162–173 (2018)

Robert D. Doverspike

Network Evolution Strategies, LLC
Tinton Falls, NJ, USA
r.d.doverspikeconsulting@gmail.com

Robert Doverspike has a PhD from Rensselaer Polytechnic Institute. He has worked at Bell Labs, Bellcore-Telcordia, been Executive Director AT&T Network Evolution Research, and founded Network Evolution Strategies (consulting with Google). He is an INFORMS Fellow, an IEEE Fellow, and the OFC Program/General Chair. He has 35 patents, is a world-renowned expert on network restoration and optimization, and has helped design some of the world's largest networks.

12. Routing and Wavelength (Spectrum) Assignment

Jane M. Simmons, George N. Rouskas

Routing a connection from its source to its destination is a fundamental component of network design. The choice of route affects numerous properties of a connection, most notably cost, latency, and availability, as well as the resulting level of congestion in the network. This chapter addresses various algorithms, strategies, and tradeoffs related to routing.

At the physical optical layer, connections are assigned a unique wavelength on a particular optical fiber, a process known as wavelength assignment (WA). Together with routing, the combination of these two processes is commonly referred to as RWA. In networks based on all-optical technology, WA can be challenging. It becomes more so when the physical properties of the optical signal need to be considered. This chapter covers several WA algorithms and strategies that have produced efficient designs in practical networks.

A recent development in the evolution of optical networks is flexible networking, where the amount of spectrum allocated to a connection can be variable. Spectrum assignment is analogous to, though more complex than, wavelength assignment; various heuristics have been proposed as covered in this chapter. Flexible (or elastic) networks are prone to more contention issues as compared to traditional optical networks. To maintain a high degree of capacity efficiency, it is likely that spectral defragmentation will be needed in these networks; several design choices are discussed.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_12

Routing a connection (or circuit) from its source to its destination is a fundamental component of network design. The choice of route affects numerous properties of a connection, most notably cost, latency, and availability. Furthermore, from a more macro perspective, routing determines the resulting level of congestion in a network, which may impact the amount of traffic that ultimately can be carried. This chapter addresses various algorithms, strategies, and tradeoffs related to routing. While much of the discourse applies to networks in general, the emphasis is on routing in the physical optical layer. Relevant network terminology and some of the fundamentals of optical systems are introduced in Sect. 12.1.

Network routing is a well-researched topic. Several commonly used routing algorithms have been developed to find an optimal path through the network with respect to a particular property. These algorithms are reviewed in Sect. 12.2. For a connection of relatively low criticality, it is sufficient to calculate a single path from source to destination. However, a failure that occurs along that path will bring the connection down. To guard against such failures, more critical connections are established with some level of protection. Some commonly used protection schemes rely on calculating two or more diverse paths between the source and destination. Optimal disjoint-path routing algorithms exist for this purpose as well, as covered in Sect. 12.3.

As noted above, the routing strategy that is utilized affects the congestion level in the network. Poor routing may lead to unbalanced load in the network, with some sections heavily utilized whereas others are lightly loaded. Such imbalances result in premature blocking in the network where it may not be possible to accommodate a new connection request due to insufficient capacity in portions of the network. Various routing strategies and their effect on network load are discussed in Sect. 12.4.

If multiple connections are being routed at once, then the order in which they are processed may also affect the blocking level. For example, it may be desirable to give precedence to geographically longer connections or connections that must traverse congested areas of the network, as it is likely more challenging to find a suitable path for these connections. Selecting the routing order is, in general, an arbitrary procedure; with rapid design runtimes, several orderings can be tested and the best result selected. Routing order is covered in Sect. 12.5.

Historically, connections in telecommunications networks have been unicast, with a single source and a single destination. However, with services such as video distribution growing in importance, connections directed from one source to multiple destinations are desirable. Establishing a multicast tree from the source to the set of destinations is more capacity efficient than utilizing multiple unicast connections. Furthermore, the optical layer is well suited for delivering multicast services. Section 12.6 discusses routing algorithms that find the optimal, or near-optimal, multicast tree.

The end-to-end path is the high-level view of how a connection is carried in the network. When examined more closely at the physical optical layer, it is necessary to assign a portion of the electromagnetic spectrum to carry the connection at each point along its path. Optical networks typically employ wavelength-division multiplexing (WDM) technology, where connections are assigned a unique wavelength on a particular optical fiber and the various wavelengths are multiplexed together into a single WDM signal on that fiber. Selecting the wavelengths to assign to each connection is the process known as *wavelength assignment*; this topic is introduced in Sect. 12.7. The combination of the routing and wavelength assignment processes is commonly referred to as RWA.

The underlying optical system technology determines the level of difficulty involved with wavelength assignment. With one class of technology, known as optical-electrical-optical (O-E-O), where connections repeatedly go from the optical domain to the electrical domain and back, the wavelengths assigned to connections on one link have no impact on the wavelengths that can be assigned on any other link. With such technology, wavelength assignment is a simple process. O-E-O technology, however, does have numerous drawbacks, as is outlined in Sect. 12.1.

In contrast, optical-optical-optical (O-O-O), or *all-optical*, technology potentially enables connections to remain in the optical domain from source to destination. While there are many advantages to this technology, it does lead to challenges in the wavelength assignment process. In practice, most networks fall somewhere in between pure O-E-O and O-O-O, where connections enter the electrical domain at intermediate points of the path, but do so infrequently. It is common practice to refer to these networks as all-optical, despite this term being a misnomer; this chapter makes use of this naming convention as well. The respective properties of O-E-O and all-optical networks drive the routing and wavelength assignment processes and are discussed further in Sect. 12.1.

Poor wavelength assignment in all-optical networks can lead to blocking, where the capacity exists to carry a new connection but no wavelength is free to assign to that connection. This is known as *wavelength contention*. Various wavelength assignment heuristics that are simple to implement yet effective in minimizing

wavelength contention have been developed, as outlined in Sect. 12.8.

Another important design decision is whether routing and wavelength assignment should be treated as separate processes or whether they should be combined into a single operation. This dichotomy is discussed in Sect. 12.9.

In an all-optical network, there are numerous practical challenges of wavelength assignment that are tied to the physical optical system. Some of the more important effects are introduced in Sect. 12.10.

A recent development in the evolution of optical networks is *flexible networking*, introduced in Sect. 12.11. In traditional optical networks, there is a fixed number of wavelengths that can be carried on each optical fiber, and each such wavelength occupies a fixed amount of spectrum. In contrast, with flexible (or elastic) networks, the amount of spectrum allocated to a connection can be variable, potentially leading to a more efficient use of the fiber capacity. Selecting the portion of the spectrum to assign to a given connection is known as *spectrum assignment*. As elucidated in Sect. 12.12, spectrum assignment is analogous to, though more complex than, wavelength assignment. Various heuristics have been proposed to specifically address the nuances of spectrum assignment.

Flexible networks tend to lead to more contention issues as compared to traditional optical networks. This is especially true if the network is dynamic, such that connections are frequently established and then torn down. Dynamism is expected to result in fragmented spectrum utilization where many small blocks of unutilized spectrum exist as opposed to large contiguous blocks of free spectrum. To maintain a high degree of capacity efficiency, it is likely that *spectral defragmentation* would need to be performed on a periodic basis, where connections are shifted to different regions of the spectrum (and possibly rerouted) to create larger chunks of available spectrum. Defragmentation techniques are discussed in Sect. 12.12.

Network design can be performed on different time scales. In *long-term* network planning, there is sufficient time between the planning and provisioning processes such that any additional equipment required by the plan can be deployed. The planning emphasis is on determining the optimal strategy for accommodating a set of demands; the runtime of the design algorithms is not critical. In *real-time* operation, there is little time between planning and provisioning. It is assumed that the traffic must be accommodated using whatever equipment is already deployed in the network, even if that sacrifices optimality. Any design calculations must be performed quickly (milliseconds to minutes time scale, depending on the requirements of the network). The long-term and real-time dichotomy is also referred to in the literature as *offline* and *online* design, respectively. Methodologies for both time scales are discussed in this chapter.

A number of algorithms are introduced in this chapter. Implementation of many of these algorithms, using the C programming language, can be found in [12.1].

12.1 Terminology

Network terminology that is relevant to the topics of this chapter is defined here. Additionally, a high-level comparison of O-E-O and all-optical networks is provided, with an emphasis on the aspects that affect routing and wavelength assignment.

Network nodes are the sites in the network that source, terminate, and/or switch traffic. Sites that serve only to amplify the optical signal are not considered nodes. Network nodes are depicted as circles in the figures contained here. *Network links* are the physical optical fibers that run between the nodes, with the fibers typically deployed in pairs. Links are almost always bidirectional, with one fiber of the pair carrying traffic in one direction and the other fiber carrying traffic in the opposite direction. Network links are depicted in figures as a single line. In discussions on routing, links are often referred to as *hops*.

The interconnection pattern among the nodes represents the *network topology*. The *nodal degree* is the number of fiber-pairs incident on a node. The nodal degree may also be defined as the number of links incident on a node. A link may be populated with multiple fiber-pairs, such that the two definitions are not always equivalent.

In a network that employs O-E-O technology (i.e., an O-E-O network), the connection signal is carried in the optical domain on the links but is converted to the electrical domain at each node that is traversed by the connection. An end-to-end path looks as depicted in Fig. 12.1a. The process of O-E-O conversion at each node cleans up the signal; more precisely, the signal is re-amplified, re-shaped, and re-timed. This is known as *3R regeneration*. (Note that passing a signal through a line optical-amplifier is considered *1R regen-*

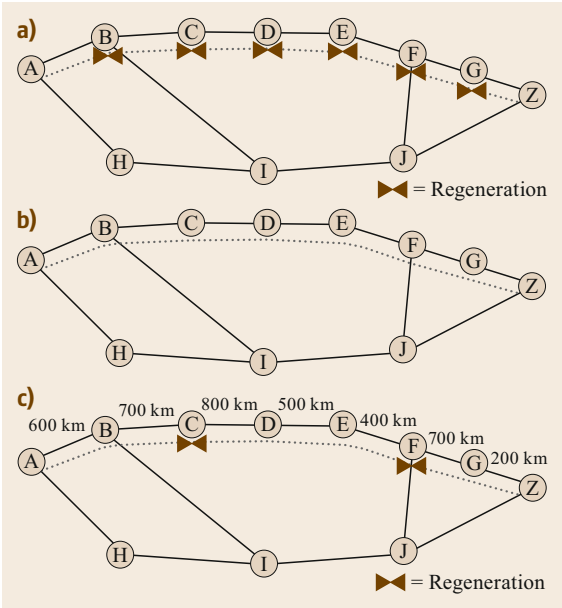


Fig. 12.1a–c The connection path from node A to node Z is shown by the *dotted line*. **(a)** In an O-E-O network, regeneration occurs at each node along the path of the connection. **(b)** In a true all-optical network, there is no regeneration along the path. The connection remains in the optical domain from end to end. **(c)** With an optical reach of 2000 km and the link distances as shown, two regenerations are required for the connection. (© Monarch Network Architects LLC)

eration, as it only re-amplifies the signal.) Regeneration equipment is needed for each individual signal that traverses a node; i.e., the WDM signal is demultiplexed into its constituent wavelengths and each one of the wavelengths that is routed through the node is passed through a regenerator. (Terminal equipment, as opposed to a regenerator, is used for a wavelength that is destined for the node as opposed to being routed *through* the node; a regenerator often comprises back-to-back terminal equipment, as discussed further in Sect. 12.7.1.) Thus, a network node may be equipped with hundreds of regenerators. (Limited *multiwavelength regeneration* has been demonstrated but is still in the early stages of research.)

The large amount of equipment required for regeneration poses drawbacks in terms of cost, power consumption, reliability, and space requirements. In O-E-O networks, regeneration is the major contributor to these factors in the optical layer. Minimizing the number of nodes that a signal must traverse is desirable, as it minimizes the number of regenerations.

Conversely, a true all-optical signal that remains in the optical domain is illustrated in Fig. 12.1b. No regenerations are present along the end-to-end path. The network nodes are equipped with elements such as *reconfigurable optical add/drop multiplexers* (ROADMs) that allow the signal to remain in the optical domain. This is known as *optical bypass* of the node.

An important property of all-optical networks is the *optical reach*, which is the distance that an optical signal can travel before it needs to be regenerated. (In practice, there are numerous factors other than distance that affect where regeneration is required [12.1]. For simplicity, we only consider distance here.) Most optical-bypass-enabled systems have an optical reach on the order of 1500–3000 km. If the length of the end-to-end path is longer than the optical reach, then the signal needs to be regenerated, possibly more than once, as illustrated in Fig. 12.1c. This is more often the case in core networks (also known as backbone networks, long-haul networks, or cross-country networks), where some of the end-to-end paths may be a few thousand kilometers in extent.

Regeneration is almost always performed in a network node as opposed to at an intermediate point along a link. Thus, regeneration typically occurs prior to the signal traveling exactly a distance equal to the optical reach. This may lead to more regeneration being required than predicted by the connection distance. For example, in Fig. 12.1c, the total end-to-end path distance is 3900 km, but with an optical reach of 2000 km, two regenerations are required, not one. Furthermore, a study performed on several realistic backbone networks indicated that the shortest path was not the path with the fewest number of required regenerations for roughly 1% of the source/destination pairs in those networks [12.1].

Taking advantage of extended optical reach and ROADMs to remove all, or at least most, of the required regeneration provides significant advantages in terms of cost, power consumption, reliability, and space. The drawback is that if a signal traverses a node in the optical domain, then it must be carried on the same wavelength into and out of the node. This is known as the *wavelength continuity constraint*. (Wavelength conversion in the optical domain is possible; however, it is costly and complex and has not been widely commercialized.) Thus, the presence of optical bypass creates an interdependence among the links of a network, where the wavelength assigned to a connection on one link affects the wavelengths that can be assigned to connections on other links. This leads to wavelength assignment being a critical aspect of network design in all-optical networks. Note that such an interdependence does not exist in O-E-O networks where regeneration at every node along a path allows the wavelengths to be assigned independently on each link.

12.2 Shortest-Path Routing Algorithms

We first consider algorithms that find a single path that is the shortest between a given source and destination [12.2]. If multiple paths are tied for the shortest, then one of the paths is found. Depending on the algorithm used, ties may be broken differently.

The class of shortest-path algorithms discussed here assumes that each link is assigned a cost metric and that the metric is additive. With an additive metric, the metric for a path equals the sum of the metrics of each link composing that path. Furthermore, there can be no cycles in the network where the sum of the link metrics around the cycle is negative. With these assumptions, a shortest-path algorithm finds the path from source to destination that minimizes the total cost metric.

These algorithms can be applied whether or not the links in the network are bidirectional. A link is bidirectional if traffic can be routed in either direction over the link. Different metrics can be assigned to the two directions of a bidirectional link. However, if the network is bidirectionally symmetric, such that the traffic flow is always two-way and such that the cost metric is the same for the two directions of a link, then a shortest path from source to destination also represents, in reverse, a shortest path from destination to source. In this scenario, which is typical of telecommunications networks, it does not matter which endpoint is designated as the source and which is designated as the destination.

There are numerous metrics that are useful for network routing. If the link metric is the geographic distance of the link, then the shortest-path algorithm does, indeed, find the path with the shortest distance. If all links are assigned a metric of unity (or any constant positive number), then the shortest-path algorithm finds the path with the fewest hops. Using the negative of the logarithm of the link availability as the metric produces the most reliable path, assuming that the failure rate of a path is dominated by independent link failures. (Note that the logarithm function can be used in general to convert a multiplicative metric to an additive metric.) Other useful metrics for routing in an all-optical network are the link noise figure and the link optical signal-to-noise ratio (OSNR) [12.1].

In spite of the link metric not necessarily being associated with distance, these algorithms are commonly still referred to as shortest-path algorithms.

12.2.1 Dijkstra Algorithm

The best known shortest-path algorithm is the Dijkstra algorithm [12.2]. The algorithm works by tracking the shortest path discovered thus far to a particular node, starting with a path length of zero for the source node. It considers the resulting length if the path is extended to a neighboring node. If the path to the neighboring node is shorter than any previously discovered paths to that node, then the path to that neighbor is updated. Dijkstra is classified as a greedy algorithm because it makes the optimal decision at each step without looking ahead to the final outcome. Unlike many greedy algorithms, it is guaranteed to find the optimal solution; i.e., the shortest path from source to destination.

12.2.2 Breadth-First-Search Algorithm

An alternative shortest-path algorithm is breadth-first search (BFS) [12.3]. It works by discovering nodes that are one hop away from the source, then the nodes that are two hops away from the source, then the nodes that are three hops away from the source, etc., until the destination is reached.

As with Dijkstra, BFS produces the shortest path from source to destination. Additionally, if there are multiple paths that are tied for the shortest, BFS breaks the tie by finding the one that has the fewest number of hops. This can be useful in all-optical networks where the likelihood of wavelength contention increases with the number of hops in a path. Note that Dijkstra does not have a similar tie-breaking property; its tie-breaking is somewhat more arbitrary (e.g., it may depend on the order in which the link information is stored in the database).

12.2.3 Constrained Shortest-Path Routing

A variation of the shortest-path problem arises when one or more constraints are placed on the desired path; this is known as the *constrained shortest path* (CSP) problem. Some constraints are straightforward to handle. For example, if one is searching for the shortest path subject to all links of the path having at least N wavelengths free, then prior to running a shortest-path algorithm, all links with fewer than N free wavelengths are removed from the topology. As another example, the intermediate steps of the BFS shortest-path algorithm can be readily used to determine the shortest path subject to the number of path hops being less than H, for any $H > 0$ (similar to [12.4]). However, more generally, the CSP problem can be difficult to solve; for example, determining the shortest path subject to the availability of the path being greater than some threshold, where the availability is based on factors other than distance. Various heuristics have been proposed to address the CSP problem [12.5]. Some heuristics have been developed

to specifically address the scenario where there is just a single constraint; this is known as the *restricted shortest path* (RSP) problem. Additionally, a simpler version of the multiconstraint problem arises when *any* path satisfying all of the constraints is desired, not necessarily the shortest path; this is known as the *multiconstrained path* (MCP) problem. An overview, including a performance comparison, of various heuristics that address the RSP and MCP problems can be found in [12.6].

12.2.4 *K*-Shortest-Paths Routing Algorithms

Routing all connection requests between a particular source and destination over the shortest path can lead to network congestion, as is discussed more fully in Sect. 12.4. It is generally advantageous to have a set of paths to choose from for each source/destination pair. A class of routing algorithms that is useful for finding a set of K possible paths between a given source and destination is the *K-shortest-paths algorithm*. Such an algorithm finds the shortest path, the second shortest path, the third shortest path, etc., up until the K-th shortest path or until no more paths exist. A commonly used K-shortest-paths algorithm is Yen's algorithm [12.7]. As with shortest-path algorithms, the link metrics must be additive along the path.

It is important to note that the K paths that are found are not necessarily completely diverse with respect to the links that are traversed; i.e., some links may appear in more than one of the K paths. K-shortest-paths routing is typically used for purposes of load balancing, not for protection against failures. Explicitly routing for failure protection is covered in Sect. 12.3.

12.2.5 Shortest-Distance Versus Minimum-Hop Routing

As stated earlier, a variety of link metrics can be used in a shortest-path algorithm. Two of the most common metrics in a telecommunications network are link distance and unity (i.e., *1*), producing the path of shortest geographic distance and the path of fewest hops, respectively. The most effective metric to use depends on the underlying network technology.

With O-E-O networks, the cost of a connection in the optical layer is dominated by the number of required regenerations. Regeneration occurs at every node that is traversed by the connection. Thus, utilizing *1* as the metric for all links in order to find the path with the least number of hops (and hence the fewest traversed nodes) is advantageous from a cost perspective. This is illustrated in Fig. 12.2 for a connection between nodes A and Z. Path 1 is the shortest-distance path at

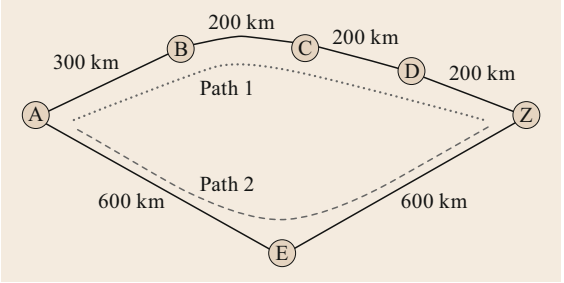


Fig. 12.2 Path 1, A–B–C–D–Z, is the shortest-distance path between nodes A and Z, but Path 2, A–E–Z, is the fewest-hops path. In an O-E-O network, where the signal is regenerated at every intermediate node, Path 2 is typically the lower-cost path (© Monarch Network Architects LLC)

900 km but includes four hops. Path 2, though it has a distance of 1200 km, is typically lower cost in an O-E-O network because it has only two hops and, thus, requires fewer regenerations.

Selecting a metric to use with all-optical networks is not as straightforward. Assuming that the number of regenerations required for a connection is dominated by the path distance, then searching for the shortest-distance path will typically minimize the number of regenerations. However, wavelength assignment for a connection becomes more challenging as the number of path hops increases. For all-optical networks, one effective strategy is to generate candidate paths by invoking a K-shortest-paths algorithm twice, once with distance as the link metric and once with *1* as the link metric. Assuming that link load is not a concern, the most critical factor in selecting one of the candidate

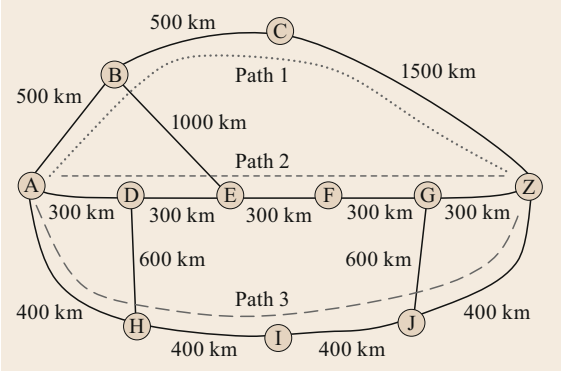


Fig. 12.3 Assume that this is an all-optical network with an optical reach of 2000 km. Path 1, A–B–C–Z, has the fewest hops but requires one regeneration. Path 2, A–D–E–F–G–Z, and Path 3, A–H–I–J–Z, require no regeneration. Of these two lowest-cost paths, Path 3 is preferred because it has fewer hops (© Monarch Network Architects LLC)

paths is typically cost. Thus, the candidate path that requires the fewest number of regenerations is selected. If there are multiple paths tied for the fewest regenerations, then the one with the fewest number of hops is selected.

This strategy is illustrated in Fig. 12.3. Three of the possible paths between nodes A and Z are: Path 1: A–B–C–Z of distance 2500 km and three hops; Path 2: A–D–E–F–G–Z of distance 1500 km and five hops; and Path 3: A–H–I–J–Z of distance 1600 km and four hops. Assume that the optical reach is 2000 km, such that there can be no all-optical segment that is longer than this distance. With this assumption, Path 1 requires one regeneration, whereas Paths 2 and 3 do not require any regeneration. Of the latter two paths, Path 3 has fewer hops and is, thus, more desirable despite it being longer than Path 2 (again, assuming that link load is not a factor).

12.3 Disjoint-Path Routing for Protection

Network customers desire a certain level of *availability* for each of their connections. Availability is defined as the probability of being in a working state at a given instant of time. The desired availability is typically specified contractually in the *service level agreement* (SLA) with the network provider. A common cause of connection failure is the failure of one or more links in the end-to-end path. Link failures are typically caused by fiber cuts or optical amplifier failures. Being able to route around the failed link allows the connection to be restored to the working state.

In some protection schemes, it is necessary to first identify which link has failed; the detour route around that failed link is then utilized. Determining the location of a failure can be time consuming. To expedite the restoration process and improve availability, network providers often utilize protection schemes where the same backup path is utilized regardless of where the failure has occurred in the original (i.e., primary) path. This allows the protection process to commence prior to the completion of the fault location process. In order to implement such a failure-independent protection scheme, also known as *path protection*, it is necessary that the backup path be completely link-disjoint from the primary path. If a very high level of availability is desired for a connection, then it may be necessary to protect against node failures in addition to link failures. In this scenario, the backup path must be completely link and node disjoint from the primary path. However, it should be noted that node failures occur much less frequently than link failures, such that node disjointness is often not required.

Shortest-path algorithms find the single shortest path from source to destination. With path protection, it is desirable to find two disjoint paths where the overall sum of the link metrics on the two paths is minimized. One of the paths is used as the primary path (typically the shorter one) with the other serving as the backup path.

It may seem reasonable to find the desired two paths by invoking a shortest-path algorithm twice. After the first invocation, the links (and intermediate nodes if node disjointness is also required) composing the first path are removed from the network topology that is

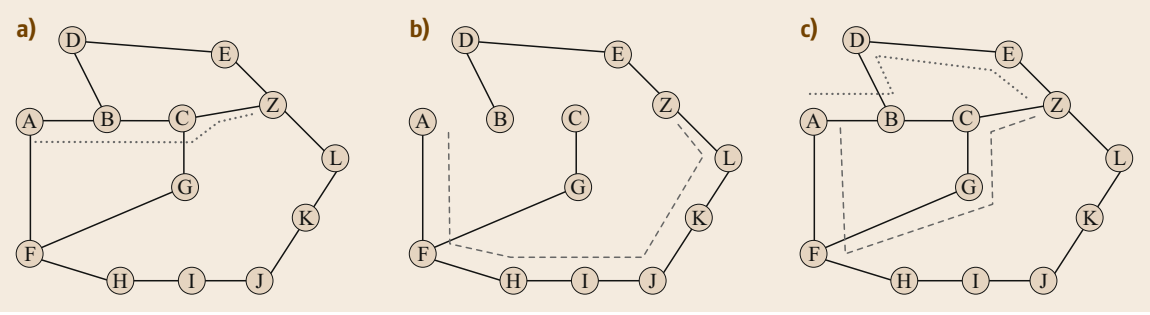


Fig. 12.4a–c The pair of disjoint paths with the fewest number of total hops is desired between nodes A and Z. **(a)** The first call to the shortest-path algorithm returns the path shown by the *dotted line*. **(b)** The network topology after pruning the links composing the shortest path. The second call to the shortest-path algorithm finds the path indicated by the *dashed line*. The total number of hops in the two paths is ten. **(c)** The shortest pair of disjoint paths between nodes A and Z is shown by the *dotted* and *dashed lines*; the total number of hops in these two paths is only eight (© Monarch Network Architects LLC)

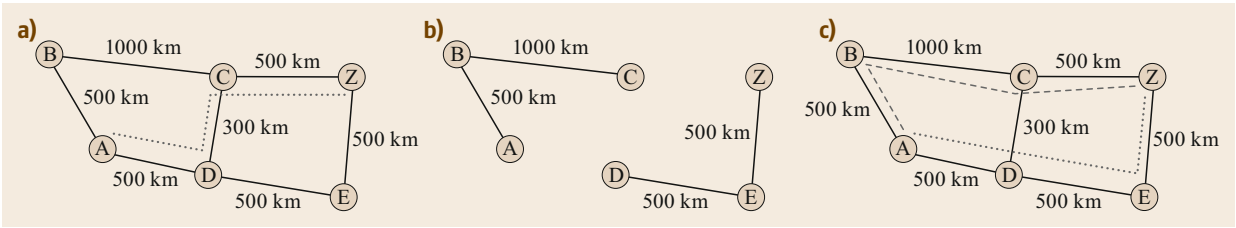


Fig. 12.5a–c The shortest pair of disjoint paths is desired between nodes A and Z. **(a)** The first call to the shortest-path algorithm returns the path shown by the *dotted line*. **(b)** The network topology after pruning the links composing the shortest path. The second call to the shortest-path algorithm fails, as no path exists between nodes A and Z in this pruned topology. **(c)** The shortest pair of disjoint paths between nodes A and Z, shown by the *dotted* and *dashed lines* (© Monarch Network Architects LLC)

used in the second call of the algorithm. Clearly, if a path is found by the second invocation, then that path is guaranteed to be disjoint from the first path that was found.

Although this strategy may appear to be reasonable, there are potential pitfalls. First, it inherently assumes that the shortest disjoint pair of paths includes the shortest path. As shown in Fig. 12.4, this is not always the case. In this example, the source is node A and the destination is node Z. Assume that this is an O-E-O network, with all links assigned a metric of *1*. The shortest path (i.e., the fewest-hops path) is A–B–C–Z, as shown in Fig. 12.4a. If the links of this path are removed and the shortest-path algorithm is invoked again, then the resulting path is A–F–H–I–J–K–L–Z; see Fig. 12.4b. The sum of the metrics of these two paths is 10. However, the shortest disjoint pair of paths is actually A–B–D–E–Z and A–F–G–C–Z, as illustrated in Fig. 12.4c; this pair of paths has a total metric of 8. Note that the shortest path, A–B–C–Z, is not part of the optimal solution.

Not only may the two-invocation method produce a suboptimal result, it may fail completely. This is illustrated in Fig. 12.5. Assume that this is an all-optical network, and the metric is the physical distance as shown next to each link. In this figure, the shortest path from A to Z is A–D–C–Z, as shown in Fig. 12.5a. If the links of this path are removed from the topology, as shown in Fig. 12.5b, then no other paths exist between nodes A and Z. Thus, the process fails to find a disjoint pair of paths despite the existence of such a pair, as shown in Fig. 12.5c: A–B–C–Z and A–D–E–Z. This type of scenario, where the two-invocation methodology fails despite the existence of disjoint paths, is called a *trap topology*.

Rather than using the two-invocation method, it is recommended that an algorithm specifically designed to find disjoint paths be used. The two most commonly used shortest-disjoint-paths algorithms are the Suurballe algorithm [12.8, 9] and the Bhandari algorithm [12.3]. Both algorithms make use of a shortest-path algorithm; however, extensive graph transformations are performed as well to ensure that the shortest pair of disjoint paths is found, assuming such a pair exists. Both algorithms require that the link metric be additive and both algorithms are guaranteed to produce the optimal result. They can be utilized to find the shortest pair of link-disjoint paths or the shortest pair of link-and-node-disjoint paths. The runtimes of the Suurballe and Bhandari algorithms are about the same; however, the latter may be more easily adapted to various network routing applications. We focus on its use here.

The Bhandari algorithm is readily extensible. For example, a mission-critical connection may require a high level of availability such that three disjoint paths are needed, one primary with two backups. If a failure occurs on the primary path, the first backup path is used. If a failure occurs on the backup path prior to the primary path being repaired, then the connection is moved to the second backup path. The Bhandari algorithm can be used to find the shortest set of three disjoint paths, if they exist. More generally, the Bhandari algorithm can be used to find the shortest set of N disjoint paths, for any N, assuming N such paths exist. (In most optical networks, however, there are rarely more than just a small number of disjoint paths between a given source and destination, especially in a backbone network.)

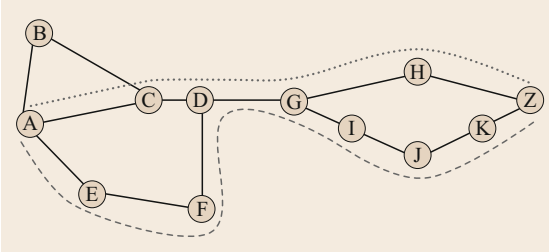


Fig. 12.6 There is no completely disjoint pair of paths between nodes A and Z. The set of paths shown by the *dotted* and *dashed lines* represents the shortest maximally-disjoint pair of paths. The paths have nodes D and G, and the link between them, in common (© Monarch Network Architects LLC)

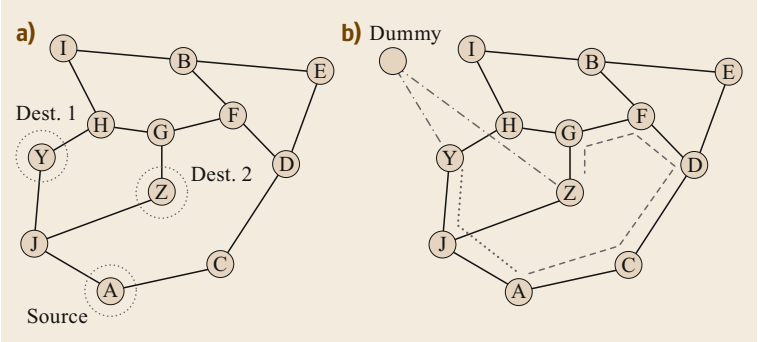


Fig. 12.7 **(a)** A disjoint path is desired between one source (node A) and two destinations (nodes Y and Z). **(b)** A dummy node is added to the topology and connected to the two destinations via links that are assigned a metric of zero. A shortest-disjoint-paths algorithm is run between node A and the dummy node to implicitly generate the desired disjoint paths, as shown by the *dotted line* and the *dashed line* (© Monarch Network Architects LLC)

In some scenarios, completely disjoint paths between the source and destination do not exist. In such cases, the Bhandari algorithm can be utilized to find the shortest *maximally-disjoint* set of paths. The resulting paths minimize the number of links (and optionally nodes) that are common to multiple paths. This is illustrated in Fig. 12.6, where the shortest maximally-disjoint pair of paths is A–C–D–G–H–Z and A–E–F–D–G–I–J–K–Z. Note that link DG is common to both paths.

12.3.1 Disjoint-Path Routing with Multiple Sources and/or Destinations

An important variation of the shortest-disjoint-paths routing problem exists when there is more than one source and/or destination, and each of the source/destination paths must be mutually disjoint for protection purposes. (Note that this is different from multicast routing, where the goal is to create a tree from one source to multiple destinations.) This protection scenario arises, for example, when backhauling traffic to multiple sites, utilizing redundant data centers in cloud computing, and routing through multiple gateways in a multidomain topology [12.1]. The example of Fig. 12.7a is used to illustrate the basic strategy. We assume that this is a backhauling example, where node A backhauls its traffic to two diverse sites, Y and Z. (Backhauling refers to the general process of transporting traffic from a minor site to a major site for further distribution.) Additionally, it is required that the paths from A to these two sites be disjoint.

In order to apply the shortest-disjoint-paths algorithm, a dummy node is added to the network topology as shown in Fig. 12.7b. Links are added from both Y and Z to the dummy node, and these links are assigned a metric of 0. The shortest-disjoint-paths algorithm is then run with node A as the source and the dummy node as the destination. This implicitly finds the desired disjoint paths as shown in the figure.

A similar strategy is followed if there are D possible destinations, with $D > 2$, and disjoint paths are required from the source to M of the destinations, with $M \leq D$. Each of the D destinations is connected to a dummy node through a link of metric 0. An extensible shortest-disjoint-paths algorithm such as the Bhandari algorithm is invoked between the source and the dummy node to find the desired M disjoint paths. Note that this procedure implicitly selects M of the D destinations that produce the shortest such set of disjoint paths.

If the scenario is such that there are multiple sources and one destination, then the dummy node is connected to each of the sources via links with a metric of 0. The shortest-disjoint-paths algorithm is then run between the dummy node and the destination. If there are both multiple sources and multiple destinations, then two dummy nodes are added, one connected to the sources and one connected to the destinations. The shortest-disjoint-paths algorithm is then run between the two dummy nodes. The algorithm does not allow control over which source/destination combinations will result.

12.3.2 Shared-Risk Link Groups

One challenge of routing in practical networks is that the high-level network topology may not reveal interdependencies among the links. Consider the network topology shown in Fig. 12.8a and assume that it is desired to find the shortest pair of disjoint paths from node A to node Z. From this figure, it appears that the paths A–B–Z and A–Z are the optimal solution. However, the fiber-level depiction of the network, shown in Fig. 12.8b, indicates that links AB and AZ are not fully

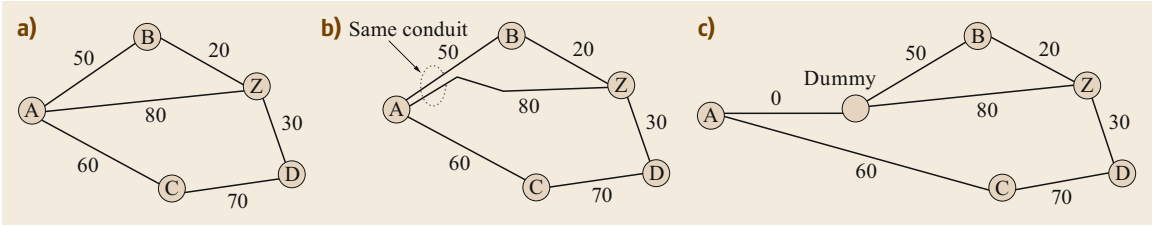


Fig. 12.8 **(a)** In the link-level view of the topology, links AB and AZ appear to be disjoint. **(b)** In the fiber-level view, these two links lie in the same conduit exiting node A and are, thus, not fully diverse. A single cut to this section of conduit can cause both links to fail. **(c)** Graph transformation to account for the SRLG extending from node A (© Monarch Network Architects LLC)

disjoint. The corresponding fibers partially lie in the same conduit exiting node A such that a rupture to the conduit likely causes both links to fail. These two links are said to constitute a *shared-risk link group* (SRLG). Links that are a member of the same SRLG are not fully disjoint. It is desirable to avoid a solution where an SRLG link is in the primary path and another link in that same SRLG is in the backup path.

For the SRLG configuration depicted in Fig. 12.8b, it is straightforward to find truly disjoint paths. First, a graph transformation is performed as shown in Fig. 12.8c, where a dummy node is added and each link belonging to the SRLG is modified to have this dummy node as its endpoint instead of node A. A link of metric 0 is added between node A and the dummy node. The shortest-disjoint-paths algorithm is run on the modified topology to find the desired solution: paths A–B–Z and A–C–D–Z. (While it is not necessary in this example, for this type of SRLG graph transformation, the shortest-disjoint-paths algorithm should be run with both link and node disjointness required [12.1].)

The scenario of Fig. 12.8 is one class of SRLG, known as the *fork configuration*. There are several other SRLG configurations that appear in practical networks [12.1, 3]. There are no known computationally efficient algorithms that are guaranteed to find the optimal pair of disjoint paths in the presence of any type of SRLG. In some cases, it may be necessary to employ heuristic algorithms that are not guaranteed to find the optimal (shortest) set of disjoint paths [12.10].

12.4 Routing Strategies

Monitoring network load is an essential aspect of network design. If too much traffic is carried on a small subset of the links, it may result in an unnecessarily high blocking rate, where future connections cannot be accommodated despite the presence of free capacity on most network links. The routing strategy clearly affects the network load. We discuss three of the most common routing strategies here.

12.4.1 Fixed-Path Routing

The simplest routing strategy is known as *fixed-path routing*. One path is calculated for each source/destination pair, and that path is utilized for every connection request between those two nodes. If any link along that path has reached its capacity (e.g., 80 connections are already routed on a link that supports 80 wavelengths), then further requests between the source/destination pair are blocked.

This strategy is very simple to implement, as it requires no calculations to be performed on an ongoing basis; all calculations are performed up front. Routing is a simple binary decision: either all links of the precalculated path have available capacity for the new connection, or they do not.

The drawback to fixed-path routing is that it is completely nonadaptive. The same path is selected regardless of the load levels of the links along that path. Such load-blind routing typically leads to uneven load distribution among the links, which ultimately may lead to premature blocking.

Despite this drawback, fixed-path routing is employed by many (if not most) network service providers, where the shortest-distance path is the one path selected for each source/destination pair. The rationale for this strategy is that it minimizes the latency of the connection. (Latency is the propagation delay from source to destination; in the optical layer, latency is typically dominated by the distance of the path.) While there are some network customers where latency differences on the order of microseconds can be critical (e.g., enterprises involved with electronic financial trad-

ing [12.11]), for most applications, a tolerance of a few extra milliseconds is acceptable. In most realistic networks, it is possible to find alternative paths that are very close in distance to the shortest path such that the increase in latency is acceptably small. By not considering such alternative paths, fixed-path routing is overly restrictive and can lead to poor blocking performance.

12.4.2 Alternative-Path Routing

A second routing strategy is *alternative-path routing*. In contrast to fixed-path routing, a set of candidate paths is calculated for each source/destination pair. At the time of a connection request, one of the candidate paths is selected, typically based on the current load in the network. For example, the candidate path that results in the lowest load on the most heavily loaded link of the path is selected. This enables adaptability to the current load levels in the network.

It is well known that alternative-path routing can lead to lower blocking levels as compared to fixed-path routing [12.12, 13]. One study performed by a service provider on its own large backbone network demonstrated one to two orders of magnitude lower blocking probability using alternative-path routing, even when limiting the length of the alternative paths for purposes of latency [12.14].

Furthermore, a large number of candidate paths is not required to achieve significant blocking improvements. Generating on the order of three candidate paths is often sufficient to provide effective load balancing. The key to the success of this strategy lies in the choice of the candidate paths. There is typically a relatively small set of links in a network that can be considered the *hot links*. These are the links that are likely to be in high demand and, thus, likely to become congested. An effective strategy to determine the expected hot links is to perform a design for a typical traffic profile where all connections in the profile follow the shortest path. The links that are the most heavily loaded in this exercise are generally the hot links.

With knowledge of the likely hot links, the candidate paths for a given source/destination pair should be selected such that the paths exhibit diversity with respect to these links. It is not desirable to have a particular hot link be included in all of the candidate paths, if possible, as this will not provide the opportunity to avoid that link in the routing process. Note that there is no requirement that the candidate paths for a given source/destination be *completely* diverse. While total path disjointness is necessary for protection, a less severe diversity requirement is sufficient for purposes of load balancing [12.15].

For completeness, we mention one variation of alternative-path routing known as *fixed alternative-path routing*. In this scheme, a set of candidate paths is generated and ordered. For each connection request, the first of the candidate paths (based on the fixed ordering) that has free capacity to accommodate the request is selected. Thus, load balancing is implemented only after one or more links in the network are full, thereby providing limited benefit as compared to fixed-path routing.

12.4.3 Dynamic Routing

With dynamic routing, no routes are precalculated. Rather, when a connection request is received by the network, a search is performed for a path at that time. Typically, path selection is based on cost and/or load. For example, consider assigning each link in an O-E-O network a value of LARGE $+ L_j$, where LARGE is a very large constant and L_j is a metric that reflects the current load level on link j. (Any load-related metric can be used, as long as the metric is additive.) Running a shortest-path algorithm with these link values will place the first priority on minimizing the number of hops in the path, due to the dominance of the LARGE component. In an O-E-O network, minimizing the number of hops also minimizes the number of regenerations required. The second priority is selecting a path of minimal load, as defined by the metric. Once the network contains several full or close-to-full links, the priorities may shift such that each link j is assigned a metric of simply L_j. With this assignment, the path of minimal load is selected regardless of the number of hops.

Dynamic routing clearly provides the most opportunity for the path selection process to adapt to the current state of the network. This may appear to be the optimal routing strategy; however, there are important ramifications in an all-optical network. By allowing any path to be selected as opposed to limiting the set of candidate paths to a small set of precalculated paths, connections between a given source/destination pair will tend to follow different paths. This has the effect of decreasing the network *interference length*, which can potentially lead to more contention in the wavelength assignment process. The interference length is the average number of links shared by two paths that have at least one link in common [12.16]. In addition, if the all-optical network makes use of wavebands, where groups of wavelengths are treated as a single unit, then the diversity of paths produced by a purely dynamic strategy can be detrimental from the viewpoint of efficiently packing the wavebands. Note that wavebands may be required with some of the technology that has been proposed for

achieving large fiber capacity in future networks, as is discussed in Sect. 12.12.5 [12.17, 18].

Another drawback to dynamic routing is the additional delay inherent in the scheme. This is especially true in a centralized architecture where any route request must be directed to a single network entity (e.g., a *path computation element* (PCE) [12.19, 20]). Delays incurred due to communication to and from the PCE, as well as possible queuing delays within the PCE, may add tens of milliseconds to the route calculation process. For dynamic applications that require very fast connection setup, this additional delay may be untenable.

To summarize this section on routing strategies, studies have shown that alternative-path routing provides a good compromise between adaptability and calculation time, and is more amenable to wavelength assignment in an all-optical network. If all candidate paths between a source/destination pair are blocked, then dynamic routing can be invoked.

12.5 Routing Order

In real-time network operation, connection requests are processed as soon as they are received. Thus, the routing process typically consists of searching for a path for just one connection (if batched routing is employed, then there may be a small number of connections to be routed at one time). In contrast, with long-term network design, routing is likely to be performed on a large traffic matrix comprising hundreds of connections. The order in which the connections are routed may have a significant impact on the ultimate level of network loading and blocking.

An effective ordering approach is to give routing priority to the source/destination pairs for which finding a path is likely to be more problematic. Thus, factors such as the relative locations of the two endpoints, whether or not a disjoint protection path is required for the connection, and whether or not the candidate paths for the connection contain a large number of *hot links* should be considered when assigning the routing order. For example, finding a path for a source/destination pair that is located at opposite ends of a network is likely to be more difficult than finding a path between an adjacent source and destination. It is typically advantageous to route such a connection early in the process when the links are relatively lightly loaded to minimize the constraints. Similarly, requiring disjoint paths for purposes of protection is already a significant constraint on the possible number of suitable paths. Giving priority to routing connections that require protection may enhance the likelihood that a feasible pair of disjoint paths can be found.

One enhancement of this strategy is to combine it with a metaheuristic such as *simulated annealing* [12.21]. A baseline solution is first generated based on the priority ordering method described above. The ordering that generated the baseline result is passed to the simulated annealing process. In each step of simulated annealing, the orderings of two connections are swapped, and the routing process rerun. If the result is better (e.g., less blocking or lower levels of resulting load), this new ordering is accepted. If the result is worse, the new ordering is accepted with some probability (this allows the process to extricate itself from local minima). The probability threshold becomes lower as the simulated annealing process progresses. If simulated annealing is run long enough, the overall results are likely to improve as compared to the baseline solution.

Other ordering schemes can be found in [12.1]. With rapid design runtimes, several orderings can be tested and the best result selected.

12.6 Multicast Routing

Video distribution is one of the major drivers of traffic growth in networks. Such services are characterized by a single source delivering the same traffic to a set of destinations. One means of delivering these services is to establish multiple unicast connections between the source and each of the destinations. Alternatively, a single multicast tree that includes each of the destinations can be established [12.22]. These two options are illustrated in Fig. 12.9. The multicast tree eliminates duplicate routing on various links of the network (e.g., link QR in Fig. 12.9) and is thus more capacity efficient. One study in a realistic backbone network showed that multicast provides a factor of roughly two to three improvement in capacity as compared to multiple unicast connections, where capacity was measured as the average number of wavelengths required on a link [12.1].

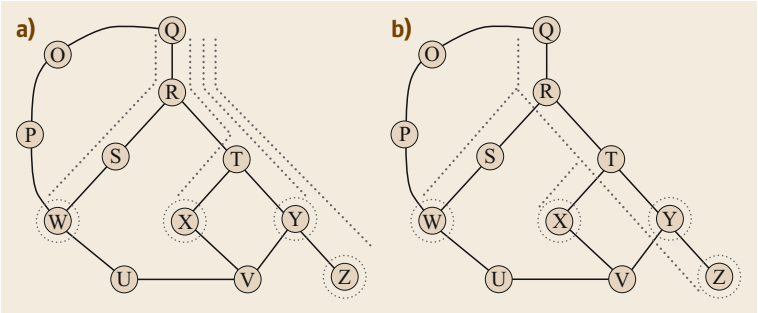


Fig. 12.9 **(a)** Four unicast connections are established between the source Q and the destinations W, X, Y and Z. **(b)** One multicast connection is established between node Q and the four destinations (© Monarch Network Architects LLC)

A tree that interconnects the source to all of the destinations is known as a *Steiner tree* (where it is assumed that all links are bidirectionally symmetric; i.e., two-way links with the same metric in both directions). The *weight* of the tree is the sum of the metrics of all links that compose the tree. As with point-to-point connections, it is often desirable to find the routing solution that minimizes the weight of the tree. Finding the Steiner tree of minimum weight is, in general, a difficult problem to solve unless the source is *broadcasting* to every node in the network. However, several heuristics exist to find good approximate solutions to the problem [12.23]. Two such heuristics are *minimum spanning tree with enhancement* (MSTE) [12.24, 25] and *minimum paths* (MP) [12.26]. Examples of these heuristics, along with their C-code implementation, can be found in [12.1]. Based on studies utilizing realistic backbone networks, MP tends to produce better results, although the relative performance of MSTE improves as the number of destinations increases [12.1]. In the case of broadcast, an algorithm such as Prim's or Kruskal's can be used to optimally find the tree of minimum weight [12.2].

The multicast algorithms enumerated above provide a single path from the source to any of the destinations. If a failure occurs along that path, connectivity with that destination is lost. Furthermore, due to the tree topology, a link failure along the tree often disconnects several destinations. Providing protection for a multicast tree, where connectivity with all destinations is maintained regardless of any single link failure, can be cumbersome and may require a large amount of protection resources.

Various strategies have been devised for providing multicast protection [12.27–29]. One approach common to several of these strategies is to make use of segment-based protection, where the multicast tree is conceptually partitioned into segments, and each segment is protected separately. A very different concept that has been applied to multicast protection is *network coding* [12.30–32]. With this approach, which has applications beyond just multicast protection, the destinations receive independent, typically linear, combinations of various optical signals rather than the individual optical signals that originated at the source. Processing at one or more nodes is required to create these signal combinations. The processing is preferably performed in the optical domain, but electrical processing may be required to generate more complex signal combinations. With proper processing of the received data, the destination can recreate the original signals. For protection purposes, the transmissions are sent over diverse paths and the signal combinations are such that if one signal is lost due to a failure, it can be recovered (almost) immediately from the other signals that are received. To mine the full benefits of network coding, there must be multiple signals that can be advantageously combined, as is often the case when routing a large amount of multicast traffic. Network coding *may* result in a more efficient use of network resources; however, even if the amount of required capacity is approximately the same as in a more conventional shared-mesh restoration approach, the recovery time is typically much faster.

Another important design aspect of multicast routing in an all-optical network is the selection of regeneration sites, where judicious use of regeneration may reduce the cost of the multicast tree. For example, it may be advantageous to favor the branching points of a multicast tree for regeneration. Refer to the tree of Fig. 12.10, where the source is node Q, and the multicast destinations are nodes W, X, Y, and Z. Assume that the optical reach is 2000 km. In Fig. 12.10a, the signal is regenerated at the furthest possible node from the source without violating the optical reach. This results in regenerations at nodes S and T. If, however, the regeneration occurs at the branching point node R as in Fig. 12.10b, then no other regeneration is needed, resulting in a lower cost solution.

12.6.1 Manycast Routing

In one variation of multicast routing, only N of the M destinations must be reached by the multicast tree,

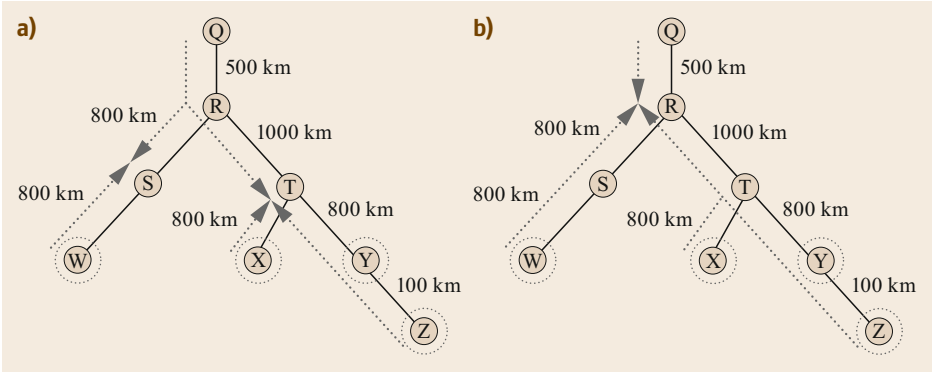


Fig. 12.10a,b Assume that the optical reach is 2000 km. **(a)** Regeneration at both nodes S and T. **(b)** Regeneration only at node R (© Monarch Network Architects LLC)

where $N < M$. The routing goal is still to produce the lowest-weight multicast tree; however, the routing algorithm must incorporate the selection of the N destinations as well. This scenario is known as *manycast*.

Manycast is useful in applications such as distributed computing. For example, there may be M processing centers distributed in the network, but an end user requires the use of only N of the processors. The goal of the end user is typically to minimize latency; it does not have a preference as to which N data centers are utilized. Optimal, or near-optimal, manycast routing, where the link metric is distance, is useful for such a scenario.

Various heuristic algorithms have been proposed for manycast routing, where the goal is to find the manycast tree of lowest weight. One effective heuristic in particular is a variation of the MP multicast algorithm [12.33].

12.7 Wavelength Assignment

Routing is one critical aspect of network design. Another important component is wavelength assignment, where each routed connection is assigned to a portion of the spectrum centered on a particular wavelength. Each time a connection enters the electrical domain, the opportunity exists (typically) to change the wavelength to which the connection has been assigned. The wavelength assignment process must satisfy the constraint that no two connections can be assigned the same wavelength on a given fiber. If this constraint were to be violated, the connections would occupy the same portion of the spectrum and interfere with each other. Note that when a link is populated with multiple fiber pairs, there can be multiple connections carried on the same wavelength on the link as long as each of the connections is routed on a different fiber.

12.7.1 Interaction Between Regeneration and Wavelength Assignment

As discussed in Sect. 12.1, all traffic that is routed through a node in an O-E-O network is regenerated. Regeneration is most commonly accomplished through the use of two back-to-back *transponders*, as shown in Fig. 12.11. A signal enters and exits a transponder in the optical domain but is converted to the electrical domain internally. The transponders interface to each other on a common wavelength (typically 1310 nm) through what is known as the *short-reach interface*. At the opposite end, the transponder receives and transmits a WDM-compatible wavelength, where the wavelength is generally in the 1500 nm range. The WDM-compatible wavelengths that are received/transmitted by the two transponders (i.e., λ_j and λ_k in Fig. 12.11) can be selected independently, thereby enabling wavelength conversion. (Wavelengths are also referred to as *lambdas*, with wavelength j represented by λ_j.)

Because of the flexibility afforded by regeneration, wavelengths can be assigned arbitrarily in an O-E-O network as long as each connection on a fiber is

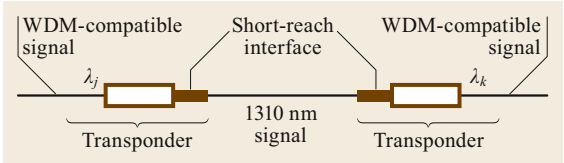


Fig. 12.11 Regeneration through the use of back-to-back transponders. The WDM-compatible signals associated with the two transponders, λ_j and λ_k, do not have to be the same, thereby enabling wavelength conversion (© Monarch Network Architects LLC)

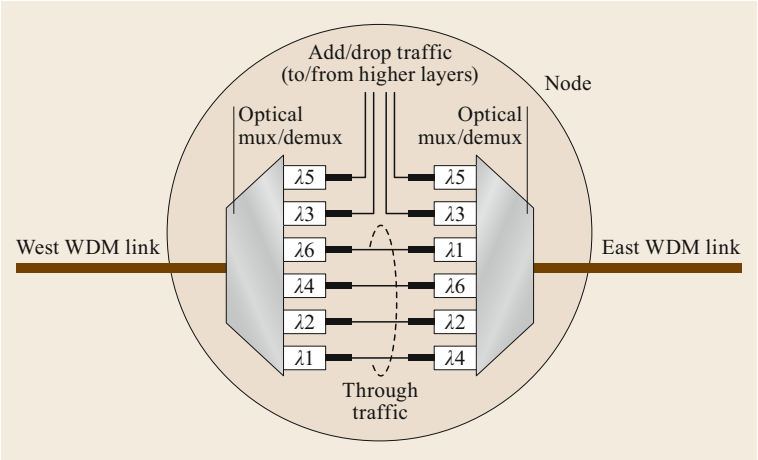
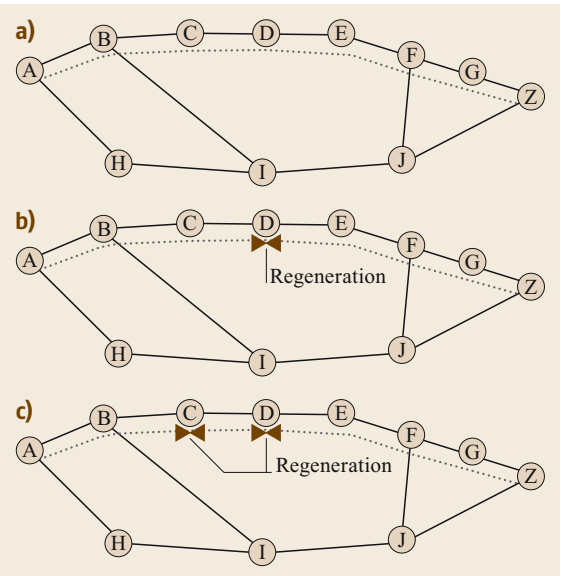


Fig. 12.12 A representative node in an O-E-O network. Wavelengths can be assigned independently to the connections on each link as long as each wavelength assigned on a fiber is unique. The lower-most connection is carried on λ_1 on the West link and carried on λ_4 on the East link (© Monarch Network Architects LLC)

assigned a unique wavelength. This is illustrated in Fig. 12.12, where the lower most connection enters the node from the *West* link on wavelength 1 and exits the node on the *East* link on wavelength 4. In the reverse direction, wavelength 4 is converted to wavelength 1. (By convention, the nodal fibers at a degree-two node are referred to as West and East; it may have no correlation to the actual geography of the node.) There is no requirement that the wavelength must be changed. As shown in Fig. 12.12, another connection enters and exits the node on λ_2.

In contrast, signals in an all-optical network are not regenerated at every node along the path. It is only required that a signal be regenerated prior to it traveling a distance that is longer than the system's optical reach.



If a signal is not regenerated at a node (i.e., it traverses the node in the optical domain), then it is carried on the same wavelength into and out of the node. More generally, if the signal traverses N consecutive nodes in the optical domain, then it is carried on the same wavelength on $N+1$ links. The wavelength assignment process consists of finding a single wavelength that is available on each of these links. We use the term *all-optical segment* to refer to a portion of a connection that rides in the optical domain without any conversion to the electrical domain.

The difficulty of assigning wavelengths clearly depends on two factors: the number of links in an all-optical segment and the utilization level on those links. As the number of links in an all-optical segment increases, the difficulty in finding a wavelength that is free on each one of the links typically increases as well. Similarly, as the utilization level of a link increases, fewer wavelengths are available, thus making it less likely that a free wavelength can be found on the entire all-optical segment.

This presents an interesting tradeoff. Each regeneration requires the deployment of two transponders, adding to the cost, power consumption, and failure rate

Fig. 12.13 **(a)** The entire end-to-end connection is carried in the optical domain. The same wavelength must be available on all seven links traversed by the connection. **(b)** The connection is regenerated at node D, creating two all-optical segments: AD and DZ. These two segments can be assigned a different wavelength. **(c)** Assume that an available wavelength cannot be found from A to D, but λ_j is available from A to C, and λ_k is available from C to D. An additional regeneration can be added at node C to make wavelength assignment feasible for this connection (© Monarch Network Architects LLC) ◄

of the associated connection. However, wavelength assignment becomes simpler the more regeneration that is present. This is illustrated in Fig. 12.13. Figure 12.13a shows a true all-optical connection that remains in the optical domain from node A to node Z. In order to find a suitable wavelength to assign to this connection, the same wavelength must be available on all seven links that compose this end-to-end path. Figure 12.13b shows that same connection with an intermediate regeneration at node D. The presence of regeneration simplifies the wavelength assignment process. One needs to find a wavelength that is free on the three links from node A to node D and a wavelength that is free on the four links from node D to node Z. There is no requirement that these two wavelengths be the same.

If wavelength assignment fails for a connection, where a suitable wavelength cannot be found for one or more of the all-optical segments composing the connection's path, then at least four design options exist. First, the connection request can be blocked. Second, it may be possible to select different regeneration locations for the connection, thereby producing a different set of all-optical segments. Wavelength assignment may be feasible on this alternate set of segments. For example, in Fig. 12.1c (this figure appeared in Sect. 12.1), the two required regenerations are chosen to be at nodes C and F, producing the all-optical segments AC, CF, and FZ. Alternatively, the two regenerations could be placed at nodes B and E without violating the optical-reach constraint (there are other alternatives as well). This choice produces a completely different set of all-optical segments: AB, BE, and EZ. If moving the regenerations does not produce a feasible wavelength assignment, or if there are no regenerations for the connection, then a third option is to route the connection on a different path and re-attempt the wavelength assignment process on the new path. As a fourth option, the connection can remain on the same path, but one or more regenerations can be added even though they are not required due to optical-reach concerns, thus incurring greater cost. This last option is illustrated in Fig. 12.13c. Assume that a wavelength is available along the DZ segment, but no single wavelength is available on all of the links of the AD segment. Furthermore, assume that λ_j is available on the links from node A to node C, and λ_k is available on the link between node C and node D. By adding a regeneration at node C (and its attendant costs), wavelength assignment for this connection becomes feasible.

Numerous studies have been performed to study the level of blocking that results due to a failure of the wavelength assignment process (assuming extra regenerations cannot be added to alleviate wavelength contention). The consensus of the majority of these studies is that *sparse* regeneration provides enough opportunities for wavelength conversion, resulting in a relatively low level of blocking due to wavelength contention [12.1, 12, 34, 35]. In a continental-scale network, the regeneration that is required based on optical reach is minimal (e.g., three regenerations in a connection that extends from the East coast to the West coast in a United States backbone network) but sufficient to achieve low levels of wavelength contention. In networks of smaller geographic extent, e.g., metro networks, the optical reach may be longer than any end-to-end connection such that regeneration is not required. However, there are other network functionalities that limit the extent of any all-optical segment. For example, low-data-rate traffic may need to be processed periodically by a grooming switch or router to make better use of the fiber capacity. Currently, grooming devices operate in the electrical domain such that the grooming process concurrently regenerates the signal. Thus, sparse grooming translates to sparse regeneration, which, in turn, allows sparse wavelength conversion. The net effect is that wavelength contention can remain low in a range of networks. Furthermore, it has been shown that just a small amount of extra regeneration effectively eliminates wavelength contention in a typical network [12.35].

It should be noted that some studies appear to indicate that wavelength contention is a major problem that results in excessive blocking. Further investigation of the details of these studies may reveal flaws in the underlying assumptions. For example, the study may not take advantage of regeneration as an opportunity to convert the wavelength. Another possible weakness is that effective wavelength assignment algorithms may not be employed. Wavelength assignment algorithms are covered in the next section.

12.8 Wavelength Assignment Algorithms

In this section, we assume that routing and wavelength assignment are two separate steps; Sect. 12.9 considers integrated approaches. It is assumed that one or more new connection requests are passed to the network design process. In the first step, each connection in the set is routed. Once a path has been selected for a connection, the required regeneration locations along that path are determined. This yields a set of

all-optical segments (i.e., the portions of the paths that lie between an endpoint and a regeneration or between two regenerations). The list of all-optical segments is passed to the wavelength assignment algorithm, which is responsible for selecting a feasible wavelength for each segment. In order for a wavelength assignment to be feasible, any two all-optical segments that are routed on the same fiber at any point along their respective paths must be assigned different wavelengths. Furthermore, the wavelength assignment must be compatible with the technology of the underlying optical system. An optical system supports a limited number of wavelengths on a fiber, thereby placing a bound on the number of different wavelengths that can be used in the assignment process. For example, a backbone network system may support 80 wavelengths on a fiber, whereas a metro network deployment may support only 40 wavelengths (there are typically fewer wavelengths needed in a metro network and 40-wavelength technology is of lower cost than 80-wavelength technology).

The wavelength assignment problem is analogous to the graph coloring problem, where each node of a graph must be assigned a color subject to the constraint that any two nodes that are adjacent in the graph topology must be assigned different colors. The objective is to color the graph using as few colors as possible. To elucidate the analogy with wavelength assignment, let each all-optical segment in a network design correspond to a node in the coloring graph. Links are added between any two nodes of the coloring graph if the paths of the two corresponding all-optical segments have any fibers in common. The resulting graph is known as the *conflict graph*, which is illustrated for a small example with four all-optical segments in Fig. 12.14. (This graph is also called the *auxiliary graph*.) Solving the graph coloring problem on this graph produces the wavelength assignment; i.e., the color assigned to a node represents the wavelength assigned to the all-optical segment corresponding to that node. The resulting wavelength assignment is such that the minimum number of wavelengths is used.

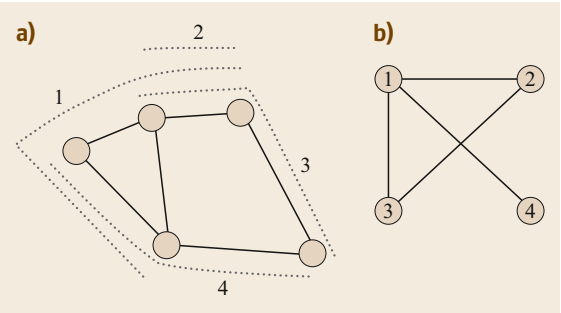


Fig. 12.14 **(a)** Four all-optical segments, each one numbered, are routed as indicated by the *dotted lines*. It is assumed that there is one fiber-pair on each link. **(b)** The resulting conflict graph, where each node represents one of the all-optical segments. A link exists between two nodes if the corresponding all-optical segments are routed on the same fiber on any link in the original graph

There are no known polynomial-time algorithms for optimally solving the graph coloring problem for general instances of the problem. This implies that there are no corresponding efficient algorithms that can optimally solve general instances of the wavelength assignment problem. Thus, heuristic algorithms are typically used. The heuristics encompass two aspects: (1) generating the order in which the all-optical segments are assigned a wavelength and (2) selecting a wavelength for each of the segments. As with routing order, there are many strategies that can be used for ordering the all-optical segments. Some of the strategies developed for ordering the nodes in a graph coloring can readily be extended to this problem, most notably the *Dsatur* strategy [12.36].

Developing heuristics to select which wavelength to assign to an all-optical segment is a well-researched topic, and numerous such heuristics have been proposed [12.37]. They differ in factors such as complexity and the amount of network-state information that needs to be monitored. In spite of the array of proposals, two of the simplest heuristics, both proposed in the very early days of optical-network research, remain the algorithms most commonly used for wavelength assignment. These heuristics are first-fit and most-used [12.38], described in further detail below. Both of these algorithms are suitable for any network topology and provide relatively good performance in realistic networks. For example, wavelength contention does not generally become an issue until there are at least a few links in the network with roughly 85% of the wavelengths used. Whether first-fit or most-used performs better for a particular network design depends on the network topology and the traffic. In general, the differences in performance are small. One advantage of first-fit is that, in contrast to most-used, it does not require any global knowledge, making it more suitable for distributed implementation.

Either of the schemes can be applied whether there is a single fiber-pair or multiple fiber-pairs on a link. Note that there are wavelength assignment schemes specifically designed for the multiple fiber-pair scenario, most notably the *least-loaded* scheme [12.12]. This has been shown to perform better than first-fit and most-used when there are several fiber-pairs per link [12.37]. As fiber capacities have increased, however, systems with several fiber-pairs on a link have become a less common occurrence.

12.8.1 First-Fit Algorithm

In the first-fit algorithm, each wavelength is assigned an index from 1 to W, where W is the maximum number of wavelengths supported on a fiber. No correlation is required between the order in which a wavelength appears in the spectrum and the index number assigned. The indices remain fixed as the network design evolves. Whenever wavelength assignment is needed for an all-optical segment, the search for an available wavelength proceeds in an order from the lowest index to the highest index. The first wavelength that is available on all links that compose the all-optical segment is assigned to the segment. First-fit is simple to implement and requires only that the status of each wavelength on each link be tracked.

Due to the interdependence of wavelength assignment across links, the presence of failure events, and the presence of network churn (i.e., the process of connections being established and then later torn down), the *indexing ordering* does not guarantee the actual *assignment order* on a particular link. Thus, relying on the indexing scheme to enforce a particular assignment ordering on a link for performance purposes is not prudent; for more details, see [12.1].

12.8.2 Most-Used Algorithm

The most-used algorithm is more adaptive than first-fit but requires more computation. Whenever a wavelength needs to be assigned to an all-optical segment, a wavelength order is established based on the number of link-fibers on which each wavelength has already been assigned in the network. The wavelength that has been assigned on the most link-fibers already is given the lowest index, and the wavelength that has been assigned on the second-most link-fibers is given the second lowest index, etc. Thus, the indexing order changes depending on the current state of the network. After the wavelengths have been indexed, the assignment process proceeds as in first-fit. The motivation behind this scheme is that a wavelength that has already been assigned on many fibers will be more difficult to use again. Thus, if a scenario arises where a heavily-used wavelength can be used, it should be assigned.

12.9 One-Step RWA

When routing and wavelength assignment are treated as separate steps in network design, it is possible that the routing process produces a path on which the wavelength assignment process fails (assuming extra regenerations are not added to alleviate the encountered wavelength contention). Alternatively, one can consider integrated routing and wavelength assignment methodologies such that if a path is selected, it is guaranteed to be feasible from a wavelength assignment perspective as well.

Various one-step RWA methodologies are discussed below, all of which impose additional processing and/or memory burdens. When the network is not heavily loaded, implementing routing and wavelength assignment as independent steps typically produces feasible solutions. Thus, under these conditions, the multistep process is favored, as it is usually faster. However, under heavy load, using a one-step methodology can provide a small improvement in performance [12.1]. Furthermore, under heavy load, some of the one-step methodologies may be more tractable, as the scarcity of free wavelengths should lead to lower complexity.

12.9.1 Topology Pruning

One of the earliest advocated one-step algorithms starts with a particular wavelength and reduces the network topology to only those links on which this wavelength is available. The routing algorithm (e.g., a shortest-path algorithm) is run on this pruned topology. If no path can be found, or if the path is too circuitous, another wavelength is chosen and the process run through again on the corresponding pruned topology. The process is repeated with successive wavelengths until a suitable path is found. With this approach, it is guaranteed that there will be a free wavelength on any route that is found. If a suitable path cannot be found after repeating the procedure for all of the wavelengths, the connection request is blocked.

In a network with regeneration, using this combined routing and wavelength assignment procedure makes the problem unnecessarily more difficult because it implicitly searches for a wavelength that is free along the whole extent of the path. As discussed earlier, it is necessary to find a free wavelength only along each all-optical segment, not along the end-to-end connection. One variation of the scheme is to select ahead of time where the regenerations are likely to occur for a connection and apply the combined routing and wavelength assignment approach to each expected all-optical segment individually. However, the route that is ultimately found could be somewhat circuitous and require regeneration at different sites than where was predicted,

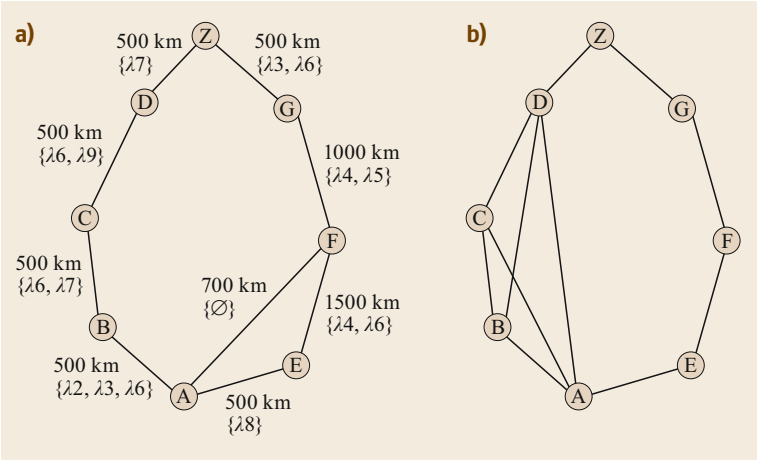


Fig. 12.15 **(a)** The true network topology, where it is assumed that the optical reach is 2000 km. The wavelengths listed next to each link are the wavelengths that are assumed to be free on the link. **(b)** The reachability graph, where a link is added between a node pair if there is a regeneration-free path between the nodes with at least one available wavelength along the path (© Monarch Network Architects LLC)

so that the process may need to be run through again. Overall, this strategy is less than ideal.

12.9.2 Reachability-Graph Transformation

A more direct integrated RWA approach is to create a transformed graph whenever a new connection request needs to be routed and assigned a wavelength, where the newly-formed graph represents feasible all-optical segments; i.e., feasible with respect to both optical reach and available wavelengths [12.1, 39]. The process to form this graph, specifically selecting the nodes that are added to this graph, depends on the timeframe of the network design. In long-term network planning, it is assumed that there is time to deploy any equipment that may be required for a design. Under these planning conditions, every network node of the real network topology appears in the transformed graph. In contrast, with real-time network operation, the required connection setup times do not allow for equipment to be deployed; only equipment that is already available in the network can be utilized for the new traffic. In this scenario, only nodes with available regeneration equipment, plus the source and the destination of the new connection request, are added to the transformed graph.

After adding the nodes to the transformed graph, a link is added between a pair of nodes in this graph only if there exists a regeneration-free path between the nodes in the true topology *and* there exists at least one wavelength that is available along the path. Even if there are multiple regeneration-free paths between a node pair, or multiple wavelengths free on a path, at most one link is added between a node pair. This transformed graph is referred to as the *reachability graph*.

An example of such a graph transformation is shown in Fig. 12.15. The true topology is shown in Fig. 12.15a, where the wavelengths that are assumed to still be available on a link are shown. The connection request is assumed to be between nodes A and Z, and the optical reach is assumed to be 2000 km. Furthermore, it is assumed that this is a long-term network design exercise and regeneration is permitted at any node. Thus, all nodes appear in the reachability graph. With these assumptions, the corresponding reachability graph is shown in Fig. 12.15b. All of the original links appear in this graph, except for link AF, which has no available wavelengths. In addition, links AC, AD, and BD are added because the respective associated paths, A–B–C, A–B–C–D, and B–C–D, are less than 2000 km and have a free wavelength (i.e., on each of these paths λ_6 is free). No link is added to represent the path E–F–G, even though λ_4 is available on this path because the path distance is 2500 km, which is longer than the optical reach.

In a real network with many nodes and wavelengths, creating the reachability graph can potentially be time consuming. A list of all node pairs where the shortest path between the nodes is less than the optical reach is maintained. These node pairs represent the possible all-optical-segment endpoints. At the time of a new connection request, a search is performed for each node pair in this list to find a regeneration-free path between the two nodes where some wavelength is available along the whole path. This determines whether a link is added between the two nodes in the reachability graph.

One strategy to do this is the topology-pruning approach described in Sect. 12.9.1, where the true topology is reduced to just those links that have a particular wavelength free. A shortest-path algorithm between the pair of nodes (i.e., the potential endpoints of an all-optical segment) is run on the pruned topology. If the distance of the resulting path is less than the optical reach, then a suitable regeneration-free path has been

found, and a link is added in the reachability graph between the node pair. If the path distance is greater than the optical reach, or no path can be found, then the process must be repeated for another wavelength. Many (if not all) of the wavelengths may need to be considered.

An alternative strategy is to run a K-shortest-paths algorithm on the true topology, where K is large enough such that *all* regeneration-free paths between a node pair of interest are found, and the resulting list is stored. At the time of a new connection request, the paths are checked for a free wavelength. If any of the paths have a free wavelength along the whole path, then a link is added between the node pair in the reachability graph. (To speed up the process, one could consider just a subset of the possible regeneration-free paths between a node pair; there may be some loss of optimality, but it is likely to be small.)

Once the reachability graph is formed, a shortest-path algorithm is run from the connection source to the connection destination to find the path in the reachability graph with the fewest hops (i.e., all link metrics are set to 1). Each hop corresponds to an all-optical segment in the true topology. If a path is found, then it is guaranteed to have the fewest number of feasible regenerations, and each resulting all-optical segment is guaranteed to have an available wavelength. In the example of Fig. 12.15, path A–D–Z is found, which corresponds to the all-optical segments A–B–C–D and D–Z in the true topology. These segments are assigned λ_6 and λ_7, respectively. Further subtleties of the reachability-graph methodology can be found in [12.1].

12.9.3 Flow-Based Methods

Global optimization techniques, such as *integer linear programming* (ILP), can be applied to the one-step RWA problem as well. ILPs typically consider the whole solution space to find the optimal solution. However, ILP methodologies often have a long runtime and are impractical except for small networks with little traffic. A more practical approach is to use efficient *linear programming* (LP) techniques (e.g., the Simplex algorithm), combined with strategies that drive the solution to integer values. Relaxing the integrality constraints enables more rapid convergence. Various techniques are applied to ultimately produce a (possibly nonoptimal) integer solution.

For example, routing a set of traffic connections can be formulated as a *multicommodity flow* (MCF) problem, where each source/destination pair in the traffic set can be considered a different commodity that needs to be carried by the network [12.40–42]. Additional variables and constraints are needed to enforce wavelength continuity. An integer solution to the problem is typically desired, which corresponds to routing each connection over just one path, using a single wavelength on a link. The integrality constraints are relaxed in the LP approach to make the problem more tractable. Despite not enforcing integer solutions, the LP can be combined with various perturbation and rounding techniques to improve the likelihood that an integer solution is found [12.41, 42]. However, even with LP relaxation techniques, there may be a greater-than-linear increase in runtime as the number of connections increases [12.42].

One approach to speed up the process is to input a set of possible paths that can be utilized by a connection between any given source and destination. This is analogous to calculating a set of candidate paths for alternative-path routing. Restricting the LP to a set of candidate paths, as opposed to allowing the LP to freely select the paths, may result in a less than optimal solution; however, with a good choice of candidate paths, the effect should be small. Another benefit to preselecting the paths is that the regeneration sites can be selected up front. This allows the wavelength continuity constraint to be specified on a per-all-optical-segment basis rather than requiring that wavelength continuity be enforced end-to-end.

Ideally, the cost function that is used in the LP encourages load balancing. Also, it is preferable if the cost function is input as a piecewise linear function with integer breakpoints as another means of pushing the LP towards an integer solution.

As suggested earlier, using a one-step RWA approach such as an LP methodology may be more expedient when adding connections to a highly loaded network. At that stage, there are few available wavelengths on each link, such that the solution space is much smaller. This should allow the LP to converge more quickly.

It is interesting to compare the results of the one-step LP-based RWA approach to those of a multistep approach, where an LP methodology is used just for the routing portion and a commonly-used graph coloring algorithm is used for wavelength assignment. The performances have been shown to be similar (depending on the cost functions used in the LPs), indicating that good results can be obtained using the simpler multistep approach [12.42]. The runtime of the multistep approach was an order of magnitude faster.

12.9.4 ILP-Based Ring RWA

Although ILP formulations have generally been considered too slow for practical RWA, a scalable ILP methodology has been proposed for ring topologies [12.43]. This methodology includes a decomposi-

tion approach that is optimal, fast for any reasonably sized ring, with a runtime that is essentially independent of the amount of traffic on the ring. This ILP methodology is, thus, a scalable one-step RWA approach for realistic ring problem instances. Further research is needed to determine whether the decomposition procedure can be extended to arbitrary mesh topologies.

12.10 Impairment-Aware Routing and Wavelength Assignment

The transmission of an optical signal is subject to an array of impairments that affect the signal quality and where the signal must be regenerated. There has been a large research effort in the area of *impairment-aware routing and wavelength assignment* (IA-RWA), leading to a number of proposed link metrics and design methodologies that account for various impairments during the routing and regeneration processes [12.44, 45].

One of the major impairments that an optical signal encounters is accumulated noise. The strength of the signal compared to the level of the noise is captured by the signal's OSNR, where signals with lower OSNR are more difficult to receive without errors. Many other optical impairments arise from the physical properties of light propagating in a fiber. For example, the propagation speed of light within a fiber depends on the optical frequency. This causes the optical signal pulses, which have a finite spectral width, to be distorted as they propagate along a fiber. This phenomenon is known as *chromatic dispersion*, or simply *dispersion*. Dispersion is a linear impairment that can typically be managed (e.g., through the use of dispersion-compensating fiber or the use of coherent technology) such that its detrimental effect is limited. Furthermore, the presence of some level of dispersion can be helpful in mitigating the effect of other optical impairments [12.46–48].

There are numerous nonlinear optical impairments that are more difficult to manage [12.49–52]. Several of these optical effects arise as a result of the fiber refractive index being dependent on the optical intensity. (The refractive index governs the speed of light propagation in a fiber.) As the optical signal power is increased, these nonlinearities become more prominent. One such nonlinearity is *self-phase modulation* (SPM), where the intensity of the light causes the phase of the optical signal to vary with time. This potentially interacts with the system dispersion to cause significant pulse distortion. *Cross-phase modulation* (XPM) is a similar effect, except that it arises from the interaction of two signals, which is more likely to occur when signals are closely packed together in the spectrum. Another nonlinear effect is *four-wave mixing* (FWM). This arises when signals carried on three particularly spaced optical frequencies interact to yield a stray signal at a fourth frequency, or two frequencies interact to generate two stray signals. These stray signals can potentially interfere with the desired signals at or near these frequencies.

In many all-optical networks, the transmission system is designed such that the power levels are low enough, or the dispersion levels are high enough, so that impairments due to adjacently propagating wavelengths are relatively small. However, there may be transmission systems where relatively high power levels are required, leading to scenarios where populating adjacent, or nearly adjacent, wavelengths in the spectrum produces non-negligible nonlinear impairments, most notably XPM. In such systems, the quality of transmission (QoT) for a given connection may depend on which other wavelengths are in use on the same fibers.

There are two methods for dealing with this scenario. The first strategy is to ensure that connections are established with enough system margin to tolerate the *worst-case* impairments that could possibly arise from populating adjacent wavelengths with other connections. This allows wavelengths to be assigned to connections without concern over interwavelength impairments. If a particular connection is deemed feasible at the time of its establishment, it should remain feasible regardless of what other connections may later be added.

In the second strategy, the effects of interwavelength impairments are calculated more precisely. The optical reach of a particular available wavelength along a given path is determined at the time a connection request is received, based on the state of the adjacent wavelengths. Consider assigning wavelength i to a new connection on a given path. If wavelengths $i-1$ and $i+1$ are not being used on the fibers that compose this path, then wavelength i may have additional optical reach, as compared to the case where a worst-case reach assumption is used. This could lead to fewer required regenerations for the new connection. The drawback is that if future connections populate wavelength $i-1$ and/or $i+1$, the performance of wavelength i may degrade below an acceptable QoT, forcing the associated connection to be assigned to a different wavelength or be rerouted, which is undesirable. If such a modifica-

tion to an existing connection is not permitted, then the strategy of maximizing the reach of wavelength i could result in future connections being blocked from using wavelengths $i-1$ or $i+1$.

The decision as to which of the two strategies to use may depend on how regenerations are handled. If regeneration is permitted in the network, then the impact of using a worst-case impairment assumption will likely be extra regenerations, because the system optical reach will effectively be reduced. However, as long as the optical reach is still reasonably long, small reductions in the reach (e.g., 2500 km versus 2800 km) do not have a large impact on the amount of regeneration required and the overall network cost [12.1].

If, however, the system requires that connections be *truly* all-optical, with no regeneration, then the policy for handling impairments may have an impact on blocking. For example, the end-to-end path distance of a new connection may be very close to the nominal optical reach. Establishing the new connection on a wavelength that is distant from any populated wavelengths may allow the connection to be successfully deployed, whereas the worst-case impairment assumption would dictate that it be blocked. This effect was examined more fully in [12.53] for a backbone network of relatively small geographic extent, where no regeneration was permitted. The two strategies outlined above were compared; i.e., either assume worst-case interwavelength impairments or calculate the interwavelength impairments more accurately based on the actual network state. In either strategy, moving an existing connection to a different path or wavelength was not permitted. The results indicated that when interwavelength impairments were more precisely calculated, the blocking rates were reduced by about an order of magnitude, due to there being a larger set of feasible paths from which to choose. However, this type of pure all-optical scenario would not arise in a network of large geographic extent, because some regeneration is needed regardless of how interwavelength impairments are treated. Interwavelength impairments are also unlikely to be an issue in a metro network, where the optical reach, even with worst-case assumptions, is typically longer than any path. Thus, the benefit of more precisely calculating interwavelength effects may not be significant in many practical networks.

Nevertheless, strategies have been developed to take interwavelength impairments into account when performing RWA for a new connection request; for example, using a cost-vector approach to routing [12.54]. Various per-wavelength components are included in the cost vector that is used for shortest-path routing, where the vector captures interwavelength impairments such as XPM and FWM. For each available wavelength on a link, the cost component for that wavelength-link combination is calculated based on the wavelengths that are already populated on that link. A modified Dijkstra routing algorithm is run with the cost vector, using the principle of dominated paths. (A path between two nodes that has all of its cost metrics higher than those of another path between the same two nodes is considered *dominated*.) Multiple nondominated paths from source to intermediate nodes are tracked, and any dominated paths are eliminated from further consideration. When the routing algorithm terminates, a scalar-generating function is applied to the final cost vector for each remaining feasible path/wavelength combination to determine which one to use.

Machine learning (as part of a *cognitive network)* has also been proposed for estimating the QoT of a new connection [12.55–57]. In one such approach, a database is maintained for a set of paths for which the QoT is known (through prior analysis, experimentation, and/or performance monitoring of live connections in the network). Each of these paths is characterized by a set of metrics (e.g., path distance, assigned wavelength, modulation format). When a potential new connection (or all-optical segment) is being evaluated, the paths in the database that are most similar to it are used to determine whether the QoT will meet the system threshold. Maintaining a proper-sized database is important to achieve the proper balance between accuracy and computation time.

12.10.1 Mixed Line-Rate Systems

An important scenario that may warrant accounting for interwavelength impairments more precisely is when multiple line rates co-propagate on a single fiber. For example, a single fiber may carry wavelengths that have been assigned to a combination of 10 and 40 Gb/s connections. The different modulation formats that are typically used for these connection rates may have a negative impact on each other. Experiments have shown that 10 Gb/s signals may have an especially detrimental effect on near-by co-propagating 40 Gb/s signals due to XPM [12.58, 59]. Furthermore, the performance penalties are severe enough that leaving enough system margin to account for the worst-case XPM would be too detrimental to the system reach. (The performance penalty induced by 10 Gb/s signals on near-by co-propagating 100 Gb/s signals is less severe; similarly, co-propagating 40 and 100 Gb/s connections are not problematic.)

To deal with this situation, a *soft* partitioning can be enforced in the wavelength assignment process, where the 40 Gb/s wavelengths are assigned from one end of the spectrum and the 10 Gb/s wavelengths are as-

signed starting at the other end. Additionally, relatively short 40 Gb/s connections, which can tolerate the performance penalty of adjacent 10 Gb/s wavelengths, can be proactively assigned wavelengths from the buffer area between the two portions of the spectrum. Note that a *fixed* partitioning of resources between the line rates is not advocated, as fixed partitioning generally leads to more blocking.

This wavelength-assignment strategy will have a tendency to segregate the conflicting line rates to minimize the performance penalties. As the network fill rate increases, and the high and low spectral ranges approach each other, the cost-vector RWA approach discussed above could be used to capture the penalties associated with adding a particular wavelength of a particular rate to a given link.

12.11 Flexible (Elastic) Optical Networks

WDM systems have historically utilized a standardized wavelength grid alignment. For example, since the 2000 timeframe, WDM *backbone* networks have typically employed a grid where the wavelengths are spaced at 50 GHz intervals, and each channel is assigned a fixed 50 GHz of spectrum, as shown in Fig. 12.16a. While the spacing has remained fixed at 50 GHz, advancements in technology have enabled the capacity of a wavelength (i.e., the line rate) to increase, rising from 2.5 to 100 Gb/s over an approximately 20-year span.

A fixed wavelength grid in combination with higher line rates has resulted in a corresponding increase in network capacity. However, it has also resulted in a greater disparity between the line rate and the rate of the client services being carried on the wavelengths. For example, there may be a significant amount of Gigabit (1 Gb/s) Ethernet services that must be carried in a network that utilizes 100 Gb/s wavelengths. In order to efficiently utilize the bandwidth of each wavelength, it is necessary to carry multiple services on one wavelength. The packing of services onto a wavelength is typically accomplished through a process known as *grooming* [12.60, 61]. Grooming is generally performed in the electrical domain using, for example, synchronous optical network/synchronous digital hierarchy (SONET/SDH) switches, optical transport network (OTN) switches, or internet protocol (IP) routers. While effective at packing the wavelengths, electronic grooming switches and routers pose major challenges in cost, size, and power consumption.

In order to eliminate or reduce the need for electronic grooming, various optical-domain grooming schemes have been proposed, such as optical packet switching (OPS) [12.62] and optical burst switching (OBS) [12.63]. However, these schemes have challenges of their own that have prevented their widespread adoption. One optical-domain grooming scheme that has gained traction is the *spectrum-sliced elastic* (SLICE) optical path architecture, originally proposed in 2008 [12.64–66]. In contrast to most grooming schemes, SLICE grooms in the frequency domain, not the time domain, thereby avoiding problematic time-based con-

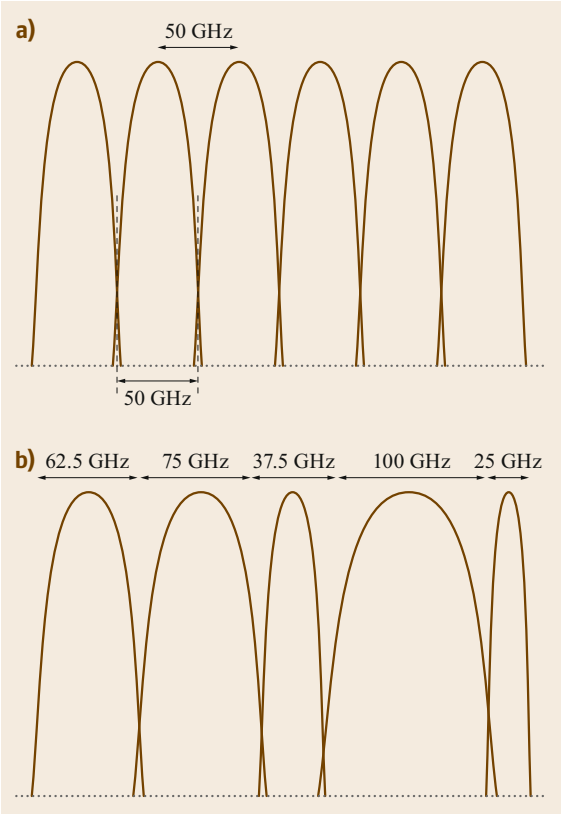


Fig. 12.16 **(a)** Fixed grid with 50 GHz spectral widths and spacing. **(b)** Gridless architecture with arbitrary spectral widths and spacing. In a practical implementation, the spectrum cannot be partitioned arbitrarily. Rather, the spectrum is likely to be logically divided into fixed-sized fine-granularity slots. Each optical channel is allocated the number of slots that it requires, as opposed to a completely unquantized amount of spectrum. This *minigrid* approach is illustrated in Fig. 12.17 ►

tention issues (electronic versus optical, and time versus frequency are separate dichotomies). The central tenet of the SLICE approach is that network client services are allocated the amount of optical spectrum that they require; i.e., the capacity of a fiber is *sliced* arbitrarily to match the requirements of the clients. (Practical limitations to this vision are discussed below.) For example, a 40 Gb/s client service is assigned 20 GHz of spectrum (assuming a spectral efficiency of 2 b/s/Hz). This approach eliminates the notion of wavelengths with fixed line rate located on a fixed grid. Networks that adhere to the SLICE approach are considered flexible, or gridless, networks. An example of this flexibility is illustrated in Fig. 12.16b, where *optical channels* of various spectral widths have been allocated.

Note that while an optical channel may be allocated less than 50 GHz worth of spectrum in order to efficiently carry a low-rate client service, it may also be allocated more than 50 GHz of spectrum. Thus, this technology can efficiently carry high-rate services as well, e.g., 400 Gb/s or 1 Tb/s clients. (It is expected that rates of 400 Gb/s and greater are likely to require more than 50 GHz of spectrum. Thus, the standardized wavelength grid was updated in 2012 to offer more flexibility. The grid plan supports any mix of wavelength spacings on one fiber, as long as each wavelength aligns with a 6.25 GHz grid, and the bandwidth assigned to each wavelength is a multiple of 12.5 GHz [12.67].)

Another component of the SLICE architecture is the ability to increase or decrease the amount of spectrum allocated to an optical channel. (An increase in spectral width requires that there be free spectrum available for expansion.) For example, a SONET/SDH-based client may dynamically adjust its service rate through the use of the *link capacity adjustment scheme* (LCAS) [12.68]. A SLICE-based network can correspondingly adjust the amount of spectrum allocated to that service. Overall, the flexibility engendered by the SLICE approach has resulted in such networks being referred to as *elastic optical networks* (EONs), which is the terminology that is used in the remainder of this chapter.

Various enabling technologies are needed to implement an EON, as briefly introduced here. First, the transmission technology must enable the deployment of tightly-packed, variable-sized optical channels. One transmission technique that has emerged as a leading candidate for EONs is the optical analog of *orthogonal frequency-division multiplexing* (OFDM) [12.69–73]. With OFDM, the optical signal is carried on a number of low-rate carriers. By increasing or decreasing the number of carriers, the bandwidth of the aggregate optical signal is modified accordingly. One advantageous property of OFDM is that, due to the lower rate of the constituent carriers, there is a greater tolerance to many fiber-based impairments. Additionally, the speed of the underlying electronics can be lower. There are other possible transmission techniques suitable for EONs, including *Nyquist-WDM* [12.74, 75].

The transmit/receive technology must also be capable of handling variable-rate, fine-granularity optical channels [12.76]. Software-controlled *bandwidth-variable transponders* (BVTs) have been developed for this purpose. Additionally, as analyzed in [12.1], the number of BVTs required in an EON network is potentially large, especially if there are numerous narrow-bandwidth optical channels routed in the network. One proposal to address this is a BVT that can be sliced into several *virtual transponders*, each of which serves one optical channel [12.76].

The network switches that route the optical channels must be compatible with the EON model. More specifically, the filter technology of such switches must be compatible with variable-granularity optical channels, where the filter shape and bandwidth can be set remotely via software [12.77–79]. This can be implemented with, for example, liquid crystal on silicon (LCoS) technology [12.80–82].

The limitations of the filtering technology impose practical implementation restrictions on EONs. First, the bandwidth of an optical channel cannot be arbitrarily fine. There is a minimum granularity bandwidth on which filters can efficiently operate (the narrower the filter, the more it deviates from the ideal sharp-edged shape). Thus, practical EONs are likely to be quasi-gridless, not totally gridless. For example, the spectrum may be divided into 320 frequency *slots*, where each slot is 12.5 GHz in bandwidth (as opposed to, for example, 80 wavelengths, each with 50 GHz of bandwidth). Each optical channel is allocated the number of slots that it requires, as opposed to a completely unquantized amount of spectrum. (This type of system is sometimes referred to as having a *mini-grid* [12.83].) One implication is that some amount of electronic grooming is still required to efficiently fill the slots.

A second limitation is that the optical channels cannot be immediately adjacent to each other. A *guardband* (likely one slot of bandwidth) is needed between the optical channels to allow the switch filters to operate on a single channel without affecting the neighboring channels. The need for guardbands, which represent wasted bandwidth, negatively impacts the utilization efficiency of the fiber capacity.

With these two limitations, an EON spectrum assignment might look as shown in Fig. 12.17. The portion of the spectrum shown has been partitioned into 30 slots. There are 4 allocated optical channels, of size 5, 3, 8, and 4 slots (left to right). A one-slot guardband is assumed.

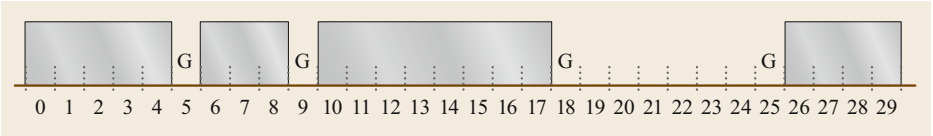


Fig. 12.17 In this depiction, the spectrum is divided into 30 spectral slots. Four optical channels are assigned on the fiber, as indicated by the *shaded boxes*; *G* indicates a guardband slot (© Monarch Network Architects LLC)

Even with these practical issues, EONs have the *potential* to: reduce, but not eliminate, the amount of electronic grooming; use bandwidth more efficiently, assuming the number of required guardbands is not excessive; and be better suited to carry a wide range of services; i.e., both low-rate and high-rate services. EONs are being aggressively pursued as an enabling technology for future networks.

12.12 Routing and Spectrum Assignment in Elastic Optical Networks

Just as routing and wavelength assignment (RWA) is a key component of the network design process in traditional grid-based wavelength networks, *routing and spectrum assignment* (RSA) is fundamental to the design of EONs [12.84, 85]. Spectrum assignment corresponds to assigning a particular set of slots to an optical channel. The restrictions are analogous to those of wavelength assignment. First, a slot on a fiber can be assigned to only one optical channel at any point in time. Second, the slots assigned to an all-optical segment must be the same along that segment; i.e., there is a *continuity constraint*. Additionally, the slots assigned must be contiguous; this *contiguity constraint* does not exist for traditional wavelength assignment (this constraint is relaxed when multipath routing is supported in the EON, as is detailed in Sect. 12.12.6).

Algorithm scalability is of greater concern with EONs because of the increased complexity. For example, there may be 320 slots to track in an EON as opposed to 80 wavelengths in a grid-based network. Furthermore, the contiguity constraint poses a significant additional challenge. Finding a path with enough available bandwidth to carry a new optical channel is not sufficient; the bandwidth must be contiguous (i.e., the slots must be consecutive). The notion of spectral fragmentation becomes an important metric that must be monitored. Because of the additional complexity, treating routing, regeneration, and spectrum assignment as separate steps is likely the approach to be used, although single-step methodologies have been proposed as well, as noted below.

With long-term network planning in an EON, a number of connection requests undergo the RSA process at one time. As with long-term RWA (Sect. 12.5), the order in which connections are routed and/or assigned spectrum can affect the ultimate network performance. For example, it is typically better to start the routing process with the connections where selecting a route is more challenging; e.g., connections that require longer paths and/or more bandwidth, or where the possible paths for a connection include more hot links. Furthermore, as in RWA, metaheuristics, such as simulated annealing, can be used to adjust the ordering to improve upon the solution [12.86, 87].

12.12.1 Routing

With respect to routing, the same options exist as for RWA, i.e., fixed-path routing, alternative-path routing, and dynamic routing. Fixed-path routing is generally undesirable due to the resulting load imbalances. Dynamic routing typically leads to several different paths being chosen between a given source and destination. This is especially problematic with spectrum assignment (SA), where greater freedom in selecting paths is likely to lead to more spectral fragmentation; i.e., it is preferable to assign spectrum along the same link sequences, so that contiguous blocks of spectrum remain free on the links.

Thus, alternative-path routing is typically favored for RSA, as it is for RWA. A set of candidate paths is calculated for each relevant source/destination pair, where the paths in the set provide diversity with respect to the expected hot links of a network. With alternative-path routing in RWA, link load is typically used to select one of the candidate paths for a new connection request. With RSA, link load does not tell the whole story. The amount of fragmentation that results from the selection of a particular route may be more important [12.88].

Two types of fragmentation arise in EONs: *vertical fragmentation* corresponds to the spectral fragmentation of a single link, whereas *horizontal* (or *spatial*) *fragmentation* considers the alignment of the available spectrum on adjacent links [12.89, 90]. In contrast, tra-

ditional wavelength-based networks suffer from just horizontal fragmentation.

Various measures have been proposed to capture both types of fragmentation. One such metric is *utilization entropy* (UE) [12.91]. Let a slot that is occupied by an existing optical channel be represented by a *1* and an unoccupied slot be represented by a *0*. For any two consecutive slots on a fiber, a status change occurs when a *1* is followed by a *0*, or vice versa. The UE for a link is defined as (total number of status changes on the link)/(total number of slots on the link −1) yielding a normalized value between 0.0 and 1.0. Higher UE indicates a greater degree of fragmentation. UE can also be calculated for a path, where the average number of status changes per slot across any two consecutive links of the path is considered.

UE essentially captures the number of spectral gaps; however, it does not explicitly consider the size of those gaps. Larger spectral gaps are more useful as they can accommodate traffic of higher bandwidth. One metric to address this compares the size of the largest available gap to the sum of the sizes of all available gaps [12.92]. To capture available gap size more formally, a Shannon-entropy fragmentation metric has been proposed [12.93]

$$H_{\text{frag}} = -\sum_{i=1}^{N} \frac{D_i}{D} \ln \frac{D_i}{D} \,,$$

where D is the total number of slots on a link, and D_i is the number of slots in the i-th block (a block is a consecutive sequence of slots that are all 1s or all 0s; i.e., a block is either fully utilized or fully available). Higher values of H_{frag} indicate more fragmentation. This metric can be extended to paths by performing a bitwise-OR of the occupancy value of a slot along each link of the path. Interestingly, simulations showed that considering the metric for each link of a path produced better results than the path-based metric [12.93]. This is largely due to the coarse nature of the bitwise-OR operation (i.e., a single *1* in a slot along the path results in a *1* for that slot).

Another metric that takes into account horizontal fragmentation considers the alignment of available slots across link pairs, where each pair is composed of one link on the path, and one link that is adjacent to that link [12.88]. The metric captures how many optical channels of size S slots can be accommodated on each such link pair, for all possible S, and weights this count according to the expected number of optical channels of that size. Other fragmentation metrics, similar in spirit to those mentioned above, are proposed in [12.89, 90].

When alternative-path routing is used in a multistep RSA approach, the candidate path with, for example, the lowest path-based fragmentation metric could be selected for a new optical connection. In an example of a one-step approach where routing and SA are handled together, the increase in the fragmentation metric is calculated for each feasible candidate path/SA combination. The path and assignment that results in the smallest increase in fragmentation is then chosen. While likely to produce improved performance as compared to the multistep approach, evaluating the various route/SA combinations could be time consuming.

12.12.2 Spectrum Assignment

In the multistep RSA approach, once a route has been selected for an optical channel, the next step is to determine where regeneration, if any, is required. Regeneration divides the path of the optical channel into a set of all-optical segments. Spectrum can be assigned to each segment independently; i.e., regeneration in EONs allows *spectrum conversion*, just as regeneration in wavelength-based networks allows wavelength conversion.

The WA problem was mapped to an instance of graph coloring, as is detailed in Sect. 12.8. This analogy can be extended to SA as well. As with WA, a conflict graph is constructed, where each vertex corresponds to an all-optical segment, and two vertices are connected by an edge if the corresponding segments have at least one network fiber in common. Additionally, for SA, each vertex is weighted by the number of slots that are required by the optical channel to which the segment belongs. The graph is colored using a *weighted* graph coloring algorithm. Any solution to the weighted graph coloring problem can be mapped to a solution for the corresponding SA problem, where the slots assigned to each vertex must be contiguous in order to enforce the spectral contiguity constraint [12.94].

Alternatively, SA can be mapped to the problem of scheduling tasks on a multiprocessor system [12.95]. Consider a set of tasks, each associated with a fixed set of processors and a required processing time. The task scheduling problem involves minimizing the *schedule length* (i.e., the time by which all tasks have been completed) subject to the following: a processor can work on at most one task at any given time; each task must be processed simultaneously by all processors in its associated set; and preemptions are not permitted [12.96–98]. By mapping each all-optical segment that needs to be assigned spectrum to a task and letting the set of links over which the segment is routed correspond to the fixed set of processors assigned to each task, the SA problem becomes an instance of task scheduling. The assigned start and end times of a task correspond to the assigned start and end slots of a segment. The objective of minimizing the schedule length corresponds to min-

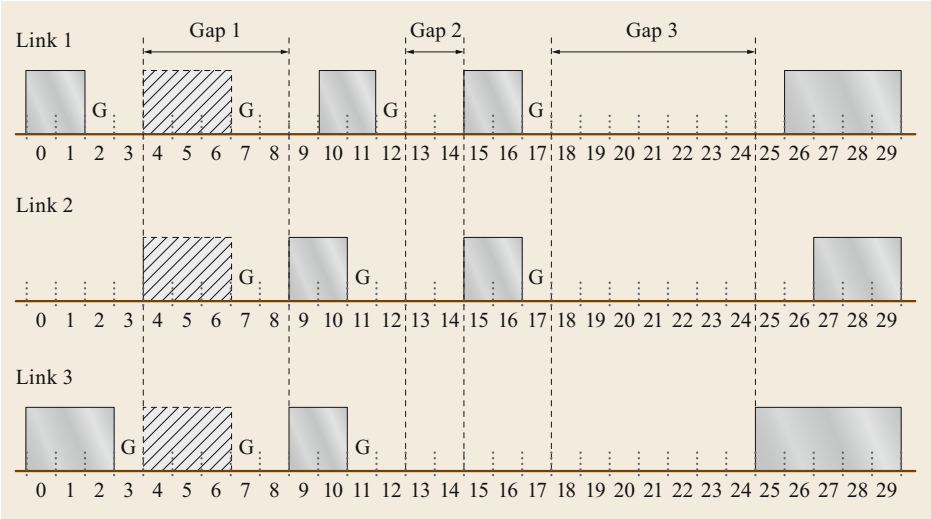


Fig. 12.18 A new optical channel, requiring a total of four slots, is routed all-optically on three links. The *solid shaded slots* indicate spectrum that has previously been assigned. Of the three existing gaps, the best-fitting one is gap 1. The new optical channel, represented by the *hatched slots*, is shown assigned to this gap. This strands bandwidth on link 1 (slot 3) and on links 2 and 3 (slot 8) (© Monarch Network Architects LLC)

imizing the highest numbered slot needed to carry any of the optical channels.

Most WA heuristics can be extended for use in SA. For example, in the first-fit heuristic, the lowest-numbered spectral gap that is wide enough to accommodate the new optical channel (or all-optical segment) is selected [12.99, 100]. In the most-used heuristic, the spectral gap that maximizes the usage of the slots in that gap across the network is selected. Despite the additional information that is considered in most-used, first-fit yielded a slightly lower blocking probability in various studies [12.86, 101].

In one proposed variation, first-fit is combined with alternative-path routing, such that the candidate path is selected that has the lowest-numbered spectral gap of sufficient width along the whole path [12.102]. This is a one-step RSA algorithm; i.e., it selects the route and spectrum assignment together. This scheme is more appropriate for networks that do not require regeneration because it inherently assumes that the same spectrum assignment is utilized along the whole end-to-end path.

Variations of first-fit have also been proposed to better address fragmentation issues. For example, in first-last-fit [12.103], optical channels that require an odd number of slots are assigned spectrum using first-fit, whereas optical channels that require an even number of slots are assigned spectrum using last-fit (i.e., the highest-indexed spectral gap is selected). This *soft* partitioning of odd and even sized optical channels results in a small improvement in blocking probability [12.104].

Another SA heuristic that attempts to minimize fragmentation is best-fit. It assigns a new optical channel to the smallest spectral gap that is large enough to accommodate it. This heuristic is illustrated in Fig. 12.18, where the new optical channel is routed on links 1, 2, and 3, and requires four slots, including the guardband slot. The three existing gaps that span all three links are as shown. Gap 2 (two slots) is too small. Of the remaining two gaps, gap 1 (five slots) is selected over gap 3 (seven slots) because it is closer in size to the required four slots. The motivation behind best-fit is appealing, yet it has been shown to yield higher blocking as compared to first-fit [12.1]. The drawback of the scheme is that the residual unused portion of the selected gap (e.g., slot 8 on links 2 and 3) is typically very small, and effectively becomes stranded bandwidth. If gap 3 had been selected, starting at slot 18, a three-slot gap would remain on all three links. (Another potential advantage of first-fit is that it tends to group the utilized slots at one end of the spectrum, leaving larger available gaps at the other end of the spectrum. However, for a dynamic network, this quasi-orderly assignment pattern will likely dissipate over time.)

As compared to best-fit, a small performance improvement can be attained with exact-fit. This heuristic preferentially assigns an optical channel to a spectral gap that is of equal size. If such a gap does not exist, the scheme reverts to first-fit.

Another challenge that arises with SA is fairness with respect to blocking. As the spectrum becomes more fragmented, leaving only relatively small unoc-

cupied spectral gaps, finding a suitable spectrum assignment for high-bandwidth optical channels becomes increasingly more difficult. One can employ various partitioning schemes to address SA fairness [12.89]. For example, the spectrum can be divided into three regions, one each for low-rate, medium-rate, and high-rate connections. If the amount of traffic that falls within each one of these classes is not known, better performance may be obtained with a soft partitioning rather than a strict fixed partitioning.

Additional SA heuristic algorithms can be found in [12.84, 85]. Simulation results indicate only small differences in performance among the various heuristics considered (with the exception of the relatively poor-performing random assignment strategy, which assigns any spectral gap of sufficient size with equal probability) [12.85, 104].

12.12.3 ILP-Based RSA

A number of ILPs have been proposed for the RSA problem, where optimal (or near-optimal) routing and spectrum assignment is desired for a set of connection requests [12.100, 105]. As with RWA, such an approach is feasible only for small RSA instances. The ILP formulations need to capture the unique assignment of spectrum on a given fiber, the continuity constraint, and the contiguity constraint. Various objective functions can be considered, including: minimizing the maximum number of slots utilized on any link; minimizing the highest-numbered slot utilized on any link; and minimizing the total number of slots utilized across the network. If sufficient bandwidth is not available to carry all of the offered traffic, then the objective is typically to minimize the amount of requested bandwidth that is blocked.

Similar to ILPs used for RWA, by restricting the possible number of paths that can be utilized between a given source and destination (referred to as a *path-based* approach), the computation time can be markedly reduced with little loss of optimality [12.86, 106]. Furthermore, by inputting the set of possible slots that can be utilized by a particular connection (referred to as a *channel-based* approach), the number of required ILP constraints can be reduced significantly, thus speeding up the runtime (although it is still not fast enough to be able to handle large-sized RSA instances) [12.107]. Most of the proposed ILPs handle routing and spectrum assignment in one step. Alternatively, the problem can be decomposed into two ILPs (R + SA) to improve runtimes, though sacrificing some amount of optimality [12.86].

Alternatively, the RSA problem can be formulated as a *mixed integer program* (MIP). For example, in [12.108], a MIP that utilizes both the path-based approach (to limit the number of paths to consider) and the channel-based approach (to limit the number of possible spectrum assignment choices) is combined with a number of advanced optimization techniques (e.g., branch-and-bound, column generation). It was shown that optimal results can be found for networks of moderate size.

12.12.4 Distance-Adaptive RSA

In order to enable higher bit rates, optical transmission modulation schemes have become more complex, thereby requiring the use of advanced digital signal processing (DSP) in the transmit/receive (TxRx) equipment. The presence of the DSP also affords the opportunity to provide greater TxRx flexibility, where transmission characteristics such as modulation format, error correction coding, and symbol rate can be modified remotely through software.

This flexibility enables important system design tradeoffs, including spectral width vs. optical reach [12.99, 102, 109, 110]. By utilizing a more spectrally efficient modulation format, the amount of spectrum allocated to a signal can be reduced, while the signal bit-rate is maintained. The disadvantage is that the optical reach of the signal is decreased. For example, a 100 Gb/s TxRx could be capable of either 3000 km optical reach with 75 GHz bandwidth, or 2000 km optical reach with 50 GHz bandwidth. For relatively short connections (or, more generally, connections that are afflicted by fewer impairments), the decrease in optical reach may not lead to any extra regeneration. Even for longer connections, the tradeoff (i.e., extra regeneration) may be worthwhile if capacity is at a premium. EONs, with their ability to dynamically adjust the spectral width assigned to an optical signal, are especially well suited to take advantage of this TxRx flexibility.

The selection of the modulation format can be considered in the network design process in order to better optimize the network with respect to factors such as cost or capacity utilization. This has given rise to distance-adaptive RSA (DA-RSA), also known as the routing, modulation level, and spectrum-assignment (RMLSA) problem, where the spectral width is tailored to the path over which an optical signal is routed. Note that the presence of a regeneration typically affords the opportunity to change the modulation scheme, just as it allows spectrum conversion [12.1, 111].

One potential drawback to a flexible modulation scheme is that there will likely be a mix of modulation formats co-propagating on one fiber. As discussed in Sect. 12.10.1, when certain combinations of modulation formats are used in adjacent regions of the spectrum,

the performance of the carried optical signals may degrade. This may reduce the optical reach or require that extra guardbands be utilized between optical channels, where the latter negates some of the spectral benefit of utilizing a more efficient modulation format. Mixed modulation formats were investigated in the context of the tradeoff between bit rate and optical reach, and a reduction in spectral benefits was noted [12.112].

Numerous algorithms that address the RMLSA problem in EONs have been proposed, e.g., [12.99, 102]. Many of these studies consider only true all-optical networks, where regeneration is not permitted. Thus, the approach often utilized is to generate a set of candidate paths and then associate each path with the most spectrally efficient modulation format that yields a sufficient optical reach. A slightly more restrictive version of this strategy considers only those candidate paths that can make use of the most spectrally efficient modulation format suitable for the shortest possible path [12.113]. This allows some amount of path diversity without sacrificing spectral utilization. An RMLSA algorithm that favors higher path OSNR, greater spectral efficiency, shorter path distance, and less slot utilization along the path (in order from the highest-priority criterion to the lowest one) is considered in [12.114]. ILPs that solve RMLSA in one step or two steps (RML + SA) have been proposed as well [12.102]. The MIP approach can also be extended to the RMSLA problem [12.108].

12.12.5 Routing, Spectrum and Core Assignment

As the capacity limits of *conventional* fiber are close to being reached [12.17, 115], several *space-division-multiplexed* (SDM) solutions have been proposed, where signals are carried on spatially separate channels. Each signal can still be a WDM signal, thus yielding a multiplicative effect in capacity. For example, in contrast to conventional single-core fiber, *multicore fiber* (MCF) supports multiple cores in a single fiber [12.116]. With C cores, it is expected that the capacity of a single fiber would increase by a factor of C.

Having multiple cores presents interesting tradeoffs with respect to wavelength and spectrum assignment. More attention has been given to spectrum assignment, as the presence of multiple cores may ease some of the challenges. The associated problem has been termed *routing, spectrum and core assignment* (RSCA) [12.117, 118]. Assuming that ROADMs allow for signals to be routed through a node on any of the cores, then the presence of multiple cores may improve the likelihood of finding contiguous available spectrum to support a new all-optical segment. Additionally, some of the cores could be restricted to supporting optical channels of a particular bandwidth. This would reduce the amount of stranded bandwidth and improve the alignment of available resources in adjacent links to provide more opportunity for optical bypass.

One drawback of the MCF approach, however, is that depending on how closely the cores are packed in the fiber, crosstalk may be present between signals in different cores [12.119]. If the crosstalk is severe enough, the wavelengths in different cores become coupled. For example, with C coupled cores, all $C\lambda$1s would need to be routed or added/dropped as a single block (essentially the C wavelengths form a *waveband*); the individual wavelengths cannot be routed independently. Thus, algorithms would be needed to optimize this coarser granularity. Note that coupling occurs (to an even greater degree) with *few-mode fiber* (FMF), which is an alternative SDM approach that parallels MCF (i.e., a small number of modes are supported on a fiber in contrast to the single mode that is present in conventional fiber in a long-haul or metro network) [12.116].

12.12.6 Multipath Routing

In the above discussion, it was assumed that a single optical channel is established to carry a connection. If sufficient contiguous spectral slots are not available to carry a new connection along a particular route, then either another route must be considered or the new connection request is blocked. EONs offer an alternative approach, where multiple smaller optical channels are utilized to carry a single connection [12.92]. For example, a new connection may require 45 GHz of spectrum, but no single path from the source to the destination exists with adequate available spectral gaps. Rather than blocking the connection request, it is possible to split it into two or more connections, each with spectral requirements of less than 45 GHz, in order to take advantage of whatever spectrum is available. Thus, the 45 GHz of spectrum required could be provided, for example, via three spectral regions of size 20, 15, and 10 GHz. The destination must be capable of reconstituting the original signal.

There are two strategies that can be utilized in this multiple-optical-channel approach. First, the optical channels can be limited to following the same path. With this restriction, the latency from source to destination will be approximately the same for all constituent channels (small variances arise due to slight differences in the speed of light in fiber as a function of the spectral frequency utilized), thereby avoiding the challenge of having to provide large buffers at the des-

tination in order to synchronize the received signals. This methodology is especially useful as the spectrum becomes more fragmented, where there exists free spectrum on a fiber, but it is divided into relatively small spectral gaps. It also can be advantageous for routing high-bandwidth connections, where it is more challenging to find a spectral gap of appropriate size.

A more aggressive approach allows splitting a connection across optical channels that are routed on different paths. This clearly provides greater flexibility. The drawback is the *delay differential* among the selected paths, chiefly arising from the differences in the end-to-end distances of the paths. The greater the differential, the larger the required destination buffers that are used to reconstitute the original signal. Various algorithms exist to find a set of paths where the delay differential between the shortest and longest paths is below an acceptable threshold [12.120–124]. In one approach [12.123], a graph is created, where the nodes represent possible paths between the source and destination. Links are added between a pair of nodes if the differential delay between the corresponding two paths is below the acceptable threshold. A search is performed to find the minimum clique on this graph such that the total bandwidth that can be carried by the set of paths represented by the clique nodes is sufficient to carry the connection.

Regardless of whether the constituent optical channels are routed along the same path or different paths, one important drawback stems from the requirement of needing guardbands between optical channels. Thus, partitioning a connection across multiple optical channels results in more guardbands being required. This effectively limits the usage of this tactic; otherwise the benefits gained by the greater spectrum-assignment flexibility will be mitigated by the additional wasted bandwidth assigned to guardbands. To address this, one can impose a minimum bandwidth size for each of the constituent optical channels [12.123]. Another strategy is to limit multipath routing to those connections requiring a bandwidth greater than some threshold. This would limit the number of excess guardbands and alleviate some of the fairness concerns regarding high-rate connections.

Another potential drawback is the number of transponders required to support the constituent optical channels. For example, a connection that is split among three optical channels would require at least six transponders (one transponder at either end of an optical channel), and possibly more for regeneration. As noted in Sect. 12.11, the challenge of requiring numerous transponders may be addressed by BVTs that can be sliced into several *virtual transponders*, each of which serves one optical channel.

Multipath routing provides benefits beyond just capacity management. It can also be incorporated as part of a protection strategy [12.125, 126]. For example, a connection may ideally require a rate of R but under failure conditions is satisfied with R', for some $R' < R$. The connection can be split across N diverse paths, where the total bandwidth of the paths is R, and the total bandwidth of any $(N-1)$ of the paths is R'. If a failure occurs such that one of the paths is brought down, the remaining paths provide adequate bandwidth. This is an example of *bandwidth squeezing restoration*, where the bandwidth may be reduced under failure conditions [12.127]. In another scheme, which utilizes shared restoration, one optical channel is used for the working path, but multipath routing is permitted for the backup resources. The added flexibility in routing the backup traffic typically results in a greater amount of sharing of the restoration resources, thus reducing the amount of required spare capacity in the network. Note that most multipath-based protection/restoration schemes rely on algorithms that specifically find *diverse* paths that satisfy the maximum acceptable differential delay [12.128].

Multipath routing is supported in many standards, e.g., *virtual concatenation* (VCAT) in SONET/SDH and OTN [12.129, 130].

12.12.7 Defragmentation

The quasi-gridless nature of EONs, combined with increasing network dynamism where connections are established and torn down on shorter time scales, produces significant spectral fragmentation. As indicated in Sect. 12.12.1, fragmentation can be present on a single link, where the available spectrum consists of narrow, noncontiguous spectral gaps. Larger gaps are preferable, as they offer the opportunity to carry a greater range of new traffic (it is assumed here that single-path routing is employed). Fragmentation can also be problematic across links, where two consecutive links in a path may have spectral gaps of sufficient size to carry new traffic, but the gaps do not align. This necessitates regeneration of the optical connection for purposes of spectrum conversion. This is illustrated in Fig. 12.19a, where it is desired to route a new 25 GHz optical channel over links 1 and 2 (explicit guardbands are not shown). While both links have an available 25 GHz gap, the gaps are not spectrally aligned. Thus, regeneration would be required at the ROADM. In a network where regeneration is not supported, the new connection would be blocked (unless another route is utilized).

Fragmentation is likely to occur in EONs even if fragmentation-aware RSA policies are utilized. This

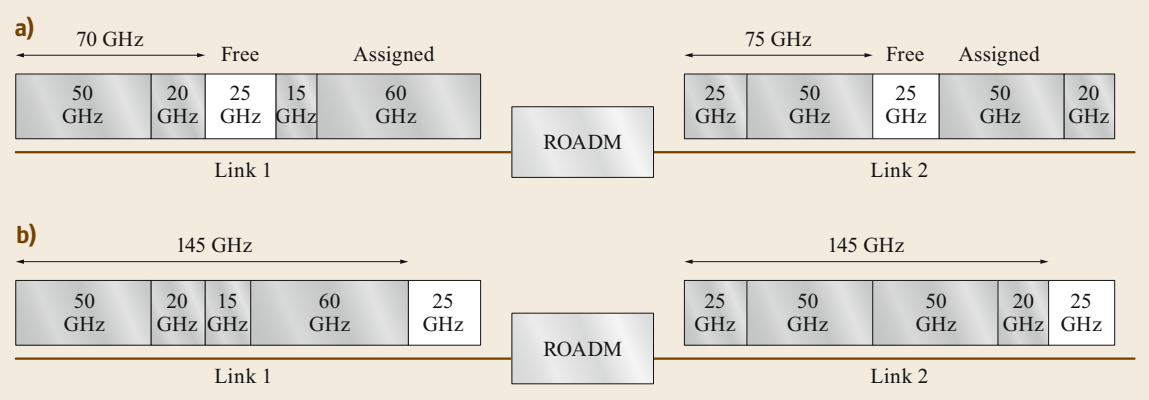


Fig. 12.19a,b Assume that a new optical channel requiring 25 GHz of spectrum needs to be routed across links 1 and 2. **(a)** Both links have an available 25 GHz gap; however, the gaps are not aligned. The optical channel cannot be routed all-optically through the ROADM, but instead would be regenerated to accomplish spectrum conversion. **(b)** After defragmentation, the assigned spectrum forms a contiguous block, leaving the available spectrum aligned (© Monarch Network Architects LLC)

mandates that defragmentation be implemented in such networks for purposes of efficient capacity management and reduced cost (less regeneration). Defragmentation involves the shifting or moving of spectrum that has already been assigned in order to aggregate the free spectrum into larger contiguous blocks on a link and align the free spectrum on adjacent links. This is shown in Fig. 12.19b, where the spectrum utilized on both links 1 and 2 is pushed into one contiguous region, leaving the available spectrum aligned at the high end of the spectrum. A new 25 GHz-channel can now be routed all-optically through the ROADM.

It should be noted that most defragmentation techniques have a tendency to realign the spectrum assignment such that the spectral slots utilized on a fiber are contiguous. However, this impedes the ability of an optical channel to increase its assigned bandwidth. Thus, defragmentation that leaves small spectral gaps between optical channels that exhibit elastic behavior may be desirable [12.131]. These small gaps could potentially be used in a time-based multiplexing scheme. For example, an optical channel on one side of the gap may expand its spectral allotment during daytime hours, whereas the optical channel on the other side of the gap expands at night. If there are times when neither optical channel is expanded, that spectrum can be assigned to low-priority traffic that may eventually be bumped.

Defragmentation is a challenging operation, as it requires the adjustment of numerous network elements; furthermore, it is operating on live traffic. Careful timing is needed to effect hitless (or close to hitless) defragmentation. The TxRx equipment at either end of the optical channel being shifted needs to be simultaneously retuned to the new spectral region. At the same time, the filters of any ROADMs along the path of the optical channel must be reconfigured to accommodate the newly-assigned spectral region, without disrupting any other existing traffic that passes through the ROADM. To ease the timing requirements, many schemes advocate the use of a *make-before-break* mechanism, where the new optical channel is established prior to the original one being removed. This depends on there being extra TxRx equipment to temporarily support an extra optical channel (or virtual or multiflow TxRx's can be used that support multiple optical channels [12.76, 132, 133]).

There are several approaches that can be taken with respect to defragmentation. In a reactive approach, defragmentation is performed only when new connection requests are blocked due to spectral fragmentation [12.134] (or when too much excess regeneration is required to accommodate a new path). The selection of the route for the new channel can, at least in part, be based on minimizing the number of conflicts with existing channels [12.135, 136].

In a proactive approach, defragmentation is performed periodically, e.g., as part of routine network maintenance, or when a network-wide fragmentation metric rises above a set threshold. One implicit periodic approach to defragmentation takes advantage of the growth/contraction cycle of the optical channels that exhibit elastic behavior. Growth is accomplished by preferentially adding slots at the lower end of the channel, whereas contraction is performed from the upper end of the channel [12.1]. This has a tendency to shift the channels to the lower end of the spectrum, resulting in more closely packed channels.

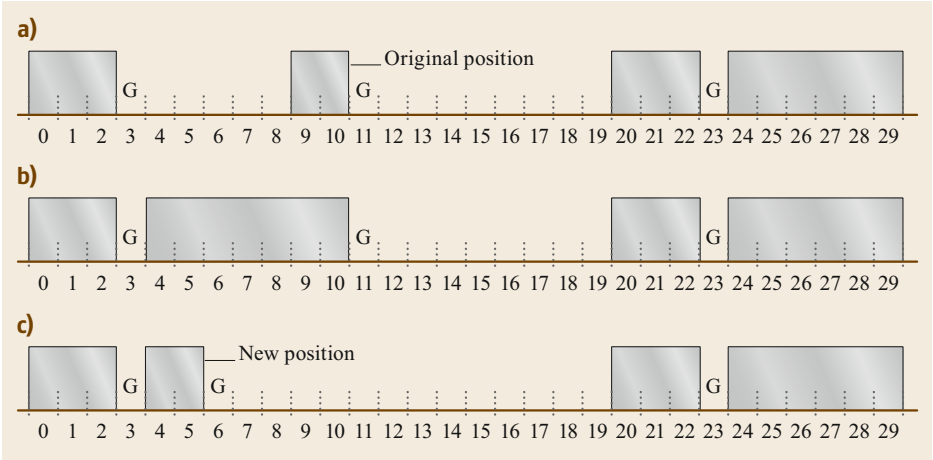


Fig. 12.20a–c Push–pull defragmentation. **(a)** The original spectral partitioning. Assume that it is desired to shift the optical channel on slots 9 and 10 to slots 4 and 5 on that same link. **(b)** In the first phase, the channel is extended to encompass the original slots, the new slots, and any slots in between. **(c)** In the second phase, the channel is contracted to encompass just the new slots, 4 and 5

Another bifurcation relates to the aggressiveness of the approach. In one class of strategies, rerouting is permitted, such that both the path and the spectrum assignment may be modified for an existing optical connection. In a less aggressive approach, rerouting of existing connections is not permitted; defragmentation is limited to shifting spectrum on a link. An example of this is a defragmentation mechanism that takes advantage of channel elasticity [12.76, 137]. This push–pull mechanism is illustrated in Fig. 12.20. Assume that the channel to be moved is initially assigned to slots 9 and 10, as shown in Fig. 12.20a. The objective of the defragmentation is to move the channel to slots 4 and 5. To accomplish this, the channel is first expanded to encompass both the original slots and the new slots, and all slots in between, as shown in Fig. 12.20b. The channel is then contracted to occupy only the desired new slots, as is shown in Fig. 12.20c. These expansion and contraction operations must be performed concurrently on each ROADM along the path of the optical channel, as well as at the TxRxs at the endpoints, so that end-to-end connectivity is never lost. Hitless operation of push–pull has been demonstrated [12.137].

While push–pull supports only a limited amount of defragmentation; i.e., spectrum can only be shifted along the same fiber, and all slots between the old and new spectrum assignments must be unassigned, it can provide a notable improvement in performance. For example, simulations have shown that it can potentially reduce the level of blocking by one to two orders of magnitude when used reactively to avoid blocking [12.138].

A third design decision is whether the defragmentation process is limited to sequential moves, where no more than one optical channel is adjusted at a given time. Or, defragmentation can involve parallel operations, where multiple optical channels are spectrally shifted and/or rerouted at one time [12.139]. Parallel operation provides more flexibility and faster operation time, although it requires accurate time synchronization across the network.

While defragmentation of an EON has been the focus here, fragmentation may also arise in the wavelength-based flexible-grid architecture supported by the new standard [12.67]. However, due to the relatively small number of bandwidths that are likely to be utilized on one fiber (e.g., perhaps 50, 62.5, and 75 GHz), the level of fragmentation should not be as severe as in an EON. Nevertheless, it is possible that defragmentation could be warranted [12.140].

12.13 Conclusion

As illustrated in this chapter, algorithms are a fundamental component of network design. Efficient utilization of network resources, which is tightly coupled to the economics of the network, relies on efficient algorithms. Furthermore, as new technologies are developed to improve network performance, new algorithms must

be developed to take full advantage of these advances. By extending previously developed design algorithms, relying to some extent on graph theory, and utilizing common sense in design decisions, the growing complexity of managing an optical network has thus far been met.

References

12.1 J.M. Simmons: *Optical Network Design and Planning*, 2nd edn. (Springer, New York 2014)

12.2 T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein: *Introduction to Algorithms*, 3rd edn. (MIT Press, Cambridge 2009)

12.3 R. Bhandari: *Survivable Networks: Algorithms for Diverse Routing* (Kluwer, Boston 1999)

12.4 R. Guerin, A. Orda: Computing shortest paths for any number of hops, IEEE/ACM Trans. Netw. **10**(5), 613–620 (2002)

12.5 T. Korkmaz, M. Krunz: Multi-constrained optimal path selection. In: *Proc. IEEE INFOCOM, Anchorage, Apr. 22–26*, Vol. 2 (2001) pp. 834–843

12.6 F. Kuipers, T. Korkmaz, M. Krunz, P. Van Mieghem: Performance evaluation of constraint-based path selection algorithms, IEEE Network **18**(5), 16–23 (2004)

12.7 J.Y. Yen: Finding the K shortest loopless paths in a network, Manag. Sci. **17**(11), 712–716 (1971)

12.8 J.W. Suurballe, R.E. Tarjan: A quick method for finding shortest pairs of disjoint paths, Networks **14**, 325–336 (1984)

12.9 J.W. Suurballe: Disjoint paths in a network, Networks **4**, 125–145 (1974)

12.10 D. Xu, Y. Xiong, C. Qiao, G. Li: Trap avoidance and protection schemes in networks with shared risk link groups, J. Lightwave Technol. **21**(11), 2683–2693 (2003)

12.11 A. Bach: The financial industry's race to zero latency and terabit networking. In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Los Angeles, Mar. 6–10* (2011), Service Provider Summit Keynote Address

12.12 E. Karasan, E. Ayanoglu: Effects of wavelength routing and selection algorithms on wavelength conversion gain in WDM optical networks, IEEE/ACM Trans. Netw. **6**(2), 186–196 (1998)

12.13 X. Chu, B. Li, Z. Zhang: A dynamic RWA algorithm in a wavelength-routed all-optical network with wavelength converters. In: *Proc. IEEE INFOCOM, San Francisco, Mar. 30–Apr. 3*, Vol. 3 (2003) pp. 1795–1804

12.14 J.L. Strand: Integrated route selection, transponder placement, wavelength assignment, and restoration in an advanced ROADM architecture, J. Opt. Commun. Netw. **4**(3), 282–288 (2012)

12.15 J.M. Simmons: Diversity requirements for selecting candidate paths for alternative-path routing. In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf., San Diego, Mar. 21–25* (2010), Paper NThA4

12.16 R.A. Barry, P.A. Humblet: Models of blocking probability in all-optical networks with and without wavelength changers, IEEE J. Sel. Areas Commun. **14**(5), 858–867 (1996)

12.17 J.M. Simmons: Technology and architectural approaches to address continued explosive growth in network traffic. In: *Proc. Int. Conf. Comput. Netw. Commun., Santa Clara, Jan. 28* (2017)

12.18 H. Hasegawa, S. Subramaniam, K. Sato: Node architecture and design of flexible waveband routing optical networks, J. Opt. Commun. Netw. **8**(10), 734–744 (2016)

12.19 A. Farrel, J.-P. Vasseur, J. Ash: A path computation element (PCE)-based architecture, Internet Engineering Task Force, Request for Comments 4655 (2006)

12.20 F. Paolucci, F. Cugini, A. Giorgetti, N. Sambo, P. Castoldi: A survey on the path computation element (PCE) architecture, IEEE Commun. Surv. Tutor. **15**(4), 1819–1841 (2013)

12.21 P.J.M. van Laarhoven, E.H.L. Aarts: *Simulated Annealing: Theory and Applications* (D. Reidel, Boston 1987)

12.22 L.H. Sahasrabuddhe, B. Mukherjee: Multicast routing algorithms and protocols: a tutorial, IEEE Network **14**(1), 90–102 (2000)

12.23 S. Voss: Steiner's problem in graphs: heuristic methods, Discrete Appl. Math. **40**(1), 45–72 (1992)

12.24 L. Kou, G. Markowsky, L. Berman: A fast algorithm for Steiner trees, Acta Inform. **15**(2), 141–145 (1981)

12.25 B.M. Waxman: Routing of multipoint connections, IEEE J. Sel. Areas Commun. **6**(9), 1617–1622 (1988)

12.26 H. Takahashi, A. Matsuyama: An approximate solution for the Steiner problem in graphs, Math. Jpn. **24**(6), 573–577 (1980)

12.27 N.K. Singhal, L.H. Sahasrabuddhe, B. Mukherjee: Provisioning of survivable multicast sessions against single link failures in optical WDM mesh networks, IEEE/OSA J. Lightwave Technol. **21**(11), 2587–2594 (2003)

12.28 X. Wang, L. Guo, L. Pang, J. Du, F. Jin: Segment protection algorithm with load balancing for multicasting WDM mesh networks. In: *Proc. 10th Int. Conf. Adv. Commun. Technol., Pyeongchang, Feb. 17–20*, Vol. 3 (2008) pp. 2013–2016

12.29 T. Panayiotou, G. Ellinas, N. Antoniades: Segment-based protection of multicast connections in metropolitan area optical networks with quality-of-transmission considerations, J. Opt. Commun. Netw. **4**(9), 692–702 (2012)

12.30 R. Koetter, M. Médard: An algebraic approach to network coding, IEEE/ACM Trans. Netw. **11**(5), 782–795 (2003)

12.31 R.C. Menendez, J.W. Gannett: Efficient, fault-tolerant all-optical multicast networks via network coding. In: *Proc. Opt. Fiber Commun./Natl. Fiber*

Opt. Eng. Conf., San Diego, Feb. 24–28 (2008), Paper JThA82

12.32 E.D. Manley, J.S. Deogun, L. Xu, D.R. Alexander: All-optical network coding, J. Opt. Commun. Netw. **2**(4), 175–191 (2010)

12.33 N. Charbonneau, V.M. Vokkarane: Tabu search meta-heuristic for static manycast routing and wavelength assignment over wavelength-routed optical WDM networks. In: *Proc. IEEE Int. Conf. Commun., Cape Town, May 23–27* (2010), https://doi.org/10.1109/ICC.2010.5502241

12.34 S. Subramaniam, M. Azizoglu, A.K. Somani: All-optical networks with sparse wavelength conversion, IEEE/ACM Trans. Netw. **4**(4), 544–557 (1996)

12.35 J.M. Simmons: Analysis of wavelength conversion in all-optical express backbone networks. In: *Proc. Opt. Fiber Commun. (OFC), Anaheim, Mar. 17–22* (2002), Paper TuG2

12.36 D. Brelaz: New methods to color the vertices of a graph, Communications ACM **22**(4), 251–256 (1979)

12.37 H. Zang, J.P. Jue, B. Mukherjee: A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks, Opt. Netw. Mag. **1**(1), 47–60 (2000)

12.38 I. Chlamtac, A. Ganz, G. Karmi: Purely optical networks for terabit communication. In: *Proc. IEEE INFOCOM, Ottawa, Apr. 23–27*, Vol. 3 (1989) pp. 887–896

12.39 O. Gerstel, H. Raza: Predeployment of resources in agile photonic networks, J. Lightwave Technol. **22**(10), 2236–2244 (2004)

12.40 D. Banerjee, B. Mukherjee: A practical approach for routing and wavelength assignment in large wavelength-routed optical networks, IEEE J. Sel. Areas Commun. **14**(5), 903–908 (1996)

12.41 A.E. Ozdaglar, D.P. Bertsekas: Routing and wavelength assignment in optical networks, IEEE/ACM Trans. Netw. **11**(2), 259–272 (2003)

12.42 K. Christodoulopoulos, K. Manousakis, E. Varvarigos: Comparison of routing and wavelength assignment algorithms in WDM networks. In: *Proc. IEEE Global Commun. Conf. (GLOBECOM), New Orleans, Nov. 30–Dec. 4* (2008), https://doi.org/10.1109/GLOCOM.2008.ECP.510

12.43 E. Yetginer, Z. Liu, G.N. Rouskas: Fast exact ILP decompositions for ring RWA, J. Opt. Commun. Netw. **3**(7), 577–586 (2011)

12.44 S. Azodolmolky, M. Klinkowski, E. Marin, D. Careglio, J. Pareta, I. Tomkos: A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks, Comput. Netw. **53**(7), 926–944 (2009)

12.45 A.G. Rahbar: Review of dynamic impairment-aware routing and wavelength assignment techniques in all-optical wavelength-routed networks, IEEE Commun. Surv. Tutor. **14**(4), 1065–1089 (2012)

12.46 C. Kurtzke: Suppression of fiber nonlinearities by appropriate dispersion management, IEEE Photonics Technol. Lett. **5**(10), 1250–1253 (1993)

12.47 R.W. Tkach, A.R. Chraplyvy: Dispersion and nonlinear effects in lightwave systems. In: *Proc. 7th Annu. Meeting IEEE LEOS, Boston, Oct. 31–Nov. 3*, Vol. 1 (1994) pp. 192–193

12.48 R.W. Tkach, A.R. Chraplyvy, F. Forghieri, A.H. Gnauck, R.M. Derosier: Four-photon mixing and high-speed WDM systems, J. Lightwave Technol. **13**(5), 841–849 (1995)

12.49 F. Forghieri, R.W. Tkach, A.R. Chraplyvy: Fiber nonlinearities and their impact on transmission systems. In: *Optical Fiber Telecommunications*, Vol. III A, ed. by I. Kaminow, T. Koch (Academic Press, San Diego 1997)

12.50 A.H. Gnauck, R.M. Jopson: Dispersion compensation for optical fiber systems. In: *Optical Fiber Telecommunications*, Vol. III A, ed. by I. Kaminow, T. Koch (Academic Press, San Diego 1997)

12.51 C.D. Poole, J. Nagel: Polarization effects in lightwave systems. In: *Optical Fiber Telecommunications*, Vol. III A, ed. by I. Kaminow, T. Koch (Academic Press, San Diego 1997)

12.52 P. Bayvel, R. Killey: Nonlinear optical effects in WDM transmission. In: *Optical Fiber Telecommunications*, Vol. IV B, ed. by I. Kaminow, T. Li (Academic Press, San Diego 2002)

12.53 K. Christodoulopoulos, P. Kokkinos, K. Manousakis, E.A. Varvarigos: Cross layer RWA in WDM networks: Is the added complexity useful or a burden? In: *Proc. Int. Conf. Transpar. Opt. Netw., Ponta Delgada, Jun. 28–Jul. 2* (2009), Paper Tu.A3.3

12.54 K. Manousakis, P. Kokkinos, K. Christodoulopoulos, E. Varvarigos: Joint online routing, wavelength assignment and regenerator allocation in translucent optical networks, J. Lightwave Technol. **28**(8), 1152–1163 (2010)

12.55 T. Jimenez, J.C. Aguado, I. de Miguel, R.J. Duran, M. Angelou, N. Merayo, P. Fernandez, R.M. Lorenzo, I. Tomkos, E.J. Abril: A cognitive quality of transmission estimator for core optical networks, J. Lightwave Technol. **31**(6), 942–951 (2013)

12.56 L. Barletta, A. Giusti, C. Rottondi, M. Tornatore: QoT estimation for unestablished lighpaths using machine learning. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Los Angeles, Mar. 19–23* (2017), Paper Th1J.1

12.57 S. Oda, M. Miyabe, S. Yoshida, T. Katagiri, Y. Aoki, T. Hoshida, J.C. Rasmussen, M. Birk, K. Tse: A learning living network with open ROADMs, J. Lightwave Technol. **35**(8), 1350–1356 (2017)

12.58 A. Bononi, M. Bertolini, P. Serena, G. Bellotti: Cross-phase modulation induced by OOK channels on higher-rate DQPSK and coherent QPSK channels, J. Lightwave Technol. **27**(18), 3974–3983 (2009)

12.59 O. Bertran-Pardo, J. Renaudier, G. Charlet, H. Mardoyan, P. Tran, M. Salsi, S. Bigo: Overlaying 10 Gb/s legacy optical networks with 40 and 100 Gb/s coherent terminals, J. Lightwave Technol. **30**(14), 2367–2375 (2012)

12.60 R. Dutta, G.N. Rouskas: Traffic grooming in WDM networks: past and future, IEEE Network **16**(6), 46–56 (2002)

12.61 R. Dutta, A.E. Kamal, G.N. Rouskas (Eds.): *Traffic Grooming for Optical Networks: Foundations, Techniques and Frontiers* (Springer, New York 2008)

12.62 D.J. Blumenthal: Optical packet switching. In: *Proc. 17th Annu. Meeting IEEE LEOS, Puerto Rico, Nov. 7–11* (2004), Paper ThU1

12.63 C. Qiao, M. Yoo: Optical burst switching (OBS)—a new paradigm for an optical internet, J. High Speed Netw. **8**(1), 69–84 (1999)

12.64 M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, T. Yoshimatsu, T. Kobayashi, Y. Miyamoto, K. Yonenaga, A. Takada, O. Ishida, S. Matsuoka: Demonstration of novel spectrum-efficient elastic optical path network with per-channel variable capacity of 40 Gb/s to over 400 Gb/s. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Brussels, Sep. 21–25* (2008), Paper Th.3.F.6

12.65 M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, S. Matsuoka: Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies, IEEE Commun. Mag. **47**(11), 66–73 (2009)

12.66 O. Gerstel, M. Jinno, A. Lord, S.J.B. Yoo: Elastic optical networking: a new dawn for the optical layer?, IEEE Commun. Mag. **50**(2), S12–S20 (2012)

12.67 International Telecommunication Union: *Spectral Grids for WDM Applications: DWDM Frequency Grid*, ITU-T Rec. G.694.1, 2.0 edn. 2012)

12.68 International Telecommunication Union: *Link Capacity Adjustment Scheme (LCAS) for Virtual Concatenated Signals*, ITU-T Rec. G. 7042/Y.1305, 2006)

12.69 I.B. Djordjevic, B. Vasic: Orthogonal frequency division multiplexing for high-speed optical transmission, Opt. Express **14**(9), 3767–3775 (2006)

12.70 W. Shieh, H. Bao, Y. Tang: Coherent optical OFDM: theory and design, Opt. Express **16**(2), 841–859 (2008)

12.71 W. Shieh: OFDM for flexible high-speed optical networks, J. Lightwave Technol. **29**(10), 1560–1577 (2011)

12.72 G. Shen, M. Zukerman: Spectrum-efficient and agile CO-OFDM optical transport networks: architecture, design, and operation, IEEE Commun. Mag. **50**(5), 82–89 (2012)

12.73 G. Zhang, M. De Leenheer, A. Morea, B. Mukherjee: A survey on OFDM-based elastic core optical networking, IEEE Commun. Surv. Tutor. **15**(1), 65–87 (2013)

12.74 G. Bosco, A. Carena, V. Curri, P. Poggiolini, F. Forghieri: Performance limits of Nyquist-WDM and CO-OFDM in high-speed PM-QPSK systems, IEEE Photonics Technol. Lett. **22**(1), 1129–1131 (2010)

12.75 G. Gavioli, E. Torrengo, G. Bosco, A. Carena, V. Curri, V. Miot, P. Poggiolini, M. Belmonte, F. Forghieri, C. Muzio, S. Piciaccia, A. Brinciotti, A. La Porta, C. Lezzi, S. Savory, S. Abrate: Investigation of the impact of ultra-narrow carrier spacing on the transmission of a 10-carrier 1Tb/s superchannel. In: *Proc. Opt. Fiber Commun./Natl. Fiber Optic Eng. Conf. (OFC/NFOEC), San Diego, Mar. 21–25* (2010), Paper OThD3

12.76 O. Gerstel: Flexible use of spectrum and photonic grooming. In: *Proc. Int. Conf. Photonics Switch., Monterey, Jul. 25–28* (2010), Paper PMD3

12.77 R. Ryf, Y. Su, L. Moller, S. Chandrasekhar, X. Liu, D.T. Nelson, C.R. Giles: Wavelength blocking filter with flexible data rates and channel spacing, J. Lightwave Technol. **23**(1), 54–61 (2005)

12.78 P. Colbourne, B. Collings: ROADM switching technologies. In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Los Angeles, Mar. 6–10* (2011), Paper OTuD1

12.79 D.M. Marom, D. Sinefeld: Beyond wavelength-selective channel switches: trends in support of flexible/elastic optical networks. In: *Proc. Int. Conf. Transpar. Opt. Netw. (ICTON), United Kingdom, Jul. 2–5* (2012), Paper Mo.B1.4

12.80 G. Baxter, S. Frisken, D. Abakoumov, H. Zhou, I. Clarke, A. Bartos, S. Poole: Highly programmable wavelength selective switch based on liquid crystal on silicon switching elements. In: *Proc. Opt. Fiber Comm./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Anaheim, Mar. 5–10* (2006), Paper OTuF2

12.81 S. Frisken: Advances in liquid crystal on silicon wavelength selective switching. In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Anaheim, Mar. 25–29* (2007), Paper OWV4

12.82 S. Frisken, G. Baxter, D. Abakoumov, H. Zhou, I. Clarke, S. Poole: Flexible and grid-less wavelength selective switch using LCOS technology. In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Los Angeles, Mar. 6–10* (2011), Paper OTuM3

12.83 G. Shen, Q. Yang: From coarse grid to mini-grid to gridless: How much can gridless help contentionless? In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Los Angeles, Mar. 6–10* (2011), Paper OTuI3

12.84 S. Talebi, F. Alam, I. Katib, M. Khamis, R. Salama, G.N. Rouskas: Spectrum management techniques for elastic optical networks: a survey, Opt. Switch. Netw. **13**, 34–48 (2014)

12.85 B.C. Chatterjee, N. Sarma, E. Oki: Routing and spectrum allocation in elastic optical networks: a tutorial, IEEE Commun. Surv. Tutor. **17**(3), 1776–1800 (2015)

12.86 K. Christodoulopoulos, I. Tomkos, E.A. Varvarigos: Routing and spectrum allocation in OFDM-based optical networks with elastic bandwidth allocation. In: *Proc. IEEE Global Telecommun. Conf. (GLOBECOM), Miami, Dec. 6–10* (2010) pp. 1–6

12.87 A.N. Patel, P.N. Ji, J.P. Jue, T. Wang: A naturally-inspired algorithm for routing, wavelength assignment, and spectrum allocation in flexible grid WDM networks. In: *Proc. IEEE Global Commun. Conf. (GLOBECOM), Anaheim, Dec. 3–7* (2012) pp. 340–345

12.88 Y. Sone, A. Hirano, A. Kadohata, M. Jinno, O. Ishida: Routing and spectrum assignment algorithm maximizes spectrum utilization in optical networks. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Geneva, Sep. 18–22* (2011), Paper Mo.1.K.3

12.89 R. Wang, B. Mukherjee: Spectrum management in heterogeneous bandwidth optical networks, Opt. Switch. Netw. **11**(A), 83–91 (2014)

12.90 M. Zhang, W. Lu, Z. Zhu, Y. Yin, S.J.B. Yoo: Planning and provisioning of elastic O-OFDM networks with fragmentation-aware routing and spectrum assignment (RSA) algorithms. In: *Proc. Asia Commun. Photonics Conf., Guanzhou, Nov. 7–10* (2012), Paper ATh2D.3

12.91 X. Wang, Q. Zhang, I. Kim, P. Palacharla, M. Sekiya: Utilization entropy for assessing resource fragmentation in optical networks. In: *Proc. Opt. Fiber Commun. Conf./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Los Angeles, Mar. 4–8* (2012), Paper OTh1A.2

12.92 Z. Zhu, W. Lu, L. Zhang, N. Ansari: Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing, J. Lightwave Technol. **31**(1), 15–22 (2013)

12.93 P. Wright, M.C. Parker, A. Lord: Minimum- and maximum-entropy routing and spectrum assignment for flexgrid elastic optical networking, J. Opt. Commun. Netw. **7**(1), A66–A72 (2015)

12.94 I. Popescu, I. Cerutti, N. Sambo, P. Castoldi: On the optimal design of a spectrum-switched optical network with multiple modulation formats and rates, J. Opt. Commun. Netw. **5**(11), 1275–1284 (2013)

12.95 S. Talebi, E. Bampis, I. Katib, G.N. Rouskas: The spectrum assignment (SA) problem in optical networks: a multiprocessor scheduling perspective, J. Opt. Commun. Netw. **6**(8), 754–763 (2014)

12.96 E. Bampis, M. Caramia, J. Fiala, A. Fishkin, A. Iovanella: Scheduling of independent dedicated multiprocessor tasks. In: *Proc. 13th Int. Symp. Algorithms Comput., Vancouver, Nov. 21–23* (2002) pp. 391–402

12.97 J.A. Hoogeveen, S.L. Van de Velde, B. Veltman: Complexity of scheduling multiprocessor tasks with prespecified processor allocations, Discrete Appl. Math. **55**, 259–272 (1994)

12.98 E. Bampis, A. Kononov: On the approximability of scheduling multiprocessor tasks with time dependent processing and processor requirements. In: *Proc. 15th Int. Parallel Distrib. Process. Symp., San Francisco, Apr. 23–27* (2001) pp. 2144–2151

12.99 M. Jinno, B. Kozicki, H. Takara, A. Watanabe, Y. Sone, T. Tanaka, A. Hirano: Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network, IEEE Commun. Mag. **48**(8), 138–145 (2010)

12.100 Y. Wang, X. Cao, Y. Pan: A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks. In: *Proc. IEEE INFOCOM, Shanghai, Apr. 10–15* (2011) pp. 1503–1511

12.101 R.J. Durán, I. Rodríguez, N. Fernández, I. de Miguel, N. Merayo, P. Fernández, J.C. Aguado, T. Jiménez, R.M. Lorenzo, E.J. Abril: Performance comparison of methods to solve the routing and spectrum allocation problem. In: *Proc. Int. Conf. Transpar. Opt. Netw. (ICTON), United Kingdom, Jul. 2–5* (2012), Paper Mo.C2.4

12.102 K. Christodoulopoulos, I. Tomkos, E.A. Varvarigos: Elastic bandwidth allocation in flexible OFDM-based optical networks, J. Lightwave Technol. **29**(9), 1354–1366 (2011)

12.103 W. Fadini, E. Oki: A subcarrier-slot partition scheme for wavelength assignment in elastic optical networks. In: *Proc. Int. Conf. High Perform. Switch. Routing (HPSR), Vancouver, Jul. 1–4* (2014) pp. 7–12

12.104 B.C. Chatterjee, E. Oki: Performance evaluation of spectrum allocation policies for elastic optical networks. In: *Proc. Int. Conf. Transpar. Opt. Netw. (ICTON), Budapest, Jul. 5–9* (2015), Paper Tu.D3.5

12.105 Y. Wang, X. Cao, Q. Hu, Y. Pan: Towards elastic and fine-granular bandwidth allocation in spectrum-sliced optical networks, J. Opt. Commun. Netw. **4**(11), 906–917 (2012)

12.106 M. Klinkowski, K. Walkowiak: Routing and spectrum assignment in spectrum sliced elastic optical path network, IEEE Commun. Lett. **15**(8), 884–886 (2011)

12.107 L. Velasco, M. Klinkowski, M. Ruiz, J. Comellas: Modeling the routing and spectrum allocation problem for flexgrid optical networks, Photonic Netw. Commun. **24**, 177–186 (2012)

12.108 M. Klinkowski, M. Zotkiewicz, K. Walkowiak, M. Pioro, M. Ruiz, L. Velasco: Solving large instances of the RSA problem in flexgrid elastic optical networks, J. Opt. Commun. Netw. **8**(5), 320–330 (2016)

12.109 B. Kozicki, H. Takara, Y. Sone, A. Watanabe, M. Jinno: Distance-adaptive spectrum allocation in elastic optical path network (SLICE) with bit per symbol adjustment. In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), San Diego, Mar. 21–25* (2010), Paper OMU3

12.110 B.T. Teipen, H. Griesser, M.H. Eiselt: Flexible bandwidth and bit-rate pro-grammability in future optical networks. In: *Proc. Int. Conf. Transpar. Opt. Netw. (ICTON), United Kingdom, Jul. 2–5* (2012), Paper Tu.C2.1

12.111 S. Yang, F. Kuipers: Impairment-aware routing in translucent spectrum-sliced elastic optical path networks. In: *Proc. 17th Eur. Conf. Netw. Opt. Commun., Villanova i la Geltru, June 20–22* (2012)

12.112 O. Rival, G. Villares, A. Morea: Impact of interchannel nonlinearities on the planning of 25–100 Gb/s elastic optical networks, J. Lightwave Technol. **29**(9), 1326–1334 (2011)

12.113 T. Takagi, H. Hasegawa, K. Sato, Y. Sone, B. Kozicki, A. Hirano, M. Jinno: Dynamic routing and frequency slot assignment for elastic optical path networks that adopt distance adaptive modulation. In: *Proc. Opt. Fiber Commun. Conf./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Los Angeles, Mar. 6–10* (2011), Paper OTuI7

12.114 H. Beyranvand, J. Salehi: A quality-of-transmission aware dynamic routing and spectrum assignment scheme for future elastic optical networks, J. Lightwave Technol. **31**(18), 3043–3054 (2013)

12.115 R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel: Capacity limits of optical fiber networks, J. Lightwave Technol. **28**(4), 662–701 (2010)

12.116 T. Morioka, Y. Awaji, R. Ryf, P. Winzer, D. Richardson, F. Poletti: Enhancing optical communications with brand new fibers, IEEE Commun. Mag. **50**(2), 31–S42 (2012)

12.117 S. Fujii, Y. Hirota, H. Tode, K. Murakami: On-demand spectrum and core allocation for reducing crosstalk in multicore fibers in elastic optical networks, J. Opt. Commun. Netw. **6**(12), 1059–1071 (2014)

12.118 A. Muhammad, G. Zervas, D. Simeonidou, R. Forchheimer: Routing, spectrum and core allocation in flexgrid SDM networks with multi-core fibers. In: *Proc. Opt. Netw. Design Model., Stockholm, May 19–22* (2014) pp. 192–197

12.119 J.M. Fini, T. Taunay, B. Zhu, M. Yan: Low cross-talk design of multi-core fibers. In: *Proc. Conf. Lasers Electro-Opt. (CLEO), San Jose, May 16–21* (2010), Paper CTuAA3

12.120 A. Srivastava, S. Acharya, M. Alicherry, B. Gupta, P. Risbood: Differential delay aware routing for Ethernet over SONET/SDH. In: *Proc. IEEE INFOCOM, Miami, Mar. 13–17*, Vol. 2 (2005) pp. 1117–1127

12.121 S. Ahuja, M. Krunz, T. Korkmaz: Optimal path selection for minimizing the differential delay in Ethernet over SONET, Comput. Netw. **50**(13), 2349–2363 (2006)

12.122 X. Chen, Y. Zhong, A. Jukan: Multi-path routing in elastic optical networks with distance-adaptive modulation formats. In: *Proc. IEEE Int. Conf. Commun. (ICC), Budapest, Jun. 9–13* (2013)

12.123 W. Lu, X. Zhou, L. Gong, M. Zhang, Z. Zhu: Dynamic multi-path service provisioning under differential delay constraint in elastic optical networks, IEEE Commun. Lett. **17**(1), 158–160 (2013)

12.124 X. Chen, A. Jukan, A. Gumaste: Optimized parallel transmission in elastic optical networks to support high-speed Ethernet, J. Lightwave Technol. **32**(2), 228–238 (2014)

12.125 A. Castro, L. Velasco, M. Ruiz, J. Comellas: Single-path provisioning with multi-path recovery in flexgrid optical networks. In: *Proc. 4th Int. Workshop Reliab. Netw. Design Model., St. Petersburg, Oct. 3–5* (2012)

12.126 L. Ruan, N. Xiao: Survivable multipath routing and spectrum allocation in OFDM-based flexible optical networks, J. Opt. Commun. Netw. **5**(3), 172–182 (2013)

12.127 Y. Sone, A. Watanabe, W. Imajuku, Y. Tsukishima, B. Kozicki, H. Takara, M. Jinno: Highly survivable restoration scheme employing optical bandwidth squeezing in spectrum-sliced elastic optical path (SLICE) network. In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), San Diego, Mar. 22–26* (2009), Paper OTh02

12.128 S. Huang, C.U. Martel, B. Mukherjee: Survivable multipath provisioning with differential delay constraint in telecom mesh networks, IEEE/ACM Trans. Netw. **19**(3), 657–669 (2011)

12.129 L. Choy: Virtual concatenation tutorial: enhancing SONET/SDH networks for data transport, J. Opt. Netw. **1**(1), 18–29 (2002)

12.130 G. Bernstein, D. Caviglia, R. Rabbat, H. Van Helvoort: VCAT-LCAS in a clamshell, IEEE Commun. Mag. **44**(5), 34–36 (2006)

12.131 S. Zhang, C. Martel, B. Mukherjee: Dynamic traffic grooming in elastic optical networks, IEEE J. Sel. Areas Commun. **31**(1), 4–12 (2013)

12.132 H. Takara, T. Goh, K. Shibahara, K. Yonenaga, S. Kawai, M. Jinno: Experimental demonstration of 400 Gb/s multi-flow, multirate, multi-reach optical transmitter for efficient elastic spectral routing. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Geneva, Sep. 18–22* (2011), Paper Tu.5.A.4

12.133 M. Jinno, H. Takara, K. Yonenaga, A. Hirano: Virtualization in optical networks from network level to hardware level, J. Opt. Commun. Netw. **5**(10), A46–A56 (2013)

12.134 T. Takagi, H. Hasegawa, K. Sato, Y. Sone, A. Hirano, M. Jinno: Disruption minimized spectrum defragmentation in elastic optical path networks that adopt distance adaptive modulation. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Geneva, Sept. 18–22* (2011), Paper Mo.2.K.3

12.135 A. Castro, L. Velasco, M. Ruiz, M. Klinkowski, J.P. Fernández-Palacios, D. Careglio: Dynamic routing and spectrum (re)allocation in future flexgrid optical networks, Comput. Netw. **56**(12), 2869–2883 (2012)

12.136 K. Wen, Y. Yin, D.J. Geisler, Sh Chang, S.J.B. Yoo: Dynamic on-demand lightpath provisioning using spectral defragmentation in flexible bandwidth networks. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Geneva, Sep. 18–22* (2011), Paper Mo.2.K.4

12.137 F. Cugini, F. Paolucci, G. Meloni, G. Berrettini, M. Secondini, F. Fresi, N. Sambo, L. Potì, P. Castoldi: Push–pull defragmentation without traffic disruption in flexible grid optical networks, J. Lightwave Technol. **31**(1), 125–133 (2013)

12.138 R. Wang, B. Mukherjee: Provisioning in elastic optical networks with non-disruptive defragmentation, J. Lightwave Technol. **31**(15), 2491–2500 (2013)

12.139 A.S. Muqaddas, M. Garrich, P. Giaccone, A. Bianco: Exploiting time synchronized operations in software-defined elastic optical networks. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Los Angeles, Mar. 19–23* (2017), Paper W4J.6

12.140 A.N. Patel, P.N. Ji, J.P. Jue, T. Wang: Defragmentation of transparent flexible optical WDM (FWDM) networks. In: *Proc. Opt. Fiber Commun./Natl. Fiber Opt. Eng. Conf. (OFC/NFOEC), Los Angeles, Mar. 6–10* (2011), Paper OTuI8

Jane M. Simmons

Monarch Network Architects
Holmdel, NJ, USA
jsimmons@monarchna.com

Jane M. Simmons has been involved in the research and development of optical networks for 20 years. She is the founder of Monarch Network Architects. Prior to founding Monarch, she was the Chief Network Architect of Corvis Corp. and worked at Bell Labs/AT&T Labs Research. She received a BS, from Princeton University, and SM and PhD degrees from MIT, all in Electrical Engineering.

George N. Rouskas

Dept. of Computer Science
North Carolina State University
Raleigh, NC, USA
rouskas@ncsu.edu

George N. Rouskas is the Director of Graduate Programs in the Computer Science Department at North Carolina State University. He received a degree in Computer Engineering from the National Technical University of Athens (NTUA), Greece, and MS and PhD degrees in Computer Science from the College of Computing, Georgia Institute of Technology. His research interests include network architectures and protocols, optical networks, network design and optimization, and performance evaluation.

13. Standards for Optical Transport Networks

Stephen J. Trowbridge

Most larger optical networks are built using a combination of standardized and proprietary technology.

This chapter provides information on how a combination of complete and functionally standardized optical interfaces are used to build an optical transport network OTN.

Ultra-long haul (1000s of km), subsea, and advanced high-bit rate ($>100\,\mathrm{Gb/s}$ per λ) coherent links in optical networks are generally *bookended* by single vendor equipment that may involve methods such as constellation-shaping and high-gain, usually proprietary, soft-decision forward error correction (FEC) to achieve the reach or signaling rates required. For these kinds of links, it is important to optimize link performance without incurring the *interoperability* penalty that generally arises from accommodating a variety of implementations in transmitter and receiver designs that have to work with each other. Burying a new fiber over thousands of km or laying a subsea cable is so costly, that it is important to squeeze every possible bit through the fiber that is already in place before incurring the cost of new fiber installations over these distances.

Even for implementations using proprietary technology to achieve maximum performance, a certain amount of standardization is required. This is referred to as functional standardization, which includes:

- A consistent set of mappings of client signals into the frame format of OTN. This allows for a client that is mapped into OTN using the equipment of one vendor to be demapped by the equipment of another vendor after the signal has been forwarded over an inter-domain *handoff* interface.
- A consistent equipment functional model is provided, including the information content (but not necessarily the format) of the signal overhead and how it is processed. This assures that equipment of different vendors can be managed or controlled in the same way to provide a consistent approach to configuration, connection establishment, fault detection, fault isolation, and measurement of network performance.

There are also fully standardized interfaces in the OTN, including:

- The interfaces for the client signals carried over the optical network (e.g., Ethernet, SONET (synchronous optical network)/SDH (synchronous digital hierarchy), fibre-channel, etc.)
- OTN line interfaces for shorter reaches or lower signaling rates where technology is more mature, market deployment volumes are higher, and the technology is not being pushed to the limits of its

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_13

capability (e.g., metro and some data-center interconnect applications).
- Inter-domain *handoff* interfaces. OTN frame formats are defined to be carried over Ethernet pluggable optical interfaces for interconnection between equipment of different vendors or between networks of different operators.

13.1 Optical Network History

Over the last 40 years, there has been remarkable transformation in the architectures, technologies, and the services carried over networks.

In the 1970s, traffic carried over networks consisted primarily of 64 kb/s services including voice, FAX, and modem data. Interconnect and switching were primarily the domain of telecom network operators. There was a clear delineation between a *line* (service interface, generally to an end user or computer host) and a *trunk* (a switch-to-switch interface). Aggregation to higher rates generally occurred on the trunk interfaces only. Transmission networks generally employed copper and microwave links. The digital frame formats for aggregate trunk interfaces were referred to as the plesiochronous digital hierarchy (PDH). There were two distinct hierarchies built from similar technologies that evolved in North America and in Europe. The North American (24-channel) hierarchy was based on DS1 (1.5 Mbit/s) and DS3 (45 Mbit/s) interfaces, while the European (30-channel) hierarchy included E1 (2 Mbit/s), E3 (34 Mbit/s), and E4 (140 Mbit/s) interfaces, with the higher-rate interfaces generally carried over coaxial cables.

Initial lightwave systems started to emerge around 1975 based on GaAs semiconductor lasers (850 nm) over multimode fiber. The first ITU-T Recommendations were G.651 (*Characteristics of multimode fiber*) and G.955 (originally G.956) (*Characteristics of optical systems operating at 850 nm suitable for the transport of PDH signals of 34–45 Mbit/s*). These early technologies provided point-to-point trunks between service switches, which generally implemented a switching granularity of 64 kb/s.

Synchronous optical transport networks (SONET and SDH) began emerging in the mid-1980s. This was enabled by several key technologies:

- Single-mode fiber operating initially at 1300 nm, later at 1550 nm, with fiber types documented in ITU-T Recommendations G.652, G.653, and G.655.
- Improved clock and PLL (phase-locked loop) technology enabled a fully synchronous network with timing of all nodes traceable to the same primary reference clock (PRC).

These technologies, together with the fact that the transmission rate of 50 Mb/s grew eventually to 40 Gb/s were much greater than the (initial) service rates of 64 kb/s enabled the birth of what has become known as the *transport network*. This is a network that allows the trunks between service switches to be flexibly and dynamically configured rather than being point-to-point physical links. The initial clients of the transport networks were PDH services ranging from 1.5/2 Mb/s to 140 Mb/s. As with PDH, there were initially differing standards for North American and European networks. SONET, initially described in Bellcore (now Telcordia, part of Ericsson) GR-253 and ATIS T1.105 (now ATIS-0900105.2008), describes the North American implementation building from the 24-channel North American interfaces. ETSI EN 300 417 described the European implementation building from the 30-channel European interfaces. ITU-T Recommendation G.707 [13.1] later described a superset of the North American and European hierarchies based on common physical interfaces. This was primarily intended to allow equipment vendors to produce common products to sell into both markets with the appropriate configuration. Cases where the same network element would support some interfaces configured using the North American hierarchy and some interfaces configured using the European hierarchy were rare.

This split between *transport network* and *service network* gave telecom network operators planning independence between service network capacity and offerings and the transmission capacity they built into their networks. Operators could also offer *private line* services to enterprise customers delivered across their transport networks; e.g., they could sell a dedicated DS1 service from San Francisco to New York to an enterprise customer, who might use this as a trunk in a company internal network between PBX (private branch exchange) switches, or as a dedicated data interconnect pipe.

The North American implementation began somewhat earlier, starting from a basic building block of a 51.84 Mbit/s frame structure referred to as STS-1, while the European implementation began with a basic building block of a 155.52 Mbit/s frame structure referred to as STM-1. The two hierarchies evolved to

Table 13.1 Common signaling rates for SONET/SDH

SONET	SDH	Bit rate (kbit/s)
STS-1	STM-0	51 840
OC-3	STM-1	155 520
OC-12	STM-4	622 080
OC-48	STM-16	2 488 320
OC-192	STM-64	9 953 280
OC-768	STM-256	39 813 120

use a common set of physical layer interfaces (even though the naming of the hierarchal levels was a factor of three different ones). The European hierarchy later added a hierarchal level known as STM-0 to match the lowest transmission rate in the North American hierarchy. The common physical layer signaling rates for SONET and SDH are indicated in Table 13.1.

The 1990s featured an evolution of SONET and SDH not only to higher signaling rates, but to supporting a wider variety of client data streams. The aggregated services were no longer just voice/FAX/modem over PDH signals aggregating 64 kbit/s channels. The key features to make SONET and SDH more *data friendly* included:

- Virtual concatenation (VCAT), which enabled creating larger client containers by logically *gluing* together multiple smaller containers to form a container that was *right sized* for a given client.
- The link capacity adjustment scheme (LCAS) allowed adding or removing members of a VCAT group in service without creating a hit to packet-based traffic.
- The generic framing procedure (GFP) [13.2] was introduced for mapping of non-PDH clients. Frame-based GFP-F is used for mapping of Ethernet frames, asynchronous transfer mode (ATM) cells, and IP/PPP packets. Transparent GFP-T is used for mapping of constant bit-rate clients (full-rate Ethernet, enterprise systems connection (ESCON), fibre connection (FICON), fibre channel, etc.).

Across a similar time frame in data local area networking, Ethernet evolved to become the dominant technology. This was introduced with a signaling rate of 10 Mb/s in 1980, growing to 100 Mb/s in 1995, 1 Gb/s in 1998, and 10 Gb/s in 2003. More is discussed about Ethernet rate evolution later. When there was a desire to extend Ethernet reach across a transport network, most initial deployments were subrate, as transport network capacity was relatively expensive. The transport network connection would have an Ethernet bridge at each end and a smaller transport link in the middle. A 10 Mb/s Ethernet signal bridged over a DS1 private line connection would be able to carry all of the packet traffic from that DS1 up to about 15% link occupancy. Gigabit Ethernet could be carried *full rate* over an SDH network by transcoding the 8B/10B encoded Ethernet signal to a GFP-T mapped 64B/65B signal and then mapping over a V4C-7V, virtually concatenating seven VC4 (155 Mb/s) SDH containers to carry the 1 Gb/s payload. For 10 GbE, IEEE 802.3 defined a special *10 GBASE-W* signal with a frame-format compatible with SONET OC-192. But until an accident of arithmetic from the historical 4× evolution of transport networks and the historical 10× evolution of data networks brought Ethernet rates and transport rates very close together, the market had never really considered that there was a reason for Ethernet rates and transport network rates to be the same, or that bridging together of local-area networks over wide-area networks needed to occur using the same transmission speeds as used within the local area network. One advantage of having an Ethernet rate and a transport rate at 10 Gb/s was that the transport network client optics and Ethernet optics could use common components, e.g., the same SFP pluggable module might support 10 GBASE-R, OC-192/STM-64, and later OTU3 client interfaces.

The next evolutionary step in optical networking occurred with the introduction of wavelength division multiplexing (WDM). This enabled increasing fiber capacity by putting multiple signals on different wavelengths (colors) of light on the same fiber. This decreased the cost of long-haul optical spans by using optical amplification (initially EDFA (erbium-doped fiber amplifiers)) for multiple-wavelength signals.

Initial WDM systems operated by multiplexing different clients (at the time, primarily SONET/SDH) in their native formats over different wavelengths of light. WDM for SDH signals was specified in ITU-T Recommendation G.692 (1998).

OTN standardization was driven by the fact that optical multiplexing fell short of the robustness normally expected in transport networking layers (e.g., error monitoring, path trace, fault isolation, and protection switching). Inherent monitoring of the clients at the network edge could not isolate faults. Client specific nonintrusive monitoring is cumbersome, expensive, and impractical. A variety of proprietary approaches emerged using techniques including pilot tones, nonassociated overhead, and digital wrappers to manage wavelengths on vendor specific WDM systems.

The goals of OTN standardization for the first version of ITU-T recommendation G.709 (2001) included:

- Carrier grade transport networking layer for optically multiplexed signals on a fiber
- Service transparency for SDH/SONET, Ethernet, ATM, IP, MPLS, and future clients

- Enhanced client-independent OAM and networking functionality for all services
- Gigabit/multigigabit level bandwidth granularity required to scale and manage multiterabit networks
- Wavelength level switching maximizing nodal switching capacity, the gating factor for reconfigurable network capacity
- Avoiding large numbers of fine granularity pipes that stress network planning, administration, survivability, and management
- First transmission technology in which each stakeholder gets its own connection monitoring (tandem connection monitoring)
- First transmission technology initially developed as a global standard: ATIS and ETSI explicitly deferred to ITU-T, so that a standard could be developed without North American and European regional variants
- The initial vision was one wavelength (λ) per client. But as λs evolve from 10 to 40 to 100 Gbit/s, efficiency dictates that you do not spend an entire λ on a low-rate client, and it is easier to build equipment with known physical layer bit rates. Therefore, OTN also includes TDM multiplexing.

The first version of G.709 in 2001 was optimized around transport of SONET/SDH, focusing on the higher-rate signals predominately used in long-haul transmission systems at the time (2.5, 10, and 40 Gb/s, corresponding to SONET OC-48, OC-192, and OC-768 or SDH STM-16, STM-64, and STM-256) [13.3].

The second major evolutionary step in OTN standardization occurred around 2010. This occurred concurrently with the IEEE P802.3ba 40 Gb/s and 100 Gb/s Ethernet project. Several market forces combined to shift the evolution of OTN standards away from the historical focus on telecom network clients and applications and away from the historical 4× steps for bandwidth growth.

Several factors affected this next phase of OTN standards development:

- While 10 G transport networking was being deployed in large volumes, the market for 40 G transport networking was challenging. In contrast to 10 G, where the client interfaces were simply reuse of Ethernet optics, 40 G client interfaces used a transport-unique VSR2000-3R2 interface that was relatively expensive. On the line side, nonreturn-to-zero (NRZ) modulation became challenging, often requiring complex link design with dispersion accommodation. A variety of new modulation formats evolved to make 40 G transmission easier, but largely this just fragmented the market and made it even more difficult to achieve economies of scale. There was initially no Ethernet client at 40 G, although there was a limited amount of proprietary 40 G POS (packet-over-SONET, essentially PPP (point-to-point protocol) mapping of packets over a SONET OC-768 frame). By the time that coherent modulation formats (initially dual-polarization quadrature phase-shift keying (DP-QPSK) emerged to make 40 G link planning easier, 100 G links were nearly as easy.
- There were no plans or proposals to evolve SONET/SDH transmission rates beyond 40 Gb/s, whereas Ethernet was actively standardizing a new rate of 100 Gb/s.
- While Ethernet was rapidly supplanting SONET/SDH as the most important client of optical networks, there was significant diversification of uses for optical networks. Storage-area networking, often provided by fibre-channel over OTN, became important, with a big surge after September 11, 2001, when many companies realized it might not be a good idea to have all copies of their data within a small geographical area, e.g., concentrated in lower Manhattan (New York City). Video distribution and mobile fronthaul/backhaul were also important markets.

The next OTN evolutionary step included a new top-end line rate of 100 Gb/s (OTU4), rather than what might have been a more historically consistent 4× next step to 160 Gb/s [13.4]. A new ODU0 container was added optimized for transport of 1 GbE using half the amount of network capacity as OC-48/STM-16. A generalized *ODUflex* container was added to be able to support virtually any client signal rate for any application within a set of fully standardized line rates. This set of steps, together with coherent DP-QPSK modulation becoming an efficient way to transmit 100 Gb/s per λ in a 50 GHz grid all the way from metro reaches (a few hundred km) to as much as 4000 km, resulted in almost ubiquitous 100 Gb/s networking for new network deployments. Up to about 12 000 km (trans-Pacific undersea cable reaches), each 100 Gb/s is generally carried using two 50 Gb/s wavelengths of binary-phase-shift-keying (BPSK) modulation, but the digital frame formats used were generally 100 Gb/s inversely multiplexed over two subcarriers, so from a network management perspective, it appeared as though everything was carried at 100 Gb/s.

Unfortunately, this nice uniformity in network design breaks beyond 100 Gb/s due to the Shannon limit. It is possible to carry more than 100 Gb/s per λ by increasing the baud rate, and by increasing the constellation complexity. For example, one can carry 200 Gb/s

per wavelength using DP-16QAM modulation at the same baud rate as 100 Gb/s DP-QPSK, but the maximum optical reach is reduced from around 4000 km to around 1000 km. How much information can be carried on a wavelength varies depending on how far the information needs to be carried, and there is no single *next* line rate beyond 100 G. This drove a different paradigm for OTN beyond 100 G:

- The digital frame format is decoupled from the bit rate per λ. While in many cases a single digital frame is carried over a wavelength, in the general formulation, a digital frame is carried by what is called an optical tributary signal group (OTSiG) with one or more group members. The same digital information might be carried over groups of different sizes depending on the distance it needs to be carried. For example, a 400 Gb/s digital frame might be carried over a single λ up to about 200 km, or carried over two λs up to about 1000 km, or over four λs up to about 4000 km.
- The digital frame can be created in virtually any size to meet network needs. As many of the clients to be carried over optical networks are multiples of 100 G (e.g., higher-speed Ethernet interfaces include 100 G, 200 G, and 400 G), the primary formulation is described with a frame structure called OTUC*n* (C = 100), of approximately $n \times 100$ Gb/s in bit rate. The payload and the overhead scale with the multiplier *n*. The contents of the digital frame may be a single client (e.g., an OTUC4 carrying a 400 GBASE-R Ethernet client), or multiple clients, TDM multiplexed. The internal structure of the OTUC*n* frame consists of $20 \times n$ tributary slots of approximately 5 Gb/s capacity, which can be flexibly allocated to the clients carried in the OTUC*n*. But frame formats that are not multiples of 100 G can also be defined, reflecting the fact that not all coherent modulation formats result in λ rates that are multiples of 100 G. Some nonmultiple of 100 G λ rates may occur naturally based on the constellation chosen. For example, if DP-QPSK supports 100 G per λ, then DP-16QAM at the same baud rate supports 200 G per λ. So, DP-8QAM at that same baud rate (which would likely go further than 1000 km but less than 4000 km) would support 150 G. Techniques like probabilistic constellation shaping allow for even finer granularities of bit rates supported per λ. While in some cases, it may be possible to support multiple of 100 G digital frame formats without multiple of 100 G λ rates (e.g., an OTSiG composed of two 150 G λs could carry an OTUC3), this is not always possible or practical. As a result, the beyond 100 G OTN formulation allows creating a container OTUC*n*-*M*, where *n* is the number of 100 G equivalent units of overhead, and *M* is the number of 5 G tributary slots supported in the frame, where $M < n \times 20$. For example, a 150 G container could be described as an OTUC2-30, with the same amount of overhead as a 200 G container, but supporting only 30 of the possible 40 tributary slots of payload capacity. Flex Ethernet, described later in this chapter, is one technique for efficiently using the bandwidth of these kinds of odd-sized digital containers.

13.2 OTN Signal Format

Optical transport networking (OTN) is a network technology that was originally developed for Telco network applications as part of the transport network providing configurable trunks between service network switches. This began as an extension of the existing SONET/SDH transport networks into WDM infrastructure, managing the network capacity in terms of digital frames carrying network clients and managing wavelengths in the network.

OTN has since been applied to many different network applications, including internet backbones, data center interconnect, storage-area networking, video distribution, and mobile network fronthaul/backhaul.

OTN has an unfortunate reputation as being complex, as features have been designed and incorporated for a wide range of network applications. Indeed, a network element designed to implement all of the features for the superset of all possible network applications would be quite complex. While large OTN crossconnect systems with a rich feature set are available, it is not necessary to build networks for simple applications out of *god boxes*. OTN is better thought of as a tool box, where simple tools can be used for simple network applications, and the more complex features are used only where necessary to address complex network scenarios. Many smaller, simpler OTN network elements are available, from pluggable transponders in small form factors to pizza boxes, with many vendors providing targeted offerings for specialized network applications. There would be no need, for example, to use a large OTN crossconnect system in a network for data-center interconnect over a point-to-point amplified WDM line system.

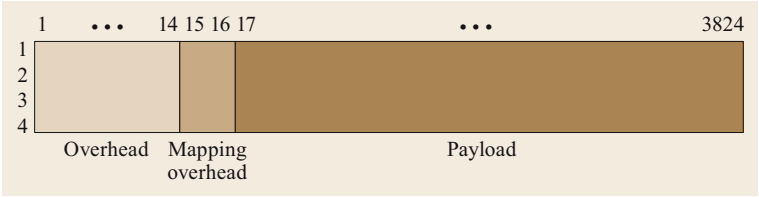


Fig. 13.1 Digital wrapper (ODU*k*) for an OTN client signal

13.2.1 Digital Wrapper

A key concept for understanding OTN is the *digital wrapper*. This wrapper, or frame format is used for two purposes: to add overhead for the purpose of managing a client signal carried over the network and to manage a wavelength in the network (or group of wavelengths carrying a single digital signal).

The wrapper carrying a client signal is illustrated in Fig. 13.1. In the standard, this wrapper is referred to as an ODU*k* (optical data unit of order *k*, where a numeric value of *k* refers to one of several specific, discrete container sizes, and *flex* refers to a container of an arbitrary size, generally tailored to the transport of a specific client).

The *row, column* format is used to assign meaning to particular bytes according to their *position* in the frame, but when transmitting across a fiber, the 15 286 bytes of the frame are transmitted left to right across the 3824 columns, then top to bottom across the four rows, followed by transmission of the next frame in the sequence.

The client signal is mapped into the payload area of this frame. There are a variety of mapping procedures that are used for different clients, but for illustration, we will first assume a bit-synchronous mapping, where 1 bit of the client signal fills 1 bit of the payload area of the OTN frame, and the entire payload area is filled. The bit rate of the wrapped signal in this case is the bit rate of the client signal multiplied by a ratio of 239/238. The added overhead occupies about 0.42% of the necessary bits to transmit the signal.

Conceptually, what we are trying to do is to send a client signal across a network. This can be envisioned as shown in Fig. 13.2. For ease of explanation, one direction of transmission is shown, although normally connections across an OTN network are bidirectional and symmetric.

The network may be simple or complex. It may be just a point-to-point WDM line system or it may be a cascade of mesh networks of different network operators. Within the network, the client may be carried on its own wavelength or it may be multiplexed with other traffic on wavelengths throughout the network. The service provided by the network may be a simplex service (e.g., the service is down if any of the network links carrying it fail) or it may be a protected service with alternate, geographically diverse routes. The underlying complexity of the network is largely hidden from the view of the client service being supported, although attributes such as whether the service is protected may be reflected in availability parameters in a service level agreement between a network operator and their customer.

In Fig. 13.1, the *mapper* function of the network ingress port maps (or fills in) the client signal into the payload area of the digital wrapper and inserts the path overhead. The demapper in the network egress port processes the path overhead and extracts the client signal from the frame.

13.2.2 Signal Monitoring Overhead

Figure 13.3 provides more detail on different areas of the overhead. Not all of the overhead is used for all applications.

The alignment overhead has the format shown in Fig. 13.4. This consists of a 6 byte fixed bit pattern that allows finding the start of the frame and a *multiframe alignment signal* (MFAS) that is simply a single byte counter that cycles from 0–255 to allow putting multibyte values into single-byte overhead fields.

The overhead used for nearly all applications is called the path monitoring overhead (PM). The location of this overhead is illustrated in Fig. 13.5.

The path monitoring information consists of 3 bytes intended to answer two questions about the service across the network: is the information connected to the correct (intended) place in the network and is the information delivered without errors? The first of these

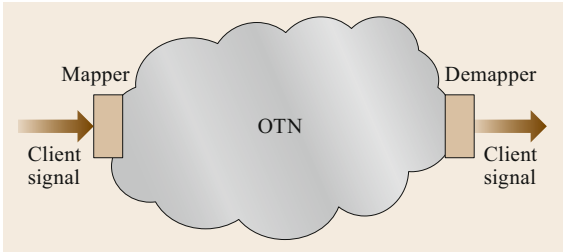


Fig. 13.2 Illustration of a client signal carried across a network

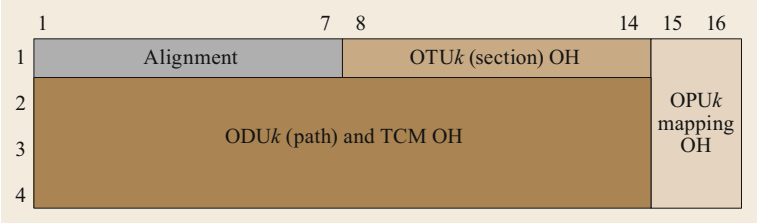


Fig. 13.3 Types and location of overhead

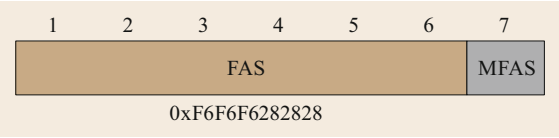


Fig. 13.4 OTN frame alignment

questions is answered using the trail trace identifier (TTI). This is a 64 byte character string that identifies the signal in the network. This character string is inserted into the first byte of the path monitoring overhead across a sequence of 64 OTN frames, using the lower-order 6 bit of the multiframe alignment counter to identify the byte positions. In Telco networks, this character string has a standardized format, including source and destination identifiers including country codes and network identifiers plus operator-specific information. But for non-Telco applications, it is just a 64 byte identifier that can be used by the management system or SDN controller to verify correct connectivity and to prevent, for example, accidentally delivering the data from one customer to the wrong customer at the other end of the connection.

The second byte is used for verifying the integrity of the information sent in the direction of transmission and contains a bit-interleaved parity (BIP) calculation across the previous OTN frame. This can be used to collect performance monitoring information by counting errors over time, or simply to expect that this value should be zero in the normal case of error-free transmission.

The third byte is used predominately for sending information in the backward direction from the demapper toward the mapper. The backward defect indication (BDI) bit is used to indicate a signal fail condition detected from the far end of the network. The backward error indication (BEI) bits are used to indicate how many BIP errors are being experienced by the far end (so the near end is able to count far-end errors). The STAT bits are used for a few other maintenance signals like the alarm indication signal (AIS) used to avoid a cascade of alarms resulting from a single failure early in a multihop connection.

This digitally wrapped client now needs to be carried across the network. Physically, the network consists of a set of optical fibers interconnecting network elements. A network element may or may not have switching flexibility; for example, an inline optical amplifier would not have any switching flexibility, whereas an add/drop multiplexer or crossconnect system would. Switching can be done in the digital (or electrical) layers or in the optical layers (switching of wavelengths). We will examine first the simpler case where any switching happens in the digital layers, so for now, we

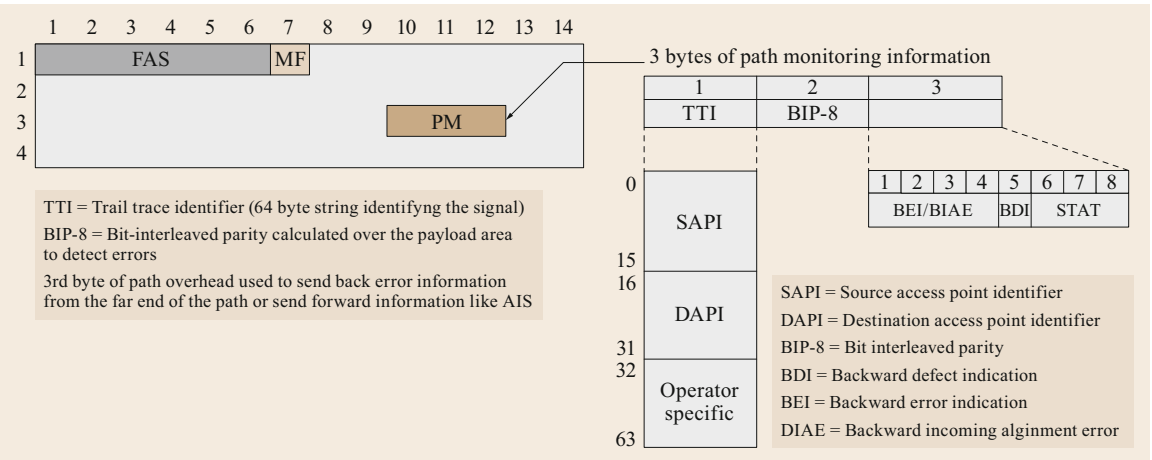


Fig. 13.5 Path monitoring information

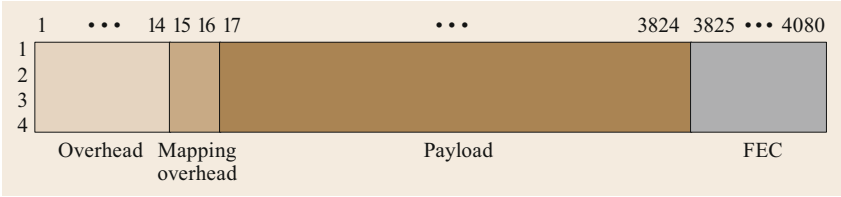


Fig. 13.6 OTU*k* frame format

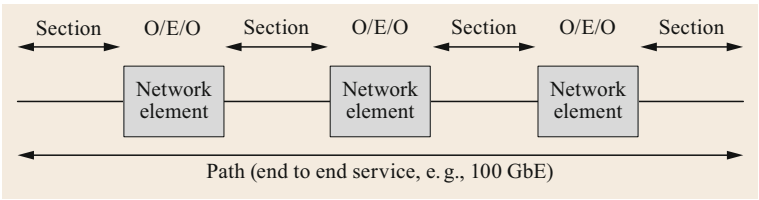


Fig. 13.7 Illustration of section and path

will assume that any network element in the middle of the path with switching flexibility has O-E-O (optical-electrical-optical) kind of processing, and the switching happens electrically.

A larger client (e.g., 100 Gb/s Ethernet) might be carried on its own wavelength, and a smaller client (e.g., 1 Gb/s Ethernet) might be multiplexed with other traffic on a wavelength. We will first discuss the case for wavelength signaling rates up to 100 G, and cover beyond 100 G a bit later.

Up to 100 G, wavelengths could use one of several discrete signaling rates (Sect. 13.1). The initially standardized rates of 2.5 G, 10 G, and 40 G were chosen around the most important SONET/SDH clients of the time. Transport of Ethernet clients reused the 10 G and 40 G signaling rates, and a signaling rate of 100 G was added for transport of Ethernet 100 GBASE-R (no SONET/SDH client at this rate). These four signaling rates were named hierarchical transmission rates using a digital frame format called the optical transmission unit of order *k* (OTU*k*), with OTU1 representing a 2.5 Gb/s signal and OTU4 representing a 100 Gb/s signal. The frame format for the OTU*k* is exactly the same wrapper format we have already seen, augmented with some physical layer transmission aspects like space for a forward error correcting (FEC) code as illustrated in Fig. 13.6. Note that some (vendor specific) soft-decision FEC codes may be framed independently of the OTU*k* rather than just framed as extra columns in the OTU*k* frame.

In the simple case, a single client is carried on the wavelength; the wrapper to manage the client and the wrapper to manage the wavelength are exactly the same wrapper, just extended with the FEC bytes and using some more fields in the overhead.

To further drill into the internals of the network, consider the case where the service is carried over a cascade of links and network elements as illustrated in Fig. 13.7.

In this case, it is no longer sufficient to know whether the service is up or down, or whether bit errors are occurring, but we would also like to know where along the path the failure has occurred, or if bit errors are being introduced, and which one of the sections is the poorly performing link responsible for the errors. For this, we can use the section monitoring overhead, whose location is indicated in Fig. 13.8.

The format and semantics of the section monitoring information is exactly the same as that of the path monitoring information, but the span is a single link between OEO points in the network. The TTI now identifies the internal points in the network where the link is connected. The BIP is calculated across the digital frame transmitted from a network element without considering any possible errors that may have been introduced upstream of that network element. Note that the section monitoring does not need to be used in the case that the service is only carried over a single optical span (e.g., in a point-to-point WDM line system).

Now consider a more complex network: for example, where the service is carried across the networks of multiple network operators, and you would like to isolate in which operator's network a failure or degradation occurs, or where there is a protected service with diversely routed working and protection paths, and a desire to independently monitor the service status and

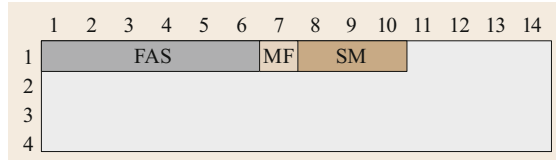


Fig. 13.8 Location of section monitoring (SM) overhead

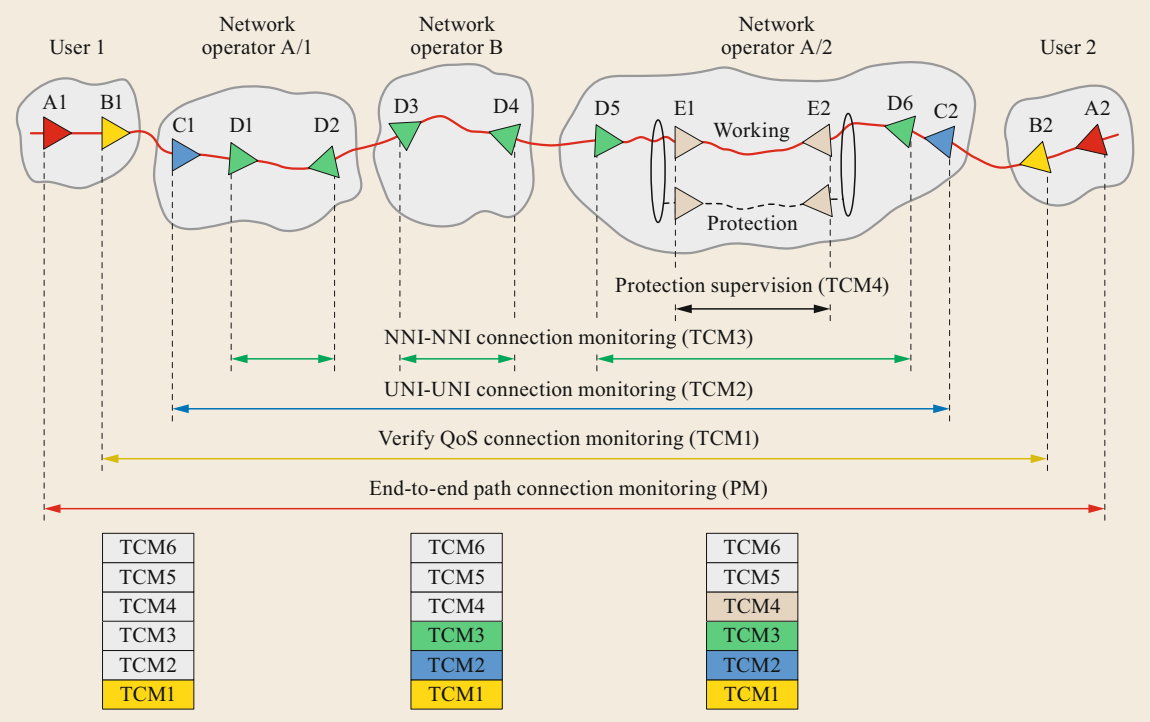


Fig. 13.9 Illustration of tandem connection monitoring

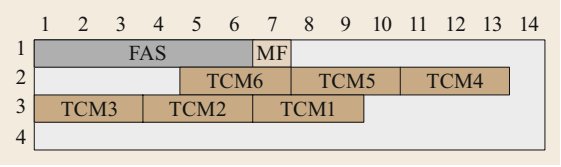


Fig. 13.10 Location of TCM overhead bytes

quality of the working and protection paths. OTN can address this as well, with a feature called tandem connection monitoring (TCM). Maintenance entities can be instantiated at any physical point in the network where monitoring of the connection is desired. An example of tandem connection monitoring is illustrated in Fig. 13.9.

Space for up to six sets of TCM overhead bytes is provided in the overhead area of the OTN frame as illustrated in Fig. 13.10, and each TCM level has 3 overhead bytes that have exactly the same format and semantics as the 3 bytes of path monitoring or 3 bytes of section monitoring that we are already familiar with.

Of course, if the network is simple and is managed within a single administrative domain, tandem connection monitoring is not required at all, but the feature is available if the complexity of the network requires it.

13.2.3 Other Overhead

Additional overhead is defined for various purposes, not used in all network scenarios. There are 6 bytes allocated for communication channels with the extent of the section or path that may be used when there is in-band management and control plane (e.g., SDN controller) communications to the network elements. There is one byte that may be used for a round-trip delay measurement feature to assure that a path is suitable for latency sensitive traffic. There are 4 bytes that may be used for protection switching, for example, SONET/SDH style linear or ring protection for operators that need 50 ms protection switching times. And there is one byte for a synchronization messaging channel that may be used in the case that an OTN interface is used for transport of time or frequency, and it is necessary to carry precision time protocol (PTP) or synchronization status message (SSM) messages over the interface.

13.3 Digital Multiplexing

The next case to consider is where the network clients are smaller than the wavelengths on which they are carried, and multiple clients are carried on wavelengths in the network. This multiplexing may happen only at the endpoints (e.g., muxponder applications, such as mapping ten Ethernet 10 GBASE-R signals over a 100 Gb/s wavelength), or intermediate nodes such as OTN cross-connect systems may groom lower-order traffic by demultiplexing clients from one interface and remultiplexing them onto another. This latter application is often what people mean when they use the term *OTN switching*, although switching can occur at several different layers (digital or optical) in the OTN hierarchy.

When multiplexing is used in the OTN, the wavelength frame format is precisely what we saw earlier in Fig. 13.6 and is referred to as the *higher-order* OTU*k*/ODU*k*, but the payload area contains a multiplex structure rather than a single client. The wrapped clients that are multiplexed inside of the higher-order container are referred to as *lower-order ODUk*. In this section, only the cases of higher-order OTU*k*/ODU*k* for $k = 1, 2, 3, 4$ (from 2.5 G through 100 G per wavelength) are discussed, while the *beyond 100 G* extension for OTN is discussed later.

The payload area of the higher-order ODU*k* is divided into several discrete units of bandwidth referred to as tributary slots (TS). One TS carries approximately 1.25 G of bandwidth. Any number of tributary slots may be assigned to a given lower-order ODU*k* depending on the bandwidth requirements. Table 13.2 indicates the number of 1.25 G tributary slots contained in each of the defined discrete wavelength rates. The tributary slots are organized within the payload area by alternating columns of the payload area across the number of tributaries into which the higher-order container is divided.

For example, to carry a 10 G client multiplexed inside of a 40 G wavelength, one would assign 8 of the available 32 tributary slots to that client. The client itself is contained in its own *wrapper* according to the format of Fig. 13.1, and that wrapped signal (including the lower-order frame alignment, overhead, and payload) is inserted into the tributary slots of the higher-order ODU*k*. The frame alignment signal of the lower-order ODU*k* is used to find the frame position of the client, which effectively *floats* and is not locked to the frame position of the higher-order ODU*k*. Note that the payload area of the higher-order ODU*k* is scrambled using a frame-synchronous scrambler, and the byte positions assigned to the lower-order ODU*k* are not generally adjacent in the higher-order frame, so the frame alignment signal for the lower-order ODU*k* will not *spoof* the frame alignment for the higher-order ODU*k*. Either an asynchronous or generic mapping procedure (GMP) mapping is used to insert stuff bytes as necessary to accommodate the difference between the space available in the tributary slots and the size of the wrapped lower-order container (Sect. 13.4).

Table 13.2 Number of 1.25 G tributary slots per wavelength

Higher order ODU*k*	# of 1.25 G TS
ODU1 (2.5 G)	2
ODU2 (10 G)	8
ODU3 (40 G)	32
ODU4 (100 G)	80

13.4 Client Mapping Methods, Tributary Slot Mappings, and Justification

The example of the wrapper described in Sect. 13.2.1 was a simple bit-synchronous wrapper, where 1 byte of the client filled 1 byte of the payload area of the OTN frame, and every byte of the payload area was filled. A consequence of this kind of mapping is that the physical layer clock of the ODU*k* is locked to the physical layer clock of the client, and the clock of the client must be at least as good as the clock required for OTN interfaces (±20 ppm from nominal).

But there are many cases of optical network clients where not every byte of the payload area is filled:

- The client may be smaller than the chosen discrete wavelength rate that has been selected to carry it.
- There may be performance or architectural reasons that the OTN container clock cannot be locked to the client clock; for example, most Ethernet clients have a ±100 ppm clock tolerance, so you would not be able to derive an OTN container clock within ±20 ppm of a nominal rate from the Ethernet clock.
- The client may not even be a constant bit-rate signal–packet traffic needs to have the space in the frame between packets filled with some kind of idle characters to distinguish the packets from what surrounds them.

Different mapping procedures are defined to address these various cases for different clients. Columns

15–16 in the OTN frame are used for any information that is needed about the way in which the client is mapped for the demapper to properly extract it from the frame.

The byte in row 4 of column 15 is referred to as the payload structure indicator (PSI). This is organized as 256 bytes that are sent repeatedly according to the index provided by the multiframe alignment signal (Fig. 13.4). The byte that is sent when MFAS = 00000000 is referred to as the payload type (PT). The payload type is used to indicate what kind of client is mapped into the OTN frame. It is important to assign a different payload type value whenever two different clients might be carried in the same size frame. For example, there is no need to have a different payload type for SDH STM-16 and SDH STM-64, as the containers are of different sizes and cannot be interconnected. But it would be necessary to distinguish whether an ODU0 carries 1 GbE, FC-100, STM-1, or STM-4, which are fundamentally different clients that are possible to map into the same size container. The remaining 255 bytes are available, when necessary, to describe the structure of the client (or of the payload). These bytes are currently defined for the case of digital multiplexing to convey how the different lower-order signals are mapped into the tributary slots of the higher-order signal.

13.4.1 Mapping Mechanisms for Constant Bit-Rate Signals

Bit-Synchronous and Asynchronous Mapping Procedures (BMP and AMP)

Bit-synchronous or asynchronous mapping may be used when the clock of the client is constrained to a very narrow range, and the difference between the size of the client and the size of the OTN payload area is not large. This is largely a legacy mapping procedure used for some initial OTN applications, but largely replaced by the generic mapping procedure (GMP) described in the next section for most newer mappings. The uses of AMP and BMP include the client mappings of STM-16/OC-48 into ODU1, STM-64/OC-192 into ODU2, and STM-256/OC-768 into ODU3. AMP is also used for the tributary slot mappings of lower-order ODU0 into higher-order ODU1, lower-order ODU1 into higher-order ODU2 or ODU3, and lower-order ODU2 into higher-order ODU3.

Fixed stuff bytes may be prescribed in certain exact positions in the payload frame to compensate (all or most) of the difference between the size of the client and the payload area.

In the case of bit-synchronous mapping, other than any defined fixed stuff bytes, the client fills all other bytes of the payload area, and the clock for the OTN container is derived from, and locked to, the frequency of the client by a fixed ratio. Asynchronous mapping decouples the clock of the client from the clock of the OTN container, and justification bytes are inserted to compensate the variable difference between the client and the OTN container clock.

Both bit-synchronous and asynchronous mappings use the same definition of the mapping overhead, and both can use the same demapper. The mapping overhead locations are illustrated in Fig. 13.11.

The bytes NJO, PJO1, and PJO2 are called (negative and positive) justification opportunity bytes, and they may contain client data or contain a stuff byte to make adjustments to account for the variation between the client and OTN container clocks. If the client clock is temporarily running above nominal, or the OTN container clock is temporarily running below nominal, it may be necessary to occasionally insert an extra payload byte into the negative justification opportunity (NJO) byte position to absorb the excess payload bytes. If the reverse occurs (the client is temporarily running below nominal, or the OTN container temporarily above nominal), the occasional shortage of payload bytes to fill the frame is compensated by sending stuff bytes rather than payload bytes into one or both of the positive justification opportunity (PJO) bytes. The PJO2 byte is not needed for all mappings.

The justification control (JC) bytes provide justification control; all three of these bytes are encoded the same way, and error correction for the justification information is provided through majority vote across the three copies of the information. The encoding of the justification control information is as shown in Fig. 13.12.

Bit-synchronous mapping is encoded identically to asynchronous mapping but with the JC bytes fixed to zero, which means that the NJO byte is always a stuff byte, and the PJO1 and PJO2 bytes are always data bytes. A common demapper can recover the client mapped using either of these mechanisms.

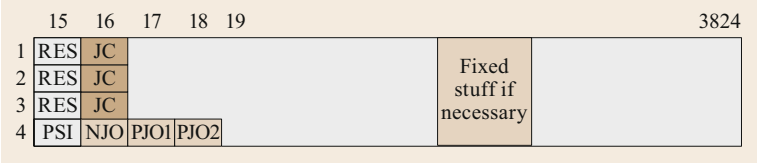


Fig. 13.11 Overhead locations for bit-synchronous and asynchronous mapping

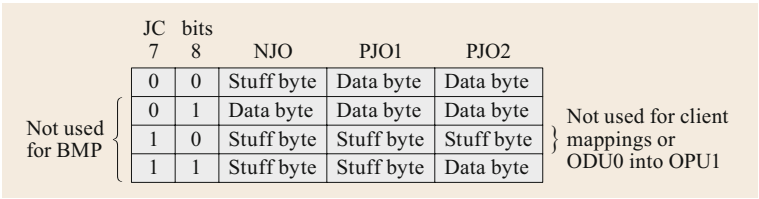


Fig. 13.12 Justification control for bit-synchronous and asynchronous mappings

Generic Mapping Procedure (GMP)

While AMP and BMP were simple in concept, they suffered a few shortcomings. Where there was a more significant difference between the size of the client and the size of the payload area, it required each mapping to calculate an appropriate number of stuff columns. When used for tributary slot mappings of clients occupying more than one tributary slot, AMP provided irregularly spaced justification opportunities across a multiframe that were difficult to manage and stay within required jitter performance. If the client had a wide clock tolerance (e.g., Ethernet signals are generally ±100 ppm, where OTN is ±20 ppm and SONET/SDH is ±4.6 ppm), more than a few stuff opportunity bytes may have been required to accommodate the variation between the client clock and the clock of the OTN frame.

As a result, GMP was designed as a single mechanism used to accommodate both the nominal bit-rate difference between the client and the payload area of the OTN frame, and the clock variations that may occur between the client and the OTN frame. There is no distinction between fixed and variable stuff byte locations.

The GMP mechanism operates by dividing the payload area (an OTN frame or a multiframe) into a certain number of GMP *words*, where each word may contain data or stuff. The words containing data are distributed as evenly as possible across the OTN frame or multiframe using a sigma/delta distribution algorithm. Correct operation depends only on the mapper and demapper knowing the number of data words that are filled in each frame or multiframe. Larger GMP word sizes are used for higher bit-rate clients and OTN frames to avoid the need for large barrel shifters in the implementation. If necessary, to meet the timing requirements of a particular client, additional timing information may also be transmitted from the mapper to the demapper, which allows the demapper to know how many client bytes (or bits) are to be emitted by the demapper during each OTN frame or multiframe period. In certain cases, the GMP word size may be as large as 80 bytes, which could produce significant mapping jitter without the mitigation provided by the additional timing information.

Table 13.3 GMP word sizes for client mappings

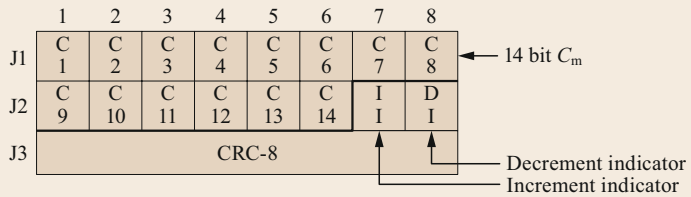
OTN container	GMP word size	Organization of OTN frame
ODU0	1 byte	15 232 GMP words of 1 byte
ODU1	2 bytes	7616 GMP words of 2 bytes
ODU2	8 bytes	1904 GMP words of 8 bytes
ODU3	32 bytes	476 GMP words of 32 bytes
ODU4	80 bytes	190 GMP words of 80 bytes plus 8 fixed-stuff columns

Table 13.3 describes the GMP word sizes for clients mapped into ODU*k* containers.

The GMP justification overhead is carried in the JC overhead in rows 1–3 of column 16 in the mapping specific overhead; the coding for the overhead is illustrated in Fig. 13.13. The information is a 14 bit binary value that indicates how many GMP words are to be filled in the next frame or multiframe. While GMP accommodates a relatively large difference between the client clock and the OTN frame size, the number of changes in words filled from one frame to the next is generally quite small. So, a coding is selected (similar to what was done for SDH pointers) to provide a coding that is robust against errors, given that an error in this information might result in demapping a client from the wrong byte positions in the frame or might result in a timing glitch for the client.

GMP is used for the mappings of lower-order ODU0 or ODUflex into tributary slots of higher-order ODU2, and for all lower-order tributaries mapped into tributary slots of ODU4. The payload area of the OTN frame (columns 17–3824) are divided into TS of approximately 1.25 G of bandwidth capacity. The assignment of byte positions to TS alternates through the payload area. Higher-order ODU4 designates the eight rightmost columns of the payload area (columns 3817–3824) as fixed stuff, as the number of bytes in the frame is not evenly divisible by 80 (the number of TS in a higher-order ODU4). GMP distribution is done over a multiframe that consists of the same number of higher-order ODU*k* frames as the number of tributary slots in the frame. Table 13.4 describes the tributary slot structure for each rate of higher-order ODU*k*.

The GMP word size is a number of bytes equal to the number of tributary slots occupied by the client. A consequence of this rule is that every lower-order tributary is filled into an area of a fixed number of GMP words across the multiframe (15 232 for higher-order ODU2 or ODU3, 15 200 for higher-order ODU4) no



C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	II	DI	Change
U	U	U	U	U	U	U	U	U	U	U	U	U	U	0	0	0
I	U	I	U	I	U	I	U	I	U	I	U	I	U	1	0	+1
U	I	U	I	U	I	U	I	U	I	U	I	U	I	0	1	–1
U	I	U	I	U	I	U	I	U	I	U	I	U	I	1	0	+2
I	U	I	U	I	U	I	U	I	U	I	U	I	U	0	1	–2
Binary value														1	1	More than +2/–2

Note: I indicates inverted C_i bit
U indicates unchanged C_i bit

Fig. 13.13 GMP justification overhead. J1–J3 contain C_m, the number of m-bit GMP words in the next frame or multiframe which contain data. The other GMP words in the frame (in excess of C_m) contain stuff. The encoding of C_m uses tricks from the encoding of SDH pointers to be resilient against bit errors, which could not only result in demapping client data from the wrong positions in the frame, but could create a client timing glitch if the demapper is trying to extract an incorrect number of GMP words of client data from a frame

matter the size of the lower-order tributary. The size of the GMP words varies with the lower-order tributary, but the number of GMP words across a multiframe does not. When a lower-order ODU*j* is mapped into multiple tributary slots of a higher-order ODU*k*, the GMP justification overhead is carried in columns 15–16 of the frame within the multiframe corresponding to the highest numbered tributary slot, and describes the number of GMP words to be filled by the lower-order tributary in the next multiframe.

In certain cases, there may be a need for fine-grain control of the timing of the mapping of a client signal. This may be challenging to do with just PLL control when the GMP word size is large (e.g., as large as 80 bytes). As a result, the GMP mechanism has the facility to, when necessary, provide additional timing information from the mapper to the demapper to allow retaining as far as possible the bit-level timing of the client.

In general, Ethernet mappings do not require this additional timing information because Ethernet jitter requirements are quite coarse (±100 ppm), but for purposes of illustration, we will describe how bit-level timing of a 100 GBASE-R Ethernet signal mapped over a 100 G line-side ODU4 could be achieved if it were required.

100 GBASE-R has a client bit rate of 103.125 Gb/s ±100 ppm. The bit rate of the payload area of the OPU4 frame (not including eight fixed-stuff columns when GMP mapping is used) is ≈ 104.137 Gb/s ± 20 ppm. The 100 GBASE-R will be mapped into an integral number of the 190 available 80 byte GMP words in each OPU4 frame. Based on the relative clock tolerances, the average number of GMP words filled in each frame could vary between ≈ 188.1314777 and ≈ 188.176634. But since an integral number of GMP words must be filled in each frame, a given frame will have either 188 or 189 filled GMP words to arrive at that average.

Some clients may require 1 byte (C_8) or 1 bit (C_1) timing granularity instead of the granularity of the GMP word size. In an actual implementation, there will be an elastic store (first in, first out FIFO) in the mapper that accumulates the *extra* client bits during the transmission of an OPU4 frame with 188 of the 640 bit GMP words filled, and the number of extra bits in that FIFO is reduced each time a frame is sent with 189 filled GMP words. Similarly, there is an elastic store

Table 13.4 Organization of higher-order ODU*k* into tributary slots

Higher-order ODU*k*	# TS	Multiframe length	Bytes/TS/ frame	Bytes/TS/ multiframe
ODU2	8	8	1904	15 232
ODU3	32	32	476	15 232
ODU4	80	80	190	15 200

Table 13.5 Example of extra timing information – 100 G client

Frame number t	Client bits $C_1(t)$ received	GMP words $C_{640}(t)$ transmitted	Bits transmitted in GMP words $640 \times C_{640}(t)$	Excess bits in ingress FIFO $\Sigma C_{1D}(t)$
1	120 419	188	120 320	99
2	120 419	188	120 320	198
3	120 419	188	120 320	297
4	120 419	188	120 320	396
5	120 419	188	120 320	495
6	120 419	188	120 320	594
7	120 419	**189**	**120 960**	53
8	120 419	188	120 320	152
9	120 419	188	120 320	251

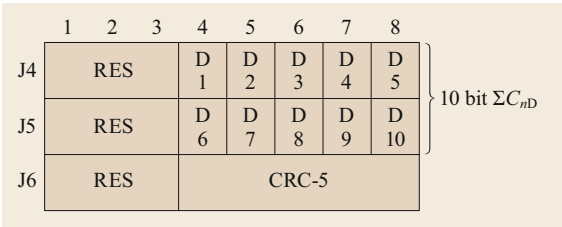


Fig. 13.14 Encoding for extra timing information (where applicable)

in the demapper, which tries to *smooth* the clock of the client from the data received across the network, and since the demapper tries to play out the client data at a constant bit rate, this egress FIFO absorbs the extra bits when an OTN frame arrives containing 189 filled GMP words, and gradually reduces its fill level while playing out the client bits at a constant bit rate when receiving an OTN frame with only 188 filled words. If both the client and the ODU4 are running at nominal frequency, the mapper will receive 120 419 bit per ODU4 frame period, and due to sending a discrete number of GMP words in each frame period, will either send 120 320 or 120 960 bit in each ODU4 frame. The extra timing information is essentially the number of excess bits held in the ingress FIFO, and is referred to as $\Sigma C_{1D}(t)$. Table 13.5 provides an example of how this timing information is used to allow the mapper to inform the demapper of the fill level in the ingress FIFO, which provides the demapper with sufficient information to play out exactly 120 419 bit of the client from its FIFO in each ODU4 frame period.

The extra timing overhead for $\Sigma C_{nD}(t)$, where $n = 1$ for bit-level timing or $n = 8$ for byte-level timing (depending on the requirements of the particular client for which the mapping is used) is shown in Fig. 13.14. These bytes are carried in rows 1–3 of column 15.

Transparent Generic Framing Procedure (GFP-T)

The more common GFP-F used for mapping of packet clients is described later, but several constant bit rate, character-oriented clients are mapped by encapsulating specific numbers of characters in a GFP frame, and if necessary, filling the space between GFP frames with GFP idles. This is particularly useful for clients such as gigabit Ethernet (1000BASE-X), where the bit rate of what is carried over the transport network can be reduced by taking advantage of specific characteristics of the client line coding.

Gigabit Ethernet uses 8B/10B line coding, so it has a physical interface signaling rate of 10/8 times the media access control (MAC) data rate, or 1.25 Gb/s. But the 10B coding is highly redundant, and not every sequence of 10 bit is a valid 10B codeword; the 10B coding includes tagging of which characters are data and which are control, and two alternate 10B encodings are available for each logical character to allow compensating for the running disparity between 0 and 1 s to provide DC balance in the overall signal. By transporting every valid character of the 8B/10B encoded stream rather than carrying every bit of an arbitrary 1.25 Gb/s signal, the transport capacity needed to carry a gigabit Ethernet signal is reduced. The mapping for gigabit Ethernet transcodes eight 10B characters from the Ethernet signal into a 65 bit block, and the demapper translates the 65 bit block back to eight 10B characters. Over SONET/SDH networks, the compression allowed gigabit Ethernet to be carried over a VC4-7v (approximately 1.052352 Gb/s of payload capacity) rather than the VC4-9v (approximately 1.353024 Gb/s of payload capacity) which would have been required without the compression. Over OTN networks, the compression allows mapping into the OTN ODU0 container, which is efficiently transported over all OTN wavelength rates. Figure 13.15 illustrates the mapping of gigabit Ethernet over ODU0.

This particular mapping does not do any Ethernet idle insertion/deletion during transcoding. The GFP frames are GMP mapped into the ODU0 container, so there is no need for the use of GFP idles or the 65B pad character. The demapper generates the equivalent stream of 10B characters at the network egress, resulting in a signal with the same average clock frequency as

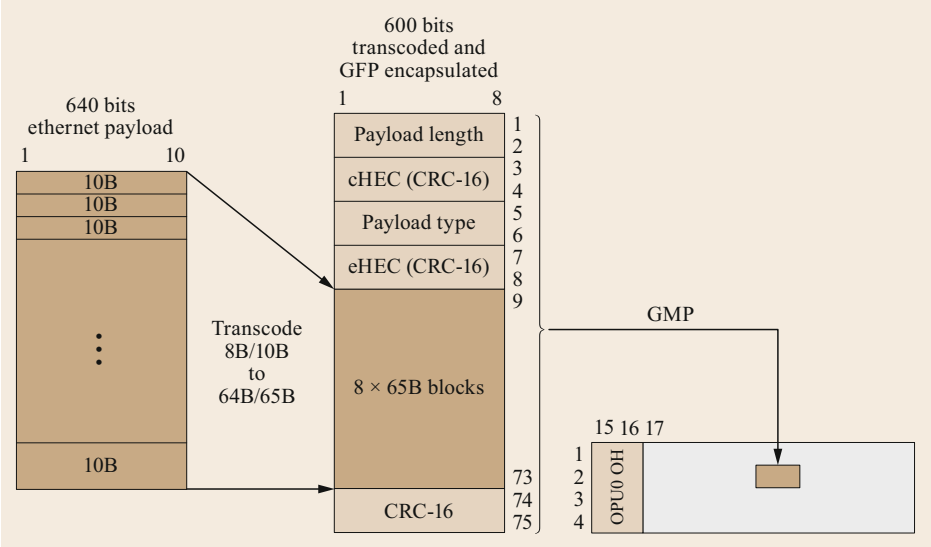


Fig. 13.15 Mapping of gigabit Ethernet over ODU0

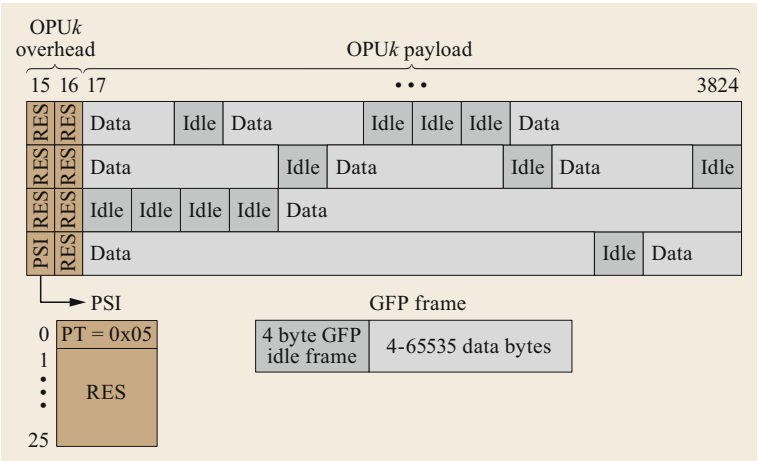


Fig. 13.16 Illustration of GFP-F mapping of packet clients over OTN

that at the network ingress, so the transport is suitable for synchronous Ethernet.

Mapping for Packet-based Clients

Frame-based Generic Framing Procedure (GFP-F). Initially, packet clients such as Ethernet packet flows, IP, and MPLS, were carried using the generic framing procedure originally developed for carrying packet clients in SDH networks. This is a simple mapping, where each packet is encapsulated in a GFP frame and GFP idles are inserted between the GFP frames as necessary to fill the payload area of the ODU*k* into which it is mapped. This mapping is illustrated in Fig. 13.16.

Idle Mapping Procedure (IMP). During the time frame of development of the Beyond 100 G OTN extension (Sect. 13.5) the possibility was considered to develop a *wide word* version of GFP-F taking into account that modern systems supporting high bit-rate interfaces generally use very wide data busses. But in a similar time frame, the trend was toward the majority of packet clients being Ethernet (particularly high bit-rate clients). Flex Ethernet was emerging (Sect. 13.5.2), and it was considered simpler to reuse the Ethernet version of 66B coding for mapping of future packet clients into the OTN. Ethernet interfaces at 40 Gb/s and higher all use the [13.5] clause 82 encoding, which aligns the start of packet to an 8 byte boundary, always implicitly encoding the packet <start> character at the beginning of a 66B block. As a result, most packet clients in B100G networks and FlexE clients are mapped, generally into ODUflex, by delineating the packets using 66B coding per [13.5] clause 82, and by adapting the packet client flow into the ODUflex container using Ethernet idle insertion/deletion.

13.5 OTN Clients

Mappings of numerous clients have been defined over OTN. As mentioned earlier, not all equipment will support every mapping; for example, early generation Telco equipment may only have supported SONET/SDH clients, while modern data center interconnect may support only Ethernet. A few important clients are described in detail in this section, followed by a summary of many of the clients that can be carried using OTN networks and technology.

13.5.1 High-Speed Ethernet

Beginning with the introduction of 10 GBASE-R Ethernet in 2003, Ethernet has been steadily replacing SONET/SDH as the most important client of optical networks.

From the introduction of 40 and 100 Gb/s Ethernet in 2010, high-speed Ethernet interfaces have generally been developed with the expectation that these interfaces would be composed of lower-speed parallel lanes, where a serial interface at those rates was not expected to be feasible or economic for some years after the introduction of an Ethernet speed. 40 Gb/s Ethernet was developed as a 4×10 Gb/s interface, and 100 Gb/s Ethernet was developed to support some initial 10×10 Gb/s variants, with an architecture to support the longer-term 4×25 Gb/s architecture. In order to support both 4-lane and 10-lane variants of 100 Gb/s Ethernet, a logical physical coding sublayer(PCS)-lane architecture was defined, with the number of PCS lanes being the least common multiple of all of the envisioned numbers of physical lanes. 100 Gb/s Ethernet is structured as twenty 5 Gb/s PCS lanes. Two PCS lanes are bit-multiplexed onto each physical 10 Gb/s lane, and five PCS lanes are bit-multiplexed onto each physical 25 Gb/s lane. Periodic lane alignment markers are inserted onto each PCS lane to allow the receiver to properly identify, deskew, and re-interleave the data from each of the PCS lanes.

When mapping a parallel Ethernet interface over OTN, it is necessary to define a canonical serialized and deskewed version of the Ethernet information to map into the OTN frame. The amount of skew that can be tolerated by an Ethernet receiver is limited to the maximum skew expected over relatively short Ethernet link distances. An Ethernet receiver would not be able to deskew a signal where the lanes have been carried independently over thousands of kilometers. Even the cascaded skew from an OTN ingress link and the OTN egress link may exceed what is expected over a point-to-point Ethernet link. As a result, the OTN mapper has been defined in a way to remove the skew between PCS lanes, so that when the PCS lanes are redistributed to physical lanes by the demapper at the OTN egress, the Ethernet receiver sees only the skew across the egress link, which is exactly the skew as might be experienced over a point-to-point Ethernet link. Another motivation for defining canonical formats for the OTN mappings of Ethernet signals is to provide independence of Ethernet interface types at the OTN ingress and egress.

Another development in high-speed Ethernet was the introduction of forward error correction (FEC). This was initially introduced for 100 Gb/s backplane Ethernet in 2014, but was found to be useful also for 25 Gb/s lanes over multimode fiber. Several 100 G MSAs took advantage of the FEC coding gain for single-mode fiber (e.g., CWDM4, and *ER4-lite*), as the FEC coding was becoming more universally available in 100 G MAC/PCS implementations. While this was an option for certain 100 Gb/s Ethernet interfaces, FEC became mandatory as IEEE began to develop 200 and 400 Gb/s Ethernet using 50 and 100 Gb/s PAM4 encoded lanes. IEEE has used Reed–Solomon codes based on a 10 bit symbol size, which provides the necessary coding gain with low-latency and low-power implementations. An RS(528, 514) FEC code is used for most 100 Gb/s Ethernet implementations, where the FEC parity occupies precisely the amount of space saved by transcoding 64B/66B codewords to a 256B/257B format. An RS(544, 514) FEC code is used for interfaces with 50 or 100 Gb/s PAM4 encoded lanes. Together with the 257B transcoding, this amounts to about 3.03% overclocking of the lanes as compared to what would be required for 66B coding without FEC.

When mapping an Ethernet interface with FEC over OTN, another consideration is that the Ethernet FEC has been chosen to correct the number of errors expected over a single point-to-point Ethernet link between an Ethernet transmitter and an Ethernet receiver. It was not designed to correct double-link errors that would accrue across the cascade of the ingress link, the OTN network itself, and the egress link. As a result, the OTN mapper will correct errors and remove the FEC code, and transcode the signal back to a 66B format for OTN transport. Across the OTN, the links are protected by an OTN FEC. At the demapper, the signal is again transcoded to 257B format and a new Ethernet FEC is calculated, which at this point only needs to be able to correct the errors that occur in the egress link.

Another trend in Ethernet standardization is to leverage the physical lane rates of one rate of Ethernet to create another. Ethernet evolution is coming to be viewed both in terms of the top-line MAC data rate as well as

the physical lane rate(s) that are used, which are reflective of the underlying technology. For example, 25 Gb/s lane rates were developed as part of 100 Gb/s standardization, but there followed an interest to use 25 Gb/s per lane technology in single-lane variants for 25 Gb/s Ethernet. The initial interest was for short-reach data center applications (e.g., server to top-of-rack), but this also became appealing as an upgrade path for single-lane 10 GBASE-R interfaces to move to 25 Gb/s signaling over the same cabling without having to move to a multilane interface. Similarly, when 50 Gb/s per lane PAM4 signaling was developed as part of 200 and 400 Gb/s Ethernet standardization, there was interest to use the 50 Gb/s signaling in single lane variants as 50 Gb/s Ethernet. So, 25 Gb/s Ethernet was developed and standardized after 100 Gb/s Ethernet, leveraging some of the same technology, and 50 Gb/s Ethernet was developed and standardized after 200 and 400 Gb/s Ethernet, leveraging some of the same technology.

Until recently, Ethernet evolution followed a path of supporting steadily increasing MAC data rates in powers of 10: 10 Mb/s Ethernet emerged in 1980, followed by 100 Mb/s Ethernet in 1995, 1 Gb/s Ethernet in 1998, and 10 Gb/s Ethernet in 2003. The *power of 10* paradigm was broken in 2010 when IEEE 802.3 standardized both 40 and 100 Gb/s Ethernet. While the *headline* 802.3 projects continue to increase MAC data rates with 200 and 400 Gb/s Ethernet emerging in late 2017, the attractiveness of leveraging physical lane rate technologies has spawned another set of MAC rates, with 25 Gb/s Ethernet emerging in 2016 and 50 Gb/s Ethernet emerging in 2018. Other specialized applications have spawned even more data rates, with 2.5 and 5 Gb/s Ethernet being introduced to connect 802.1ac wireless access points over legacy cabling or for storage applications. So, as many new Ethernet MAC rates have been introduced after 2010 as were introduced in the previous 30 years.

13.5.2 Flex Ethernet

After a period of speed convergence between Ethernet and transport networks, with (approximately) aligned rates of 10, 40, and 100 Gb/s, the flexibility of coherent modulation formats and encountering the Shannon limit so that wavelength rates in transport networks start to vary with reach rather than a regular, monotonic increase creates a situation where more technical approaches are necessary to efficiently use network capacity. If the transport network uses 100 Gb/s wavelengths, you could interconnect data centers with 100 Gb/s Ethernet using transponders. But as the coherent modulation and reach constraints create more *interesting* wavelength rates like 150 or 250 Gb/s, it becomes a challenge to efficiently fill these wavelengths with Ethernet traffic.

One approach would be to aggregate these wavelengths in the transport network to the next logical multiple of 100 Gb/s, as discussed earlier. For example, an OTUC3 (300 Gb/s) could be inversely multiplexed over an OTSiG composed of two 150 Gb/s wavelengths, or an OTUC5 (500 Gb/s) could be inversely multiplexed over an OTSiG composed of two 250 Gb/s wavelengths. But this was not felt to be general-purpose enough to cover all use cases.

As a result, flex Ethernet (FlexE) [13.6] was developed, initially to address some of the speed differences between Ethernet rates and transport network rates. This technology allows for a decoupling of Ethernet MAC rates from Ethernet PHY (physical layer device) rates. The first version of the flex Ethernet implementation agreement supports groups composed of from 1 to n 100 GBASE-R Ethernet PHYs (with $n \times 100$ Gb/s of capacity). There is work underway to be able to create FlexE groups of 200 GBASE-R or 400 GBASE-R Ethernet PHYs. FlexE MAC client(s) can be created at rates of 10, 40, or multiples of 25 Gb/s. The capacity of the FlexE group is divided into twenty 5 G calendar slots per PHY according to a round-robin allocation of 66B block positions, and FlexE clients are inserted into the 66B block positions to which they are assigned.

For example, the situation of wanting to efficiently use a 150 Gb/s wavelength in a transport network with Ethernet traffic could be addressed by creating a FlexE group of two 100 GBASE-R Ethernet PHYs, and assigning a 150 Gb/s FlexE MAC client to 30 of the available 40 calendar slots across the two PHYs.

Besides the ability to create *interesting* virtual Ethernet rates to match flexible coherent line interface formats, the flex Ethernet mechanism can be used for simple non-LAG (link aggregation group) bonding [13.7], and for channelization to support multiple smaller Ethernet virtual PHYs across a group of physical PHYs. Together with transport networking, channelization can be used in a way that not every FlexE client carried over a FlexE group between a router and transport equipment needs to be connected to the same place in the transport network.

The FlexE group structure is illustrated in Fig. 13.17.

The FlexE structure creates a logical stream of 66B blocks on each PHY that conforms to Ethernet rules—each 66B block is a legal 66B block, even though the sequence of blocks on a PHY does not constitute a legal 66B block stream. Since each 66B block is a legal block, all of the lower layers of the IEEE 802.3 stack can be used, which might include 257B transcoding of those blocks and addition of FEC.

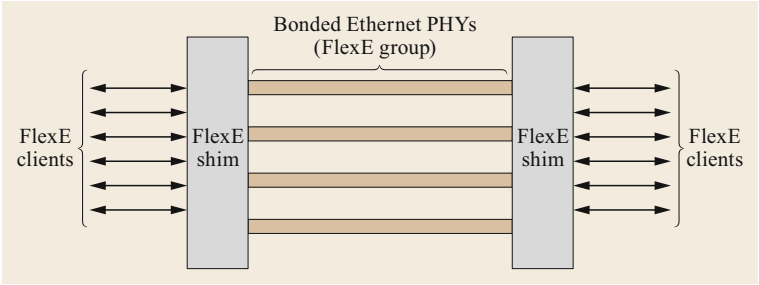


Fig. 13.17 FlexE group structure

The sequence of 66B blocks on each PHY is illustrated in Fig. 13.18.

The FlexE overhead is used to convey information from the FlexE mux to the FlexE demux concerning the group configuration and how clients are mapped into calendar slots, so that it can be assured that the FlexE demux is properly demapping FlexE clients from the slots into which they are mapped. The position of the FlexE overhead blocks is also used as a kind of a marker to allow deskewing of the PHYs of a FlexE group, so that clients that span more than one PHY can be properly reconstructed. Figure 13.19 illustrates the structure of a master calendar that runs across all PHYs of the group. The 66B block positions of the master calendar are distributed round-robin, twenty 66B blocks at a time (one calendar cycle) to each of the FlexE PHYs, with periodic insertion of the FlexE overhead in the same position on each of the PHYs.

Figure 13.20 illustrates how the calendars of each PHY of the FlexE group are re-interleaved by the FlexE demux, using the position of the FlexE overhead block as a marker to remove skew between the PHYs to keep the block order correct for FlexE clients that span more than one PHY.

Besides simple non-LAG bonding, FlexE can be used in several *router to transport* use cases.

The simplest use case is referred to as *FlexE Unaware*. In this use case, an end device such as a router creates the FlexE group. The PHYs of the FlexE group are carried over the transport network as if they were 100 Gb/s Ethernet clients using a PCS codeword transparent mapping. The transport network is unaware that the PHYs belong to a FlexE group. However, to minimize skew, all of the PHYs of the FlexE group must be carried over the same fiber route. Even with that routing restriction, the FlexE implementation must be able to tolerate skew between the PHYs of the group that are significantly greater than Ethernet inter-lane skew, since the propagation of each wavelength is affected by chromatic dispersion. For the cases where the wavelengths are significantly separated on the fiber (e.g., one wavelength is at the short end of the S band, and another wavelength is at the long end of the L band), and the link distances are long (trans-Pacific links could be as long as 12 000 km), a skew budget of 10 μs is recommended to address these kinds of applications.

Another use case is referred to as FlexE terminating, where the extent of the FlexE group is created by the router and terminated by the transport network equipment which extracts and carries the clients. A FlexE terminating configuration could be used to efficiently fill flexible coherent line interfaces; for example, if the wavelength is 150 G, a FlexE group composed of two 100 GBASE-R PHYs could carry a single 150 G FlexE client in 30 of the available 40 calendar slots. The mapping of this client over the transport network maps the stream of 66B blocks of the FlexE client into an ODUflex container. Idle insertion and deletion per usual Ethernet processing is used to convert from the FlexE clock domain to the ODUflex clock domain. In addition to efficient filling of coherent wavelengths that do not correspond to Ethernet PHY rates, a FlexE terminating configuration may be used together with channelization where not all of the FlexE clients in the group are intended to be connected to the same place in the network. An illustration of this channelization flexibility is shown in Fig. 13.21.

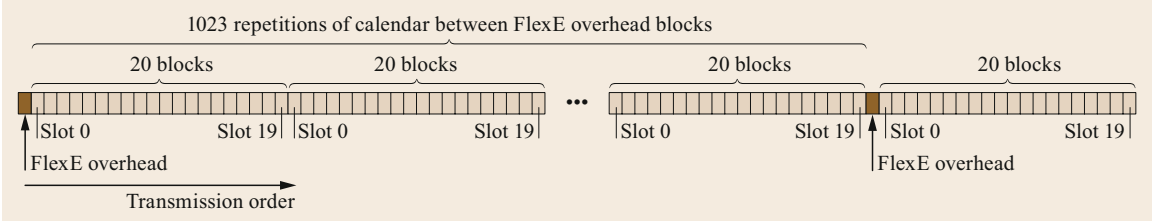


Fig. 13.18 Sequence of 66B blocks carried on each PHY of a FlexE group

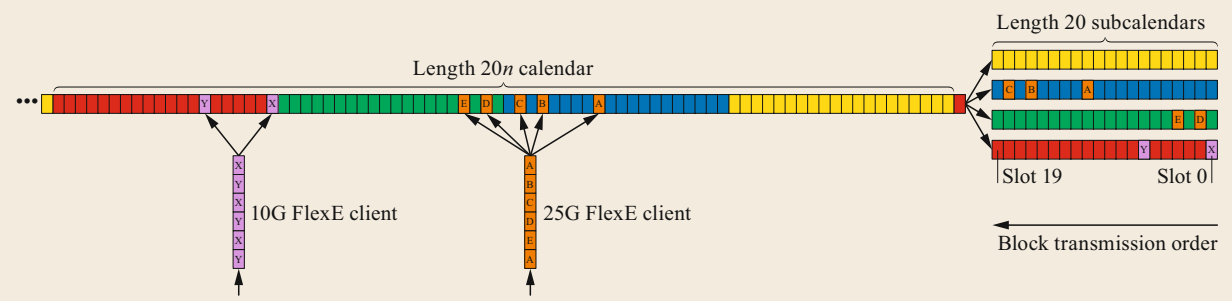


Fig. 13.19 Distribution of a master calendar to the PHYs of a FlexE group

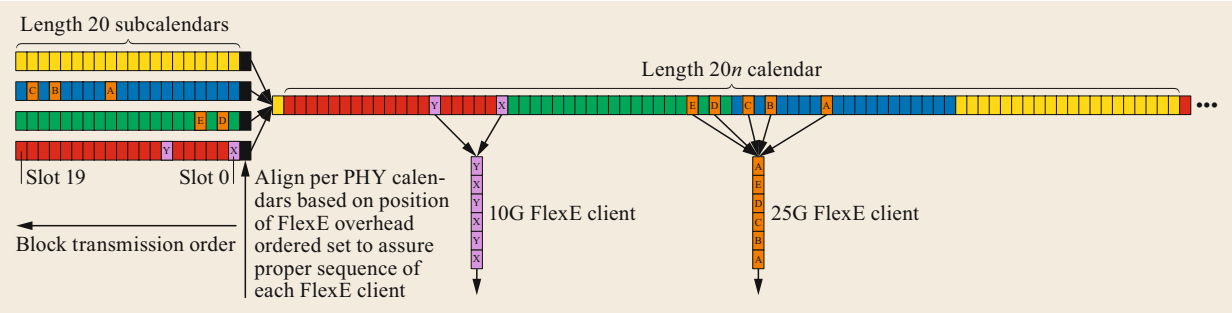


Fig. 13.20 FlexE calendar re-interleaving and extraction of FlexE clients

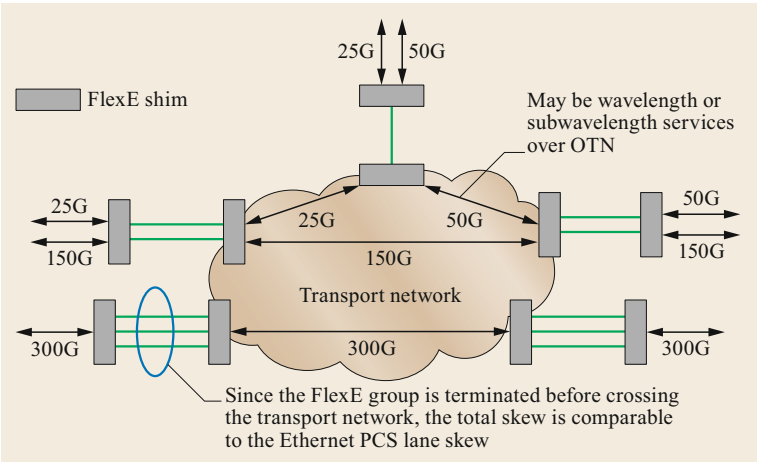


Fig. 13.21 FlexE terminating use case with channelization

A final router to transport use case is referred to as FlexE aware. An example of where this configuration might be useful is wanting to carry 300 G over two 150 G wavelengths. In the case of modern OTN equipment with the right granularity implementing the *beyond 100 G* extensions (as was discussed earlier), this might be addressed by creating an OTUC3 OTN container that is carried over an OTSiG consisting of the two 150 G wavelengths, and to carry the 300 G ODUflex over the OTUC3. But if the two 150 G wavelengths are not on the same transponder line card, and there is no hardware in the right place to be able to terminate the entire FlexE group, this can be addressed by creating a FlexE group consisting of four 100 GBASE-R PHYs. The four PHYs are organized into two 2-PHY subgroups, and each subgroup is configured marking some (10, in this example) of the calendar slots as *unavailable*. The transport network uses a special mapper to interleave the PHYs of the FlexE subgroup while discarding the unavailable calendar slot positions. The demapper will restore the full rate for each of the PHYs by inserting Ethernet error control blocks into the (not

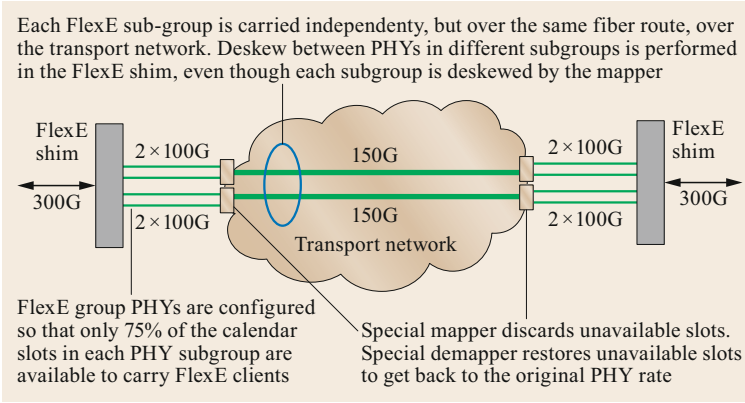


Fig. 13.22 FlexE aware configuration

carried) unavailable calendar slot positions. This configuration is illustrated in Fig. 13.22.

13.5.3 OTN Client Mapping Summary

Tables 13.6–13.10 list major clients that can be supported by OTN.

13.5.4 Mobile Network Applications of OTN

Since OTN has served as a general Telco network interface, it has naturally found widespread application for mobile backhaul. But with evolution of the architecture of the radio access network (RAN), OTN technologies have been used ever closer to the network edge. Many 4 G mobile networks use an interface called CPRI (common public radio interface) between the baseband unit (BBU) and the remote radio unit (RRU). CPRI options 1 through 10 are defined with signaling rates ranging from 614.4 Mb/s through 24.33024 Gb/s. Initially, these were very short-reach interfaces, for example from the bottom to the top of a cell tower. But a trend has emerged where some operators would like to centralize the BBU functionality serving a number of towers and extend the CPRI interface over a WDM network.

Table 13.6 Telco network client mappings into OTN

Client	Bit rate	Mapping
STM-256/OC768	39.81312 Gb/s ± 20 ppm	AMP or BMP into ODU3
STM-64/OC192	9.95328 Gb/s ± 20 ppm	AMP or BMP into ODU2
STM-16/OC48	2.48832 Gb/s ± 20 ppm	AMP or BMP into ODU1
STM-4/OC12	622.08 Mb/s ± 20 ppm	GMP[a] into ODU0
STM-1/OC3	155.52 Mb/s ± 20 ppm	GMP[a] into ODU0
CM-GPON	2.48832 Gb/s ± 20 ppm	AMP into ODU1
CM-XGPON	9.95328 Gb/s ± 20 ppm	AMP into ODU2

[a] GMP additional timing information with 1 bit granularity needed to meet SDH jitter requirements

Table 13.7 Ethernet mappings into OTN

Client	Bit rate	Mapping
1000BASE-X (Gb Ethernet)	1.25 Gb/s ± 100 ppm	65B transcoded, GFP-T/GMP into ODU0
10 GBASE-R	10.3125 Gb/s ± 100 ppm	Numerous: most common are BMP into ODU2e or GFP-T into ODU2
25 GBASE-R	25.78125 Gb/s ± 100 ppm	FEC corrected, trans-decode 257B to 66B, BMP into ODUflex
40 GBASE-R	41.25 Gb/s ± 100 ppm	1024B/1027B transcoded, GMP mapped into ODU3
50 GBASE-R	53.125 Gb/s ± 100 ppm	FEC corrected, trans-decode 257B to 66B, BMP into ODUflex
100 GBASE-R	103.125 Gb/s ± 100 ppm	FEC corrected with trans-decode 257B to 66B if necessary, GMP into ODU4
200 GBASE-R	212.5 Gb/s ± 100 ppm	FEC corrected, trans-decode 257B to 66B, remove AMs, add rate compensation blocks, BMP into ODUflex
400 GBASE-R	425 Gb/s ± 100 ppm	FEC corrected, trans-decode 257B to 66B, remove AMs, add rate compensation blocks, BMP into ODUflex

Table 13.8 Storage/data center interfaces

Client	Bit rate	Mapping
SBCON/ESCON	200 Mb/s ± 100 ppm	GMP into ODU0
FC-100	1.06525 Gb/s ± 100 ppm	GMP into ODU0
FC-200	2.125 Gb/s ± 100 ppm	GMP into ODU1
FC-400	4.25 Gb/s ± 100 ppm	BMP into ODUflex
FC-800	8.5 Gb/s ± 100 ppm	BMP into ODUflex
FC-1200	10.51875 Gb/s ± 100 ppm	1024B/1027B transcoded, BMP mapped into ODU2e
FC-1600	14.025 Gb/s ± 100 ppm	BMP into ODUflex
FC-3200	28.025 Gb/s ± 100 ppm (after trans-decoding and removal of FEC)	BMP into ODUflex

Table 13.9 High-performance computing interfaces mapped over OTN

Client	Bit rate	Mapping
Infiniband SDR	2.5G ± 10 ppm	BMP into ODUflex
Infiniband DDR	5G ± 10 ppm	BMP into ODUflex
Infiniband QDR	10G ± 10 ppm	BMP into ODUflex

Table 13.10 Video distribution interfaces mapped over OTN

Client	Bit rate	Mapping
DVB_ASI	270 Mb/s ± 100 ppm	GMP into ODU0
SDI	270 Mb/s ± 100 ppm	GMP into ODU0
1.5 G SDI	1.485 Gb/s ± 10 ppm 1.485/1.001 Gb/s ± 10 ppm[a]	GMP into ODU1
3 G SDI	2.97 Gb/s ± 10 ppm 2.97/1.001 Gb/s ± 10 ppm[a]	BMP into ODUflex
6 G SDI	6 Gb/s ± 10 ppm	BMP into ODUflex
12 G SDI	12 Gb/s ± 10 ppm	BMP into ODUflex
24 G SDI	Under development	

[a] NTSC variants supporting different frame rates

Since CPRI was designed as a very short-reach interface, the performance requirements are very strict. Jitter is limited to 2 ppb. The latency limits are very tight, so propagation delay limits the maximum reach to 10–12 km. In addition to overall latency limits, the asymmetry in the latency is strictly limited. There are several CPRI reach-extension products in the market, and while some use OTN frame formats with OTN mapping methods, including BMP, GFP-T, and GMP, the methods used to meet the CPRI performance requirements are all currently single vendor. While CPRI extension links may be carried over relatively short distances over metro reconfigurable optical add-drop multiplexer (ROADM) networks, they are all single-vendor bookended links and single optical spans from a transmitter to a receiver without electrical regeneration. Some of the mapping methods that use OTN frame formats and mapping techniques are described in an informative document [13.8].

Architectures and technologies for 5 G radio access networks are still a matter of discussion and debate. The functional split points are likely to be different, with a fronthaul segment between a distributed unit (DU) and the remote radio unit (RRU), a *middle-haul* segment between a centralized unit (CU) and the DU, and backhaul behind the DU. Different latency requirements apply to the fronthaul, middle-haul, and backhaul. A new eCPRI specification has been developed for use over Ethernet networks, with less challenging performance requirements than the CPRI interfaces used in 4 G networks, and new time-sensitive networking (TSN) techniques are being developed as part of the IEEE 802.1CM project to address 5 G fronthaul requirements. Other technologies including PON (passive optical networking) are under consideration by some operators for 5 G fronthaul applications.

13.6 OTN Beyond 100 G

As described earlier, up through 100 G, the practice was to evolve OTN through the addition of larger hierarchal layer containers, e.g., when ODU4 was added to reflect a new 100 G container which was capable of carrying 100 GBASE-R Ethernet as well as a multiplex of other, smaller traffic (e.g., 10 G, 40 G). Each time a hierarchal level was added, the signaling rate per wavelength was increased, generally with similar reach to the lower-rate previous generation technology. Coherent DP-QPSK with 100 G per wavelength was able to achieve about 4000 km reach in a 50 GHz dense wavelength division multiplexing (DWDM) grid, so there was an immediate increase of 2.5× in spectral efficiency. Initial thinking was that the next step beyond 100 G would be a new, higher-rate container, likely called ODU5, with a signaling rate expected to be around 400 G per wavelength.

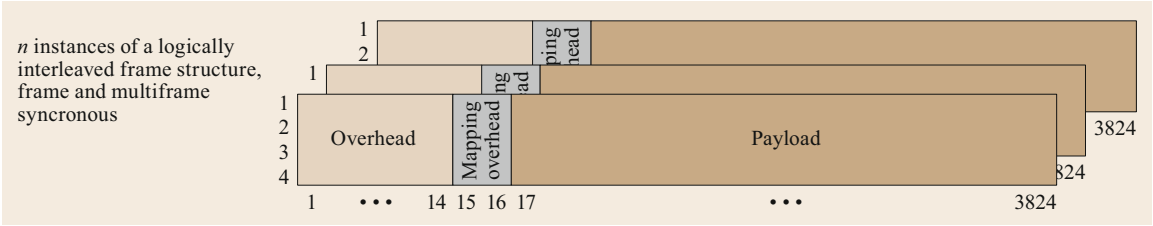


Fig. 13.23 OTUC*n* structure

Unfortunately, the Shannon limit interfered with that idea, as even going to 200 G per wavelength would reduce maximum reach from 4000 km to around 1000 km, and 400 G per wavelength would reduce reach further, perhaps to no more than about 200 km without electrical regeneration. This led to the current paradigm of trading off modulation complexity, baud rate, and the amount of spectrum allocated in a flex-grid network against the desired reach in network design. The flexibility provided by digital signal processor (DSP)-driven coherent modulators and demodulators, together with interesting techniques like probabilistic constellation shaping to get ever closer to the Shannon limit has resulted in an environment where there is no single *next wavelength rate* beyond 100 G. However, the services that need to be supported keep growing in size (e.g., with IEEE 802.3 defining 200 and 400 Gb/s Ethernet) creating the need for increasing digital container sizes to carry these network clients, while dealing with a physical layer infrastructure that has moved away from a paradigm where wavelengths operate at a small number of discrete signaling rates.

A key step in resolving this conflict was to decouple the digital container from the wavelength(s) that might carry it. A 400 G digital container to transport 400 GbE might be transported via a single wavelength over very short distances, but may require two wavelengths or subcarriers to go 1000 km, or four wavelengths or subcarriers to go 4000 km. But the digital container is the same in all cases. The group of wavelengths that carries a single digital container is referred to as an optical tributary signal group (OTSiG), and the digital container is designed without worrying about how many members are in that group. The addition of FEC (hard or soft decision) is considered to be part of the adaptation of the digital container to the OTSiG, and the FEC parity is no longer considered to be extra columns locked to the OTN frame.

As higher-rate services to be carried are generally multiple of 100 G services, the formulation of the container is something called an OTUC*n*, where C = 100 and *n* indicates the number of 100 G instances that form the container (e.g., OTUC3 = 300 G). This is a general-purpose container for interfaces with a digital frame format with a size greater than 100 G, although depending on desired reach, the wavelength rate may not exceed 100 G. This is based on *n* interleaved instances of a 100 G frame format, as shown in Fig. 13.23.

The characteristics of the OTUC*n*/ODUC*n* structure are as follows:

- It is formed of *n* logically interleaved instances of a 100 G frame format. The 100 G frame format is similar to the OTU4 format without FEC, and slightly larger than OTU4, as it is expected to be able to carry at least *n* lower-order ODU4 (e.g., carrying 100 GBASE-R Ethernet) inside of an ODUC*n*.
- This is a higher-order (section) container only. No clients are mapped directly into the ODUC*n*. Clients are always mapped first into a lower-order container, which is then carried in the tributary slots of the ODUC*n*. For example, 400 GBASE-R Ethernet will be mapped into a 400 G ODUflex, which then might be mapped into all of the tributary slots of an ODUC4.
- To reduce the complexity of these framer devices, a coarser tributary slot granularity has been adopted than for ODU*k* where $k \leq 4$. The tributary slot granularity is 5 G (with 20 TS supported per 100 G slice), but the number of tributaries is limited to 10 per 100 G slice. The purpose of this granularity is to allow for efficient transport of both 10 G tributaries and 25 G tributaries, not that there is any important client expected at 5 G.
- For efficient transport of much smaller clients (e.g., 1 GbE, STM-1, STM-4), an intermediate layer of multiplexing will be used (e.g., ODU2 or ODU4).

A consequence of the interleaved structure is that growing to larger container sizes also increases the number of overhead bytes per frame. The allocation of these overhead bytes is nearly identical to that for OTU*k*/ODU*k* for $k \leq 4$, with the following modifications:

- The frame alignment overhead appears in every 100 G slice.

- Certain overhead bytes are only carried in the first 100 G slice of the ODUC*n* container: in particular, the management channels, the synchronization messaging channel, the protection switching bytes, and delay measurement.
- Within the path, section, and TCM monitoring bytes, the trail trace identifier (TTI) is only carried in the first 100 G slice. The BIP is calculated on a per-100 G slice basis, with the backward error counters also on a per-100 G slice.
- Note that tandem connection monitoring for ODUC*n* has less applicability than for ODU*k*. This will support regenerators and *carrier's carrier* configurations. ODUC*n* is not considered to be a switchable entity.

13.7 Optical Media Layer Management

The sections above describe the digital aspects of what is carried on an individual wavelength (or wavelength group, in the case of OTN beyond 100 G). But an important part of managing an optical network is managing and configuring the optical media layer network over which those wavelengths are carried. The optical media itself has no inherent structure; unlike digital multiplexing, wavelengths on a fiber are unaware of the existence of, or their relationships to, other wavelengths on the same fiber. But a hierarchy of maintenance entities is created that allows the optical media to be managed using a similar paradigm used for managing the digital layers. Figure 13.24 gives an overview of the optical maintenance entities.

Just as with the digital layers, an optical wavelength (or wavelength group, in the case of beyond 100 G) may be carried across a network composed of a series of network elements and fibers between the optical modulator (transmitter) and optical demodulator (receiver). The network elements can consist of optical multiplexers and demultiplexers, network elements such as amplifiers or equalizers that have no wavelength routing flexibility, or network elements like ROADMs or WSSs that have wavelength routing flexibility. The legacy term for a wavelength that carries an OTU*k* for $k \leq 4$ is OCh. This once stood for *optical channel*, but it was later evident this was an unfortunate term as what it referred to was actually a signal rather than a channel, so the expansion of the acronym was later dropped. For beyond 100 G signals, the group of (one or more) wavelengths that carries an OTUC*n* is referred to as an optical tributary signal group (OTSiG).

An optical transmission section (OTS) is a maintenance entity that is used across something like a single fiber span between adjacent points of visibility in the network. There may be unmanaged elements internal to the OTS (e.g., a length of dispersion compensating fiber), but all signals inserted at one end of the OTS are expected to emerge from the other end of the OTS.

An optical multiplex section (OMS) is a maintenance entity that is used across a series of network elements that have no routing flexibility for payload wavelengths, so all payload wavelengths inserted at one end are expected to emerge from the other end. There may be intermediate network elements that are visible and manageable, such as amplifiers.

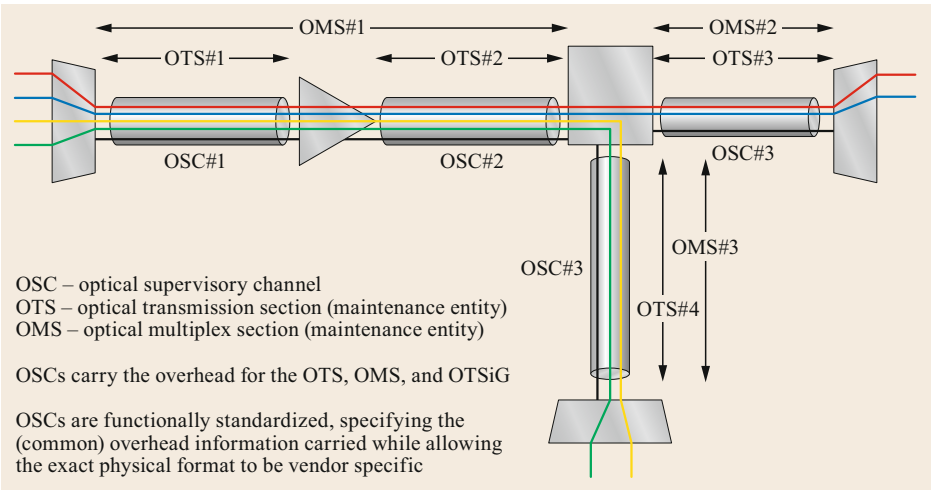


Fig. 13.24 Overview of optical media layer maintenance entities

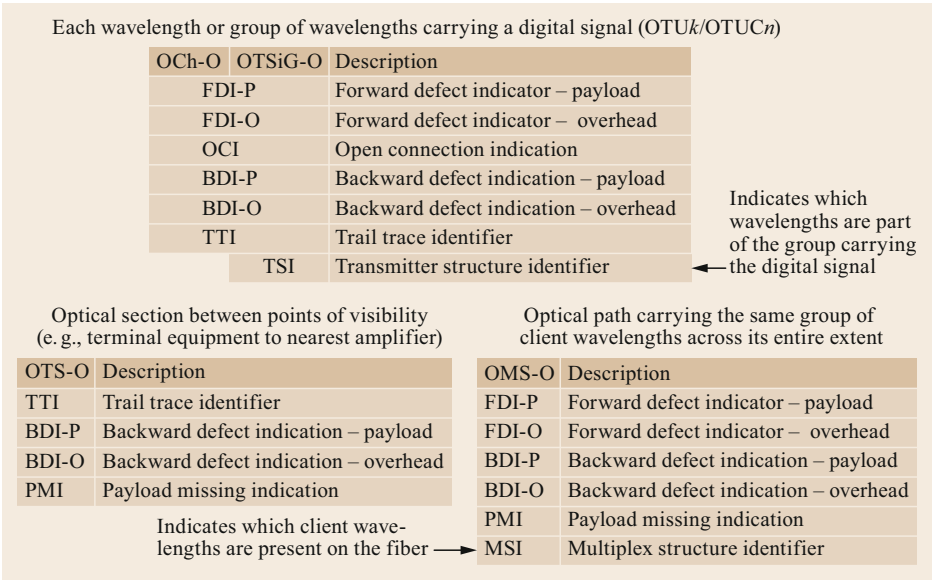


Fig. 13.25 Optical media layer overhead carried in the OSC

An optical supervisory channel (OSC) is present across each OTS in the optical media network to convey maintenance overhead related to the signals and OTS/OMS maintenance entities that traverse the particular OTS. This is generally carried on its own wavelength in a vendor-specific format. The information content of the OSC is functionally standardized, allowing a common management information view of a network composed of multiple single-vendor subnetworks.

Each of the client signals and maintenance entities can be considered to have a payload part (carried by the payload wavelengths) and an overhead part, which is nonassociated overhead carried in the OSC and relayed through each network element to the terminating point of the relevant maintenance entity. For example, in Fig. 13.24, the OCh overhead for the red wavelength is generated at the multiplexer on the left, is relayed through the amplifier and the ROADM, and is terminated at the multiplexer on the right. The overhead for OMS#1 that is composed of the red, blue, green, and yellow wavelengths between the multiplexer on the left and the ROADM is generated at the multiplexer on the left, relayed through the amplifier, and is terminated at the ROADM.

The physical layer defects detectable in the optical media layer may include loss of signal (LOS) on the OSC, LOS, on an OMS or OTS maintenance entity basis based on a measurement of received optical power, or LOS on an OCh or OTSiG payload based on received optical power on a per-wavelength basis. The overhead carried in the OSC for each of the maintenance entities described is illustrated in Fig. 13.25. The semantics is similar to that for digital layer overhead, intended to ensure proper interconnection of the optical fibers in the network via the trail trace identifiers, and a way for the near end to determine that the far end is receiving the transmitted signal correctly.

13.8 OTN Client Interfaces

In some cases, the client interfaces for the OTN network (e.g., Ethernet interfaces) appear directly on the OTN network elements themselves, and the line side interfaces are colored, DWDM interfaces. However, in some cases, it is useful to be able to have a single OTN signal on a fiber, so that the OTN signal monitoring, management, and perhaps features such as network protection, can be extended, perhaps across a user-network interface (UNI) all the way to the customer edge. These are sometimes called *grey optics* interfaces as a casual name for the case where there is a single OTU*k* carried over a fiber that cannot be optically multiplexed with other OTN signals. There must be an O-E-O conversion from the *grey* client signal to a colored line signal before optical multiplexing could be done.

13.8.1 Serial OTN Client Interfaces

At 10 Gb/s, the fact that SONET/SDH OC-192/STM-64, OTN OTU2, and Ethernet 10 GBASE-R all had

similar rates of operation allowed OTN client interfaces to use the same optical components as those developed for Ethernet, creating a situation where for many years, the least expensive OTN interfaces were operating at 10 G.

40 Gb/s emerged much earlier in transport networks than in Ethernet (SONET/SDH around 1998, OTN in 2002, and Ethernet not until 2010). The first multivendor 40 Gb/s client interface was [13.9] VSR2000-3R2, which was a serial, C-band interface operating at up to 2 km reach, which remained stubbornly expensive, partly due to not sharing any component manufacturing volumes with Ethernet.

13.8.2 Parallel OTN Client Interfaces

Inspired by the cost-disparity between 10 G and 40 G OTN client interfaces, when the IEEE P802.3bs project developed 40 and 100 Gb/s Ethernet based on parallel interfaces, the industry undertook to design an OTN frame format that could reuse Ethernet 40 G and 100 G parallel interface modules for OTN client interfaces. What was required was to *stripe* the OTN frame into a number of optical transport lanes (OTLs) equal to the number of PCS lanes used for Ethernet operating at the same approximate rate (the OTN rate of operation for these modules is slightly higher). OTU3 (40 G) signals are striped into four OTLs, and OTU4 (100 G) signals are striped into 20 OTLs, which can then be bit-multiplexed as necessary onto the appropriate number of physical lanes. For example, 5 of the 20 OTLs of an OTU4 are bit-multiplexed onto each physical lane of a 4×25 G pluggable module designed for 100 GBASE-R. The nomenclature for a multiple lane interface for an OTU*k* is OTL*k.x*, where *x* represents the number of physical lanes over which the OTU*k* is striped. For example, OTL4.4 represents an interface of four physical lanes (using a 4×25 G Ethernet module), where each physical lane carries five bit-multiplexed OTLs.

The way the OTU*k* frame is striped across multiple lanes is by distributing the OTU*k* frame to the lanes round robin in 16 byte increments. At each OTU*k* frame boundary, the lanes are rotated so that the bytes that were sent to lane 1 on the previous frame are sent to lane 2, the bytes that were sent to lane 2 on the previous frame are sent to lane 3, and so forth. This produces the pattern shown in Fig. 13.26 for the OTU3 and OTU4 frames.

A consequence of this striping method is that bytes 1–16 of each OTU*k* frame will appear on each OTL once per 4×3824 bytes (exactly the size of an OTU*k* frame), and the receiver can use a frame alignment method that is nearly identical to that used for recovering any OTU*k* frame from a line interface. The relative position of the frame alignment signal on each lane allows the receiver to reconstruct the original OTU*k* frame.

While the OTN bit rate of operation for 40 GBASE-R and 100 GBASE-R modules is slightly higher than the Ethernet bit rate of operation, the performance is maintained from the fact that the initially standardized 40 GBASE-R and 100 GBASE-R interfaces operated without FEC, whereas the OTN frames used RS(255, 239) error correction.

ITU-T optical specifications corresponding to the use of Ethernet pluggable optical modules for OTL3.4 (40 G OTU3) and OTL4.4 (100 G OTU4) can be found in [13.10] and [13.9].

13.8.3 Flexible OTN Client Interfaces (FlexO)

As described in Sect. 13.5, OTN evolution beyond 100 G did not continue the track of defining a new hierarchal level with a new discrete line interface rate every time technology enabled a higher speed. The

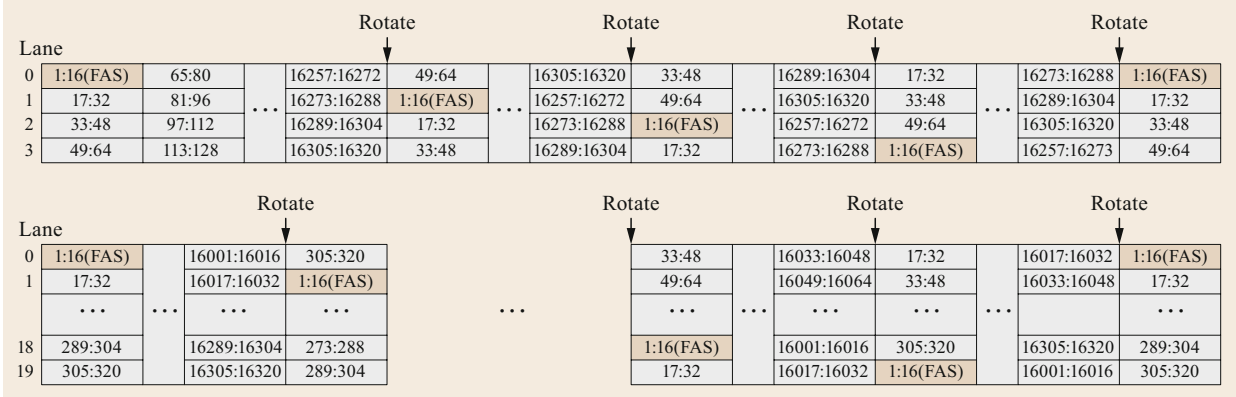


Fig. 13.26 Distribution of OTU3 and OTU4 frames to parallel lanes

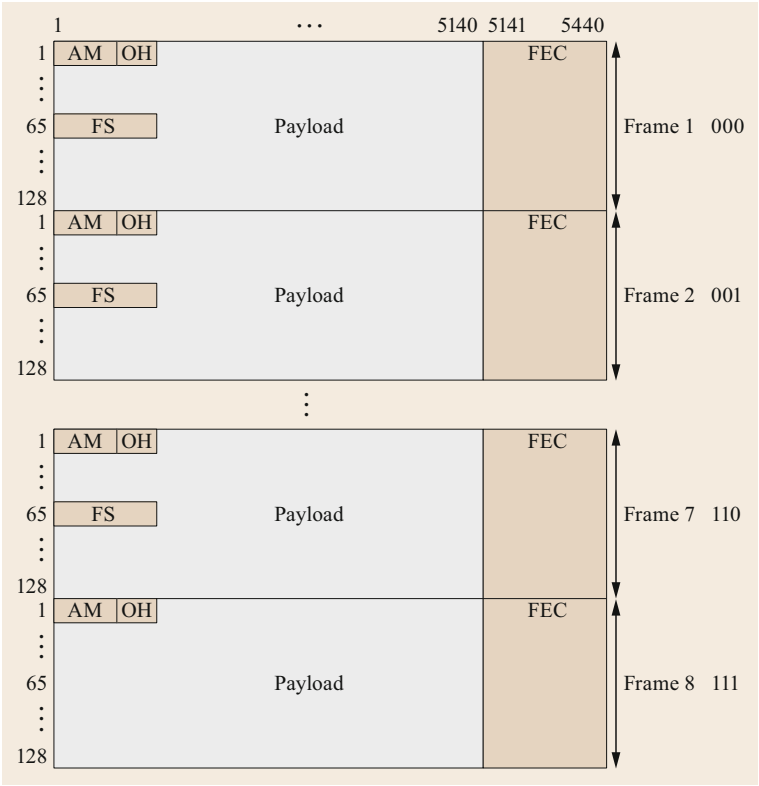


Fig. 13.27 FlexO frame and multiframe format

OTUC*n* formulation permitted creation of a variety of line interface rates to meet the service rate and reach requirements of a particular application. IEEE 802.3 Ethernet has continued the trend of specifying new, discrete interface rates, with the next two beyond 100 G being 200 G and 400 G. The issue became how to create OTN client interfaces at greater than 100 G rates, taking advantage of Ethernet pluggable optics, where the line interfaces exist at an increasing variety of rates. The answer was to define an OTN frame format for an OTUC*n* that could be distributed across *n* 100 G Ethernet interfaces, much in the style of flex Ethernet (Sect. 13.5.2).

For FlexO over 100 GBASE-R Ethernet modules, the *n* 100 G slices of the OTUC*n* are dis-interleaved, and each 100 G slice is encapsulated in a FlexO frame. The FlexO frame is designed based on a 100 GBASE-R Ethernet frame with FEC, using Ethernet-style alignment markers to stripe each 100 G interface over four FEC lanes on a 10 bit Reed–Solomon symbol basis. The FlexO frame, illustrated in Fig. 13.27, is organized as a 5440 bit by 128-row structure, repeating in an 8-frame multiframe. Each frame has Ethernet-style alignment markers, FlexO overhead, and an RS(544, 514) FEC (the same one used by Ethernet) based on 10 bit symbols. The first seven frames of each 8-frame multiframe contain 1280 bit of fixed stuff (indicated by FS in the diagram). The amount of fixed stuff is calculated such that each 100 G PHY operates at nearly the same bit rate as the OTL4.4 interface for OTU4 described in Sect. 13.8.2. This allows module vendors to design and test 100 G pluggable modules for support of only two bit rates of operation (Ethernet and OTN), with the OTL4.4 format used for OTU4 and the FlexO format used for each 100 G slice of an OTUC*n*.

A 2018 update of [13.11] adds formats for FlexO over 200 G and 400 G Ethernet pluggable modules. These modules are supported by interleaving two or four of the 100 G *FlexO Instances* used over a 100 G module over a 200 G or 400 G module. Also introduced is the possibility on higher rate modules to leave some of the 100 G FlexO Instances as unequipped. For example, it is possible to support a 300 G OTUC3 client interface over two 200 G optical modules with one unequipped FlexO instance. As mentioned with 100 G above, ITU-T optical interfaces corresponding to the use of 200 G and 400 G pluggable Ethernet modules can be found in [13.10] and [13.9].

13.9 Conclusions

OTN and related standards for Ethernet and flex Ethernet provide a rich toolset to address a wide range of network applications.

Simple network applications (e.g., point-to-point data center interconnect) require only a few of these tools, e.g., a single mapping of an Ethernet client over a line interface with path monitoring. Simple implementations, e.g., pluggable transponders, are available to serve these simple network applications.

Larger, more complex networks, e.g., multiservice networks spanning the administrative domains of multiple network operators, may require using additional capabilities defined within the OTN suite of standards to perform traffic management, service assurance, and fault isolation. Equipment used in these types of networks may range from *pizza box* type network elements at the network edge to collect traffic, to ROADMs, to large OTN crossconnect systems where digital grooming of traffic may be required.

OTN is not inherently complex, but some of the networks in which it is used are. OTN can be as simple or as complex as it needs to be for any given network application. A summary of specifications enabling the use of the same commonly available pluggable modules for Ethernet or OTN client interfaces is provided in Table 13.11.

Table 13.11 Use of ethernet pluggable modules for OTN client interfaces

Ethernet interface [13.5]	ITU-T optical specification	ITU-T frame format
40 GBASE-LR4 (clause 87)	[13.10] C4S1-2D1	[13.1] STL256.4
40 GBASE-ER4 (clause 87)	[13.10] C4L1-2D1	[13.12] OTL3_4
100 GBASE-LR4 (clause 88)	[13.9] 4I1-9D1F	[13.12] OTL4.4
100 GBASE-ER4 (clause 88)	[13.9] 4L1-9C1F	[13.11] FOIC1.4
CWDM4 MSA	[13.10] C4S1-9D1F	
4WDM MSA 40 km	[13.9] 4L1-9D1F	
200 GBASE-FR4 (clause 122)	[13.10] C4S1-4D1F	[13.11] FOIC2.4
200 GBASE-LR4 (clause 122)	[13.9] 4I1-4D1F	
400 GBASE-FR8 (clause 122)	[13.9] 8R1-4D1F	[13.11] FOIC4.8
400 GBASE-LR8 (clause 122)	[13.9] 8I1-4D1F	

13.10 Interoperable OTN Line Interfaces

As observed earlier in the chapter, optical networks may be constructed using a combination of fully standardized interfaces, and functionally standardized proprietary interfaces. Longer reach (1000s of km or subsea) and higher signaling rate interfaces (given the current state of the art, greater than 100 G) are normally proprietary interfaces, where there is a need not to sacrifice any performance by including extra margin to account for different vendor designs of the transmitter and receiver.

Shorter reach, lower signaling rate interfaces are feasible to be fully standardized. The current state of technology development is that 10 G NRZ or 100 G DP-DQPSK (dual-polarization differential quadrature phase-shift keying) optical signals carried over an amplified metro ROADM network with reaches up to approximately 450 km using 50 or 100 GHz grid spacing. The *black link* methodology is used for specifications of these interfaces, as it is impractical to exhaustively specify the performance of every piece of equipment that may appear between the transmitter and receiver, and while real network configurations have been used to help establish parameter values, the standard only specifies the transfer function that is expected to be supported between the transmitter and the receiver at a single-channel reference point. The optical specifications for these interfaces can be found in [13.13].

The digital frame formats for 10 G NRZ are specified in [13.12]. As 100 G DP-DQPSK requires a higher-gain FEC, the format to carry an OTU4 over a single 100 G wavelength using a G.698.2 application code is specified in [13.14]. The format for carrying an OTUC*n* over *n* 100 G single-channel interfaces using G.698.2 application codes is specified in [13.15].

References

13.1 ITU-T Recommendation G.707/Y.1322: Amendment 2. Network node interface for the synchronous digital hierarchy (SDH) (2009)

13.2 ITU-T Recommendation G.7041: Generic framing procedure (2016)

13.3 K. Kazi (Ed.): *Optical Networking Standards: A Comprehensive Guide for Professionals* (Springer, Berlin, Heidelberg 2006)

13.4 S. Frigerio, A. Lometti, J. Rahn, S. Trowbridge, E.L. Varma: Realizing the optical transport networking vision in the 100 Gb/s era, Bell Labs Tech. J. **14**(4), 163–192 (2010)

13.5 IEEE Standard 802.3-2018: Standard for ethernet (2018)

13.6 Optical Internetworking Forum: *Flex Ethernet Implementation Agreement – Version 2.1* (OIF, Fremont 2019)

13.7 IEEE Standard 802.1AX-2020: Standard for Local and Metropolitan Area Networks – Link Aggregation (2020)

13.8 ITU-T Supplement 56 to the G-series of Recommendations: OTN transport of CPRI signals (2016)

13.9 ITU-T Recommendation G.959.1: Optical transport network physical layer interfaces (2018)

13.10 ITU-T Recommendation G.695: Optical interfaces for coarse wavelength division multiplexing applications (2018)

13.11 ITU-T Recommendation G.709.1: Flexible OTN short-reach interface (2017)

13.12 ITU-T Recommendation G.709: Interfaces for the optical transport network (2016)

13.13 ITU-T Recommendation G.698.2: Amplified multichannel dense wavelength division multiplexing applications with single channel optical interfaces (2018)

13.14 ITU-T Recommendation G.709.2: OTU4 long-reach interface (2018)

13.15 ITU-T Recommendation G.709.3: Flexible OTN long-reach interfaces (2018)

Stephen J. Trowbridge

Optical Networking Business Unit
Nokia
Boulder, CO, USA
steve.trowbridge@nokia.com

Dr Stephen J. Trowbridge is Consulting Director of Standardization at Nokia. He received his BS (EE&CS), MS (CS), and PhD (CS) from the University of Colorado, Boulder. He has been active in optical networking standardization since 1995. He is chairman of ITU-T Study Group15. He is Editor of the Optical Internetworking Forum (OIF) Flex Ethernet implementation agreement. He was named a Bell Labs Fellow in 2014.

14. Traffic Grooming

Rudra Dutta, Hiroaki Harai

A particular thread of research in optical networking that is concerned with the efficient assignment of traffic demands to available network bandwidth became known as traffic grooming in the mid-1990s. Initially motivated by the distinctly different network characteristics of optical and electronic communication channels, the area focused on how subwavelength traffic components were to be mapped to wavelength communication channels, such that the need to convert traffic back to the electronic domain at intermediate network nodes, for the purpose of differential routing, was minimized. Over time, it broadened to include joint considerations with other network design goals and constraints. It was influenced in turn by existing technology limitations, and in turn served to influence continuing technology trends. Traffic grooming has had a significant effect on both the research and practice of transport networking. It continues to be a meaningful area not just in historical terms, but as a wealth of techniques that can be called upon for considering the traffic engineering problem afresh as each new development at the optical layer, or change in economic realities of networking equipment or traffic requirements, redefines the conditions of that problem.

14.1 Factors Motivating Traffic Grooming

In the few years after its inception, research on the original traffic grooming problem grew very quickly to a significant volume. This indicates that the time was right for such studies, and in retrospect, a confluence of a number of factors can be seen to have brought this about. Before describing the original or classical grooming problem in Sect. 14.2, we briefly outline these factors in Sect. 14.1, to provide understanding of the context in which that problem was posed. It is necessary, for this perspective, to retrace some commonly known network history of this period of emergence.

The development of Internet traffic that was instrumental in introducing the traffic grooming problem occurred at similar times, but not precisely the same times, in various parts of the world; the rest of this section is written from a typical perspective, since similar trends were generally obtained over time across the world.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_14

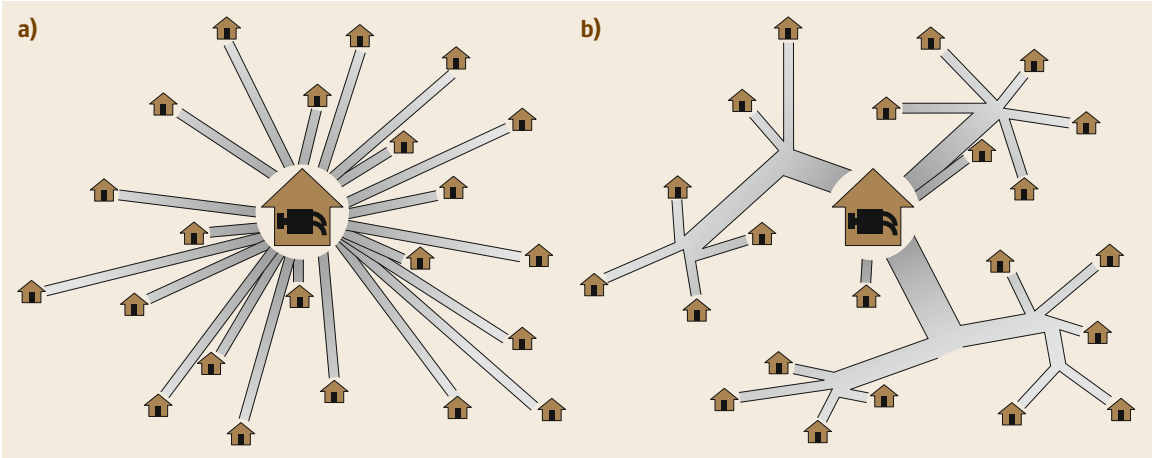


Fig. 14.1a,b The last-mile problem: (**a**) an extreme solution, (**b**) a more balanced solution

14.1.1 The Last-Mile Problem and Multicommodity Flow

In designing and dimensioning any capacitated network, be it a network of communication channels, fluid flow pipes, road transport, or any other, the so-called *last-mile problem* is a ubiquitous one. Consider a city water distribution system, where a single pumping station pumps water into distribution pipes, which eventually supply water to individual households. The last-mile problem refers to the fact that for any such distribution system, the complexity of the pipe system, and the number of individual pipe segments, will be dominated by the final stage of the distribution, the so-called *last mile* on the way to individual houses. Figure 14.1 shows the basic dilemma. In Fig. 14.1a, we show an extreme possibility of what the pipe system may look like. In this solution, individual small pipes are laid direct from the central pumping station to each individual house. This solution is clearly untenable from a scalability point of view. However, even if it were practicable for a given scale, it still might not be the most cost-effective solution, depending on the cost of lengths of pipe (and laying them), and the cost of junctions that would be required to allow a large-bore pipe to feed multiple smaller-bore pipes. Fig. 14.1b shows a more general solution, in which a combination of pipes of different bores, together with junction boxes, is used to perform the same distribution. Typically, one would assume that larger-bore pipes cost more per unit of length, but that the cost increases less rapidly than the capacity of the pipe (i.e., a double capacity pipe costs less than twice as much per meter); this is likely to be true of both the pipe itself and ancillary (sometimes larger) costs such as laying the pipe. Thus, cost savings are obtained by introducing aggregation of the commodity to be distributed (water, in this example) at the larger-scale end. On the other hand, such a solution requires pipe junctions. Obviously, similar problems can be posed for *collection* systems rather than *distribution* systems, as when wastewater is gathered from households for transport to one central wastewater treatment plant. Lastly, the essential nature of the problem remains unchanged if the number of pumping stations (or wastewater treatment plants) is a few, rather than one.

In every real-world problem, there are other costs and constraints on possible implementations, but most if not all of them may be abstracted into (i) constraints regarding the points at which branching (disaggregation) or aggregation is possible, (ii) costs of lengths, and (iii) costs of junctions. This is amenable for abstraction into a graph problem. An optimal solution may be found to this abstract problem for any combination of these costs; for many natural cost measures, the scalability considerations will cause an optimal solution to employ some degree of aggregation.

In practice, the locations at which branching is possible, and which pairs of locations are feasible to lay a pipe between, may often be constrained by other issues such as zoning, real-estate prices, availability of utilities, etc. In such a case, it is often more meaningful to consider those locations and adjacencies as being given parameters for a particular problem instance, and the only problem to be solved is how much capacity to provision for each possible pipe, so that sufficient water may flow from the pumping stations to each consumer (or from each consumer to the wastewater treatment plants). As is obvious, this is equivalent to the graph problem of *flow maximization* of a single commodity, a problem of polynomial complexity that is readily solvable optimally by algorithms such as the *Ford–Fulkerson* method [14.1]. (As we have stated it above,

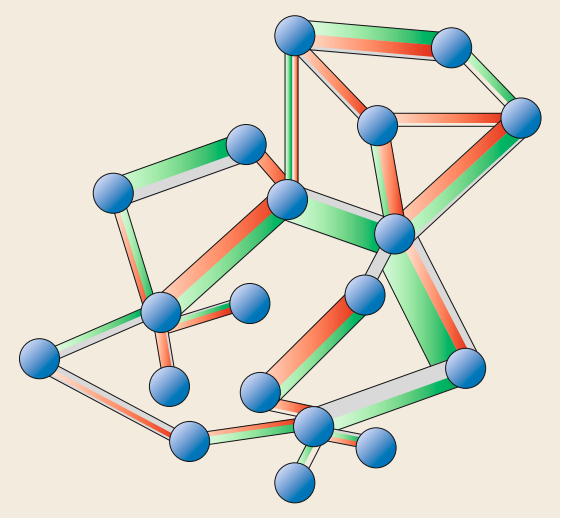


Fig. 14.2 The multicommodity flow problem

the problem is the *decision* version of deciding whether given flow demands can be satisfied, rather than the corresponding *optimization* version of discovering the maximum possible flow. But as with most problems, the two versions are readily interconvertible.) This remains true even if the last-mile problem is on both the production and the consumption side, i.e., there are many small sources as well as many small destinations, and if the demand/supply of each destination/source are allowed to be independent parameters. Such a problem is a *single-commodity* flow problem, since all the commodity (water) is of a single type, and any unit of it, wherever produced and however routed, will equally well satisfy any unit of consumption demand by any consumer. Further, with a single-commodity problem, there are hopes that a large network instance can effectively be decomposed into multiple smaller network instances, by imposing locality constraints (so that consumers are fed by producers that are local to them, rather than ones far off).

However, the essential difficulty of the problem changes drastically if there are multiple commodities that are distinct, so that the demand for one commodity cannot be satisfied by any quantity of a different commodity. With the same formulation as above, maximizing flows or satisfying each consumer's demands for each commodity, the multicommodity flow (MCF) problem, is well known to be NP-complete, even with no more than two commodities [14.2]. Figure 14.2 illustrates a solution to an instance of a two-commodity flow problem, and it is readily apparent that it is far more difficult, even intuitively, to judge the relative quality of two such solutions.

In practice, for a small number of commodities, approaches based on single-commodity solutions followed by reconciliation of flow capacities can yield reasonably good solutions at no more than polynomial cost. However, this is precisely the aspect in which a large-scale entity-to-entity messaging system, such as a communication network, is particularly at a disadvantage. When every endpoint (source or producer) expects to be able to address messages to any other endpoint (destination or consumer), every such ordered pair of nodes in the network gives rise to a unique commodity. Thus, in a network that connects N endpoints, the number of distinct commodities is $N(N-1)$. The idea of locality also disappears, since many more source-destination pairs are far off from each other than pairs that are close. Hopes for simple solutions that approach optimality are increasingly misplaced as N increases. It is for this reason that the last-mile problem is particularly vicious in increasing the complexity of the design task for any-to-any communication networks, such as the telephone system, postal system, and the Internet.

14.1.2 Synchronous Optical Network/ Synchronous Digital Hierarchy Optical Transport

The synchronous optical network (SONET) and synchronous digital hierarchy (SDH) are related standards, resembling each other so closely as to be essentially the same, especially for the purpose of our current discussion. The former was advanced from an industry consortium, standardized jointly with the International Telecommunication Union (ITU), and gained more ground in the United States; the latter was more a European standard, largely standardized by the ITU. In what follows, we shall refer to SONET/SDH, or simply SONET, to indicate their common characteristics. Originally defined exclusively for ring networks, SONET transport was generalized later to be capable of serving networks of general topologies. They are more fully described in previous literature such as [14.3]. In this section, we merely refer to the specific characteristics that motivate our discussion.

From the grooming point of view, SONET/SDH is a rather important entity. Starting with the introduction of the T-carrier system in the United States in 1961 (and other similar hierarchies in other parts of the world), the originally analog transmission systems for telephony adopted digital systems. This transition naturally defined the adoption, from the very beginning, of a *base rate* defined as the digitization of a single voice line. Successively higher rates were naturally formed by multiplexing an increasingly larger num-

ber of lower-rate channels, thanks to the hierarchical nature of the telephony architecture; this is the origin of the term *digital hierarchy*. All such digital standards, until the emergence of SONET, were designated *plesiochronous digital hierarchies* (PDHs), which indicates that all components of a system have the same clock rate (within a bounded delay), but may have different phases. The technological inability of such systems to guarantee a phase synchronized system-wide clock implied that at every interface between communicating network elements there was need for potential synchronizing activity at the beginning of every transmission, utilizing a preamble or hunting phase, and consequent buffering. Such systems might employ optical transmission, but it could only be for short spans; between points of such resynchronization requirements.

SONET/SDH represented a move into a *synchronous* architecture, for the first time, in the mid-1980s. This meant that SONET/SDH was able to perform multiplexing in a strictly time division multiplexed manner. At every network element, input and output clocks were synchronized; thus tight synchronization of all the channels at all the different rates was possible across an entire SONET network. Jitter and phase difference could certainly occur at the ingress to the SONET network, but this could be taken care of by introducing the appropriate amount of buffer delay at the ingress. However, once they had entered the network, the payloads and frames remained synchronized throughout. This approach made it possible for the network designer to consider the issue of hierarchical multiplexing (of lower-rate traffic streams into higher-rate channels) as one that was well and truly abstracted by the network. Further, SONET introduced the concept of *concatenation* to efficiently carry higher rates than the base rate, as well as *virtual tributaries* to efficiently carry subrate payloads. Together, these mechanisms provided *virtual concatenation* (VCAT) capability. In this sense, in SONET/SDH it was possible for the first time to consider multirate capability as a service that the network provided. Later, developments in next-generation SONET (NG-SONET) would introduce mechanisms naturally built over these capabilities, such as the *link capacity adjustment scheme* (LCAS), which allows VCAT channels to be resized to higher or lower rates without requiring stopping and starting traffic flows, and thus provides additional powerful abstractions to the network designer.

Most importantly, in an essential sense, synchronization allowed *networking* in a true optical sense even with aggregated traffic. Prior to SONET, an optical communication signal arriving via a fiber at an intermediate node could be forwarded optically to an outgoing fiber only if it was an analog signal; for digital signals such as in a PDH, the requirements of reacquiring synchronization to the digital signal implied the need to process and buffer the bits in such a signal, thus requiring so-called opto–electro–optic (OEO) conversion: the incoming optical signal would be converted to electronic form, processed by an electronic part of the network element capable of recognizing bits and bytes, and reconverted to optical form for transmission on an outgoing fiber. In SONET, if the constituent bits of a frame did not need to be changed at an intermediate node, that frame could be passed through the node as an optical signal, without OEO conversion, since it would remain synchronized. This would not only eliminate latency, but electronic switching equipment incomparably costlier than simple optical forwarding. Of course, such a potentiality was not possible to realize in the original single-wavelength SONET networks, in which each fiber carried only a single optical signal. It had to wait until the advent of multiwavelength SONET rings, utilizing wavelength division multiplexing (WDM), and the further development of standards such as optical transport networking (OTN) to fully utilize the capability of such *optical bypass*.

It is interesting to note that at the time, this feature of SONET was not seen as a capability of the network to provide variable rate service to customers, since all customers were considered to have a unique single-rate device—the telephone. The flexible aggregation/disaggregation capability represented by SONET's virtual tributaries, and all-optical forwarding, enabled by synchronous operation, were seen as internal features of the network. Nevertheless, the potential of an optical network to groom traffic can be traced back to this development.

14.1.3 Transport Networks, DWDM and Virtual Topology

As we have previously pointed out, the last-mile problem is an expression of the simple truth of large distribution networks; for scalability, a hierarchy of aggregation is inevitable in such networks. This in turn implies that higher capacity is required for the links higher up in the hierarchy, since in general more traffic is aggregated onto those links. As long as the same basic technology is used for the transmission of bits at various levels of the hierarchy, the only way to obtain higher capacity at higher levels of the hierarchy is to lay multiple links, or cables, of the type that make up the lowest levels of the hierarchy. This, coupled with the much larger distances that such links would have to span (hundreds of miles as opposed to hundreds of feet), made the links by far the larger fraction of the cost of the network at these higher levels of hierarchy, as compared

to the intermediate nodes. This was the state of the network at the time of the inception of the network, when the telephone network was the unique communication network that achieved nearly planetary span; this continued to be the case even after the emergence of the Internet (after the successive experimental and gestational phases of ARPANET and NSFNet), before the use of optical fiber communication became practical and commercially available. Accordingly, at the time a clear split in the network business was perceived by all concerned—network operators and engineers, equipment manufacturers, and researchers. This split has been characterized as that of the *transport network* and the *traffic network* by [14.4], and for the present discussion we adopt these terms.

On the one hand, network service providers with large networks of national footprint, or backbone networks, dealt with networks composed of these very costly, high-level links. In consequence, bandwidth was scarce, and every unit of bandwidth was precious. In the early days of planetary data networking, the proprietary telephone networks were the only real at-scale bandwidth providers at national and international scales, and a model of paying for bandwidth at this level very naturally came into being, since the bandwidth to be used for Internet traffic was bandwidth that required significant capital expenditure to install and operating expenditure to maintain and operate, and could otherwise profitably be used in carrying voice traffic. Data networking over leased telephone lines consisted of leased lines operated by business customers to run their organizational networks, using protocols such as X.25 or switched network access (SNA), and later TCP/IP (Transmission Control Protocol/Internet Protocol). Such sizable business customers used leased lines for their Internet access, and individual customers soon followed with modem access (through second- and third-tier service providers). Traffic on these *transport networks* was, therefore, consolidated traffic, and could be considered more or less stable, if only because the total traffic was likely to be limited by some part of the backbone network.

On the other hand, the local networks that served individual customer premises were composed of a larger number of endpoints, distributed over a comparatively tiny geographical span. Installing bandwidth was neither costly nor difficult—often consisting of nothing more than patching a few more cables in some closet. Accordingly, in these *traffic networks*, bandwidth was nearly free, and almost always overprovisioned; consequently, the network links were typically sparsely utilized, and traffic was very bursty.

Traffic engineering, or the attempt to control traffic by the use of routing as well as other mechanisms such as timing, shaping, policing, etc., is clearly more appropriate for the transport networks, since traffic is more predictable (thus traffic engineering is *possible*), and bandwidth a costly resource that must be well utilized (thus traffic engineering is *necessary*). Accordingly, a good deal of attention, in both the research and practitioner communities, went into traffic engineering for backbone networks. This line of work was the natural successor for traffic engineering and routing in telecommunications networks. For the purpose of engineering traffic for transport networks, or planning future deployments or resource provisioning, it was not useful to consider the traffic volumes offered by the traffic networks explicitly—rather, it was natural to model the aggregated traffic offered to the transport networks as essentially stationary distributions, in which the variations and burstiness of individual traffic networks had been smoothed out by superposition of many such traffic streams, as well as shaping/smoothing due to admission control and the long-haul bottleneck.

This situation exhibited only a special case of a general truth in any large, complex, distributed system: such a system, unless very precisely designed and maintained, is going to have one or more bottlenecks at any time, although such bottlenecks can move over time. The introduction of optical communication introduced precisely such a move. The bitrate capacity of even the earliest commercial fiber-optic transmission systems represented a jump of more than an order of magnitude over that of the largest bundle of long-haul electronic transmission channels, and kept improving frequently and significantly, a trend that is still continuing without much abatement at the time of writing. Further, while the manufacturing needs for increasingly more sophisticated fiber were very exacting, after the original installed capacity of manufacturing plants, the mass-production cost of fibers quickly shrank to very low values, and bundles of fibers were laid even for links where the bandwidth of a single fiber was already more than the traffic projected to be carried by the link for current conditions. One of the most significant jumps in fiber bandwidth was the development of dense wavelength division multiplexing (DWDM) technology, allowing the use of many simultaneous independent communication channels to be realized over a single optical fiber, by the use of multiple transmitting lasers and receiving photodetectors operating at slightly different center frequencies, or wavelengths. The term *dense* was applied only to distinguish such systems from predecessor systems that multiplexed two (or a few) such wavelengths on a single fiber, typically to achieve duplex communication. DWDM systems, in contrast, achieved 16 wavelength channels on a fiber early on, and 40 or 80 soon thereafter; currently so-called ultradense systems can achieve over 300 channels.

Almost overnight, the transport networks had shifted to a regime where the bandwidth of the long-haul links were no longer the bottleneck—the bandwidth capability of the backbone suddenly outstripped the current demand by far. On the one hand, to some extent this weakened the position that backbone networks were inevitably bandwidth constrained, and that careful resource constrained dimensioning and design was necessary for these networks. On the other hand, there were several factors that tended to reinforce the existing view. Obviously, the very increase in transport network bandwidth made the gulf between the magnitude of individual traffic flows and backbone network capacity even higher, making it truer than ever that the individual flows of traffic contributed to by traffic networks were insignificant with respect to the transport network design, and were only important in aggregate. Secondly, despite the eventual low price of fiber, backbone network operators were initially faced with a large capital expenditure to install fiber, as well as the costly optical switching equipment. Thus the transport network became a costlier resource than ever, and the need to use it most efficiently so as to maximize revenue remained paramount.

One of the reasons that a DWDM optical switch is costlier is that every fiber carries W wavelengths, and each such wavelength channel on each incoming (and outgoing) fiber needs its own optoelectronic line terminating equipment, analogous to SONET's add-drop multiplexer (ADM), not to mention the multiplexer/demultiplexer that each outgoing/incoming fiber requires to separate the signals on different wavelength channels. Even a single wavelength channel on a single fiber was quite a high-speed channel by the standards of the time, and the corresponding electronic equipment to cross-connect these electronic channels of traffic (analogous to SONET's digital cross-connect) was costly. In contrast, if all the traffic on a particular incoming wavelength channel were bound for the same outgoing wavelength channel on some outgoing fiber, then the entire wavelength channel could be forwarded as an optical channel in the switch, eliminating a large number of ADMs, and using only comparatively cheaper optical devices. This gave rise to the research area of *logical topology* or *virtual topology* design, which produced a rich literature of its own, starting with early work such as [14.5, 6]. This body of research has been extensively reviewed in the literature, and is dealt with elsewhere in this handbook as well. Below we provide a minimal definition, purely in order to be able to mark the contrast with the traffic grooming problem, in a later section of this chapter.

The essential problem is shown schematically in Fig. 14.3. Given a set of *lightpaths*, or optically continuous wavelength channels, that were desired to be established between node pairs of the given backbone network, the task was to realize them by picking a route for each, as well as the wavelength it would be realized on, so that the solution could be feasibly achieved by the optical switches the nodes were equipped with. For this reason, the problem is also known as *routing and wavelength assignment*, and is described more fully elsewhere in this handbook. Both G_p and G_v are allowed to be multigraphs, since more than one fiber link may be laid, or more than one lightpath may be desired, between the same pair or nodes. Each lightpath acted like an end-to-end physical circuit, being implemented by optical equipment that operated at the subbit level, and offered essentially constant end-to-end latency; hence the terms of logical/virtual topology. Indeed, backbone network providers looked to such topologies to leverage their costly new DWDM optical network to provide higher revenue by essentially providing multiple customers with networking service, each with their own virtual, customized network topology.

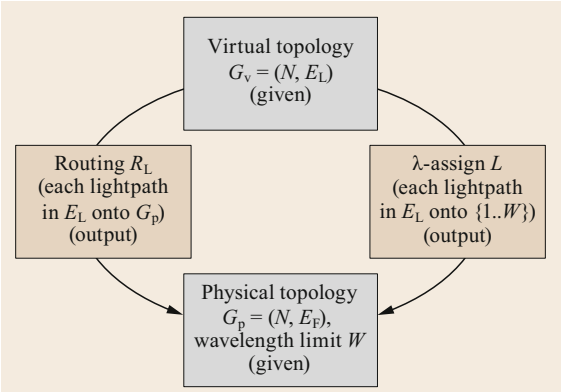


Fig. 14.3 The virtual topology problem

With the virtual topology problem, the design view thus continues to be that for transport networks, the design input is aggregated (and stable) traffic demands (the lightpaths); no explicit representation of lower-level traffic is attempted.

The gains represented by an efficient virtual topology design (that minimized the need for OEO conversion of traffic, ideally eliminating it by routing each lightpath in the demand completely optically end-to-end) would be only fully realized if the large number of ADMs potentially saved by virtualizing the topology could be actually eliminated from the optical switches, thus making it possible for much simpler and cheaper switches to realize the network. However, with such hardware switches it is only practical to customize a network with such nodes if the traffic demand (desired lightpath graph G_v) is very nearly static, and

a network solution put in place in hardware can be expected to serve the demand for a long enough time to make the investment worthwhile. With the continuing assumption that the aggregate traffic demand at the backbone was stable, this seemed practical, but it is obvious that for dynamic virtual topologies to be instantiated, it would be necessary to flexibly utilize the limited number of ADMs at the switch. The development of reconfigurable optical ADMs (ROADMs) a few years later represented precisely this capability, so that each of a limited number of ADMs could be used to terminate or passthrough any wavelength channel between incoming and outgoing fibers, and thus virtual topologies could be modified on demand. As we will see later, this provided a required piece of the grooming problem.

14.1.4 Access Networks and the Rise of Traffic

As we previously mentioned, the early data networks of national or international scale were provided largely by telecommunications operators (except for ARPANET created on an experimental basis), and the early users of data communications obtained services from them either on an individual basis, using modem calls, or organizationally in bulk, using leased lines, and later, leased topologies. Individually, the modem calls represented a vanishing fraction of both the network usage in total number of calls, and the total data network usage. (Some individual users were privileged to have T1 connections to their organizational network access; we count them among the organizational use model.) However, this evolved very quickly, powered by the growth of the cheaper, more powerful, ubiquitous personal computer (PC). It is at this time that the open, publicly available and realizable suite of Internet protocols began to first catch up with, then far outstrip, closed proprietary data network protocols.

Holding times for the fictitious *telephone calls* (actually representing dial-in by individual users to their providers for their data communication needs) were orders of magnitude higher than true voice calls, and with growing numbers, upset the model of network traffic demand that telecommunications operators had depended on for decades. The number of individual users, and hence traffic flows (commodities in terms of the MCF problem we introduced in Sect. 14.1.1), exploded from dozens to a few hundred, and quickly to millions and beyond (until now it stands at billions). With increasingly faster computing available in PCs, a large variety of applications capable of producing and consuming data at an increasing rate sprang up. Through the mechanisms of dial-up, Internet traffic (often referred to as *packet traffic*, to distinguish from the classical *circuit traffic* of voice calls) started accounting for a nonnegligible amount of the traffic on the backbone, but on the backbone itself Internet traffic received its share of the bandwidth only on lines leased by Internet service providers (ISPs), who typically leased such bandwidth conservatively. The need for intermediate network scales, between typical local area networks and the wide area (national or international) networks, became paramount, and it became a rewarding business model to provide such networks.

There is another factor related to the reprovisioning of backbone networks with optical equipment, as we mentioned in Sect. 14.1.3. Providers able to make the transition became capable of much faster networking, while those that decided against it took on the mantle of a lower-tier network, capable of consolidating less of the exploding traffic demand. There was also a natural correlation with the locality of such network providers; networks that were limited in geographical span naturally took on the mantle of regional networks. This created a hierarchy of network operators, or *tiers*, typically indicated by terms such as tier 1, tier 2, etc. networks, with the tier 1 operators providing the widest backbones, and representing the highest level of traffic aggregation. Networks at the lowest level of consolidation and aggregation became known as access networks.

Although the access networks consolidated far less traffic than the backbone networks they were feeding that traffic into, this volume of traffic was significantly higher than the traffic flows of the individual, or small business, user. Thus, the traffic demands on the backbone networks underwent a fundamental change; they were still too small to be considered individual units of traffic to be routed by individual wavelength channels on the optical networks, but not small enough that they could be considered an undifferentiated fluid of packets. Such *subwavelength traffic demands* provided the final piece of the traffic grooming puzzle.

14.2 The Original Traffic Grooming Problem

In this mix of the state of technology, the promise of technology to come, the prospect of exponentially growing traffic, and the available traffic engineering algorithms, arose the concept of traffic grooming. The underlying observation of traffic grooming is that with the advent of DWDM, there was a mismatch in the cost of the equipment and operations necessary to serve the traffic flows in a large network with communication

utilizing optical transmission. Bit for bit, optical fiber, optical transmission, optical switching, and even fiber installation could be considered cheap—but only if considered in bulk. Conversely, electronic processing of the data for the purpose of making the forwarding decision loomed as a large cost, if it was to be accomplished even at a fraction of the data speeds made possible by the optical communication technology. Thus the last-mile problem reappeared with a vengeance, and with a two-fold aspect: the meteoric rise of electronic traffic endpoints and volumes, coupled with the extreme contrast of the costs of electronic versus optical processing. Without an efficient and practical solution of the problem, the only choice was between the following two extreme scenarios:

1. Perform minimal or no multiplexing of low-speed electronic traffic into optical wavelength channels. The cost of optical equipment, bit-for-bit, would become very high in this solution; it would severely underutilize the optical bandwidth, and use of optical transmission systems would simply not be justified.
2. Perform as much multiplexing as possible, packing each wavelength, but at each intermediate routing node, reconvert traffic into electronic form, suitable for traditional routing and forwarding processing. Optical channels could still be used for transmissions along individual links, but the bandwidth of such channels would become a double-edged sword—the increased bandwidth would require very high capacity of electronic routing at each intermediate node. Any cost savings due to the use of optical transmission (as opposed to traditional electronic) would become negligible in comparison. The use of optical channels would be irrelevant to the cost of the network.

Thus, the very future of optical transport in enabling the envisioned future planetary data communication network seemed threatened. Consciously or otherwise, many optical networking researchers converged toward answering this challenge, and for a time, traffic grooming enjoyed a somewhat large proportion of the attention among optical network design problems.

In the mid-1990s, research work addressing the efficient use of wavelength resources as well as electronic routing resources in multiplexing electronic traffic onto optical wavelength channels started appearing in literature, and by the end of the 1990s, the term *grooming* became generally accepted and used to refer to this problem. The decade that followed saw quite exhaustive research into this area, and focused on descriptions of the problem that were variations on the same basic problem. This literature has been adequately surveyed and described elsewhere [14.7–11], and we do not attempt another complete description here. In the rest of this section, we describe the essential grooming problem, as originally studied.

14.2.1 Essential Grooming Problem Description

In this original form, the traffic grooming problem can be seen as the problem of aggregating subwavelength traffic onto high-speed lightpaths, minimizing the optoelectronic equipment cost of the network. It can be considered an optimization problem, and simply described as in Fig. 14.4. Contrasting this with Fig. 14.3, which describes the virtual topology problem mentioned in the previous section, we can see that while the virtual topology design problem may start with the number of lightpaths between every source and destination (the lightpath demand matrix) as a given input, in the grooming problem the corresponding input is the subwavelength traffic demand matrix, and the grooming problem extends the virtual topology design problem by requiring (i) that the lightpath demand matrix itself be computed, and (ii) that the subwavelength traffic demands be mapped onto the lightpath demand matrix. Thus, traffic grooming is a joint two-layer routing problem; however, in optimization terms, the constraints of

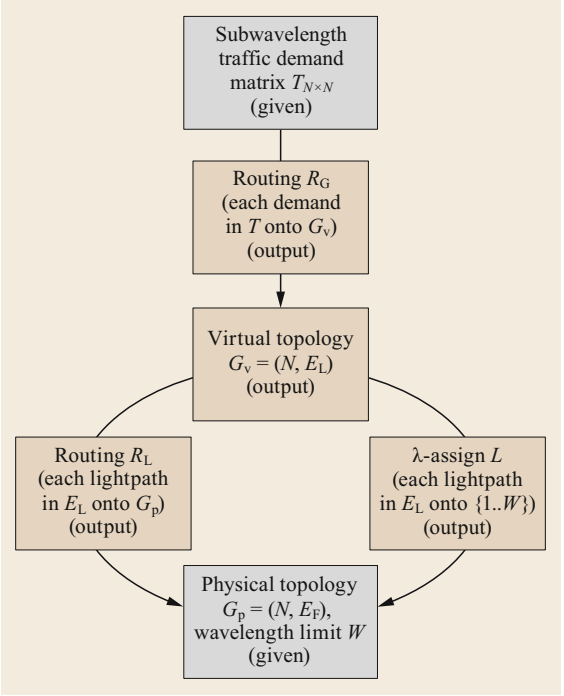


Fig. 14.4 Traffic grooming problem

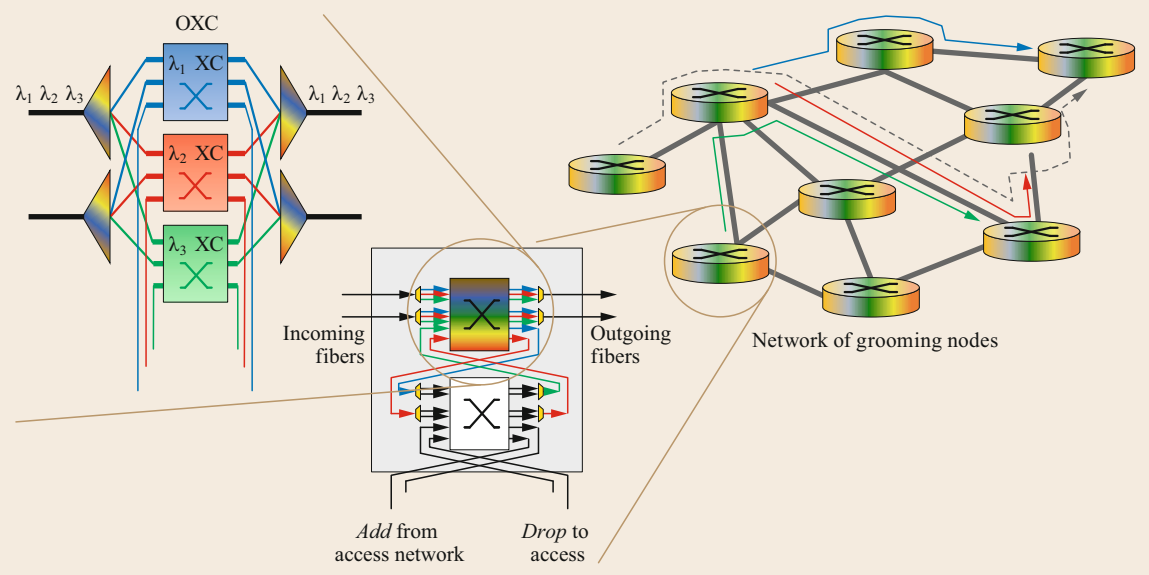


Fig. 14.5 Schematic switch architecture for optical and electronic switching

the two layers are different (the wavelength routing layer has the same wavelength continuity requirement as before). In the simplest form of the grooming problem, the impact on the objective function comes from only one of the routing layers—the cost is assumed to be proportional to the amount of traffic that has to be electronically forwarded from one lightpath onto another, whenever a subwavelength traffic component traverses a sequence of multiple lightpaths in its journey from source to destination. The virtual topology problem (routing as well as wavelength assignment) thus turns out to be subproblems of the traffic grooming problem. This can be seen as the mathematical reflection of the observation we made above—that with traffic grooming research, optical network design researchers subsumed the responsibility of addressing the needs of the end-to-end traffic, not just the needs of the transport network.

The schematic structure of a generic optoelectronic switch that embodies the two-layer routing capability, and hence is suitable for use in a traffic grooming network, is shown in Fig. 14.5. We hasten to point out that this is merely a schematic representation of the capabilities of any such switch, not necessarily the physical architecture of such a switch. The middle of Fig. 14.5 shows the individual grooming switch, while on the right we see its use in the broader networking context, and the left shows (schematically) the details of only the optical part of the switch.

The grooming switch can be seen to be a combination of two cross-connecting fabrics. The top part of the switch as diagrammed is an optical cross-connect (OXC), capable of demultiplexing each of the W wavelengths from each incoming fiber, and switching each wavelength independently to some outgoing fiber. This independent switching capability is represented in Fig. 14.5 by W separate optical switches that are *colorless* (such equipments provide the same functionality independent of the wavelength of the optical signals injected into them, and do not modify the wavelength of the signals they process), each devoted to the switching of all the lightpaths of a given wavelength at that network node. Any of the wavelength channels may also be dropped to a digital cross-connect (DXC) that forms the lower part of the grooming switch. Similarly traffic from the DXC can be added and transmitted out on some wavelength channel on some outgoing fiber as desired.

The lower part of the grooming switch represents the DXC, and therefore all the OEO conversion capability at this network node. It can similarly demultiplex subwavelength traffic streams digitally, and switch individual traffic flows (analogous to timeslots, or virtual tributaries, of SONET) from different dropped wavelength channels independently of outgoing subwavelength traffic flows multiplexed into outgoing wavelength channels. Individual subwavelength traffic flows can be added from the access networks served by this backbone network node, or dropped into those access networks.

The key observation is that the cost of this grooming switch is disproportionately accounted for by the digital part of the switch, although it handles traffic at a much lower rate, and (with good design) a much lower total

volume of traffic, than the optical part, due to the asymmetry in cost between optical switching and electronic processing/switching capabilities.

14.2.2 Mathematical Programming Formulation—Static Problem

From consideration of the multicommodity flow problem, it is easy to see that both the virtual topology problem, and the traffic grooming problem independent of the virtual topology subproblem, are NP-complete. Precise formulations of the problem are thus not generally useful in the sense that they do not practicably yield optimal solutions by automatic solving. Nevertheless, such formulations of the problem are essential, not only in obtaining a precise mathematical description of the problem, and to distinguish between specific variants, but also in yielding insights toward specific heuristic or approximation approaches. Further, generally available techniques for approximate solutions of large mathematical programming problems can be applied if such formulations are available. Many such formulations have been provided in the literature, and different formulations have differing advantages in terms of amenability to automatic solution, or to relaxation techniques, or other techniques such as the introduction of cutting planes. Typically, such formulations are either posed as integer linear programs (ILPs) or mixed integer linear programs (MILPs). Here, we simply provide one formulation, for the sake of completeness; it is crafted for ease of exposition rather than computational efficiency. It follows the classic formulation of the basic communication networking problem of [14.12], and the previous formulation of the grooming problem of [14.8].

In this formulation of the problem, we represent both physical fiber links, and lightpaths, to be directed. This is the more general representation; an undirected representation is in essence a constraint for fiber links and lightpaths to exist only in bidirected (and similarly routed) pairs, and possibly also a requirement for traffic demands to be similarly symmetric. We also assume each link of the physical topology to be a single fiber link. The parameters include the physical topology, specifically, p_{lm} is the physical *link indicator*; it indicates whether a link exists in the physical topology from node l to m. We use the following notation for lightpaths:

b_{ij} is the *lightpath count*, i.e., the number of lightpaths established in the solution, from node i to j

$b_{ij}(l,m)$ is the number of such lightpaths whose physical routing traverses the physical link from node l to m

$c_{ij}^{(k)}(l,m)$ is the *link lightpath wavelength indicator*; it is 1 if such a lightpath uses the wavelength k over the physical link from l to m, 0 otherwise.

All of the above variables are integers, and the last set binary integers. The remaining variables relate to traffic, they are also all represented as integers since we assume that traffic demands, as well as allocated flows, are all expressed as multiples of some basic rate (and therefore quantized). C denotes the bandwidth of a single wavelength channel in terms of that basic rate. Input parameters of the form $t^{(sd)}$ embody the demands of various source-destination node pairs on the network; $t^{(sd)}$ is the amount of traffic that needs to flow from node s to d. Collectively, these form the traffic matrix $\mathbf{T} = [t^{(sd)}]$. All such traffic demands must be carried by lightpaths of the virtual topology in the solution. We use the variable t_{ij} to denote the aggregate traffic carried by the lightpath (or lightpaths) from node i to j, and $t_{ij}^{(sd)}$ denotes the amount of this traffic that is due to the demand $t^{(sd)}$. With N denoting the number of nodes in the physical topology, the domain for each of i, j, s, d, l, m is $\{0, \ldots, N-1\}$. The number of wavelength channels W supported by each directional physical link is assumed to be given as a parameter. The domain for k is $\{0, \ldots, W-1\}$. Accordingly, we can formulate the traffic grooming problem as follows. The constraints are separated into subsets that embody the different subproblems (SP) in the grooming problem as described before.

Minimize (one of the following functions):
Total number of lightpaths

$$\sum_{i,j} b_{ij} \,, \tag{14.1}$$

total amount of electronic switching

$$\sum_{s,d,i,j} t_{ij}^{(sd)} - \sum_{s,d} t^{(sd)} \,, \tag{14.2}$$

maximum number of lightpaths at a node

$$\max_i \left[\max \left(\sum_j b_{ji}, \sum_j b_{ij} \right) \right] . \tag{14.3}$$

Subject to:
Physical topology constraints

$$b_{ij}(l,m) \le b_{ij} p_{lm} \,, \quad \forall i,j,l,m \,, \tag{14.4}$$

$$c_{ij}^{(k)}(l,m) \le p_{lm} \,, \quad \forall i,j,k,l,m \,. \tag{14.5}$$

Lightpath routing SP constraints

$$\sum_{l=0}^{N-1} b_{ij}(m,l) - \sum_{l=0}^{N-1} b_{ij}(l,m) = \left\{ \begin{array}{ll} b_{ij}, & m = i \\ -b_{ij}, & m = j \\ 0, & m \neq i,\ m \neq j \end{array} \right\} \quad \forall m,i,j\,, \tag{14.6}$$

$$\sum_{i,j} b_{ij}(l,m) \leq W\,, \quad \forall l,m\,. \tag{14.7}$$

Lightpath wavelength assignment SP constraints

$$\sum_{k=0}^{W-1} c_{ij}^{(k)}(l,m) = b_{ij}(l,m)\,, \quad \forall i,j,l,m\,, \tag{14.8}$$

$$\sum_{i,j} c_{ij}^{(k)}(l,m) \leq 1\,, \quad \forall k,l,m\,, \tag{14.9}$$

$$\sum_{l=0}^{N-1} c_{ij}^{(k)}(m,l) - \sum_{l=0}^{N-1} c_{ij}^{(k)}(l,m) \left\{ \begin{array}{ll} \leq b_{ij}, & m = i \\ \geq -b_{ij}, & m = j \\ = 0, & m \neq i,\ m \neq j \end{array} \right\} \quad \forall i,j,k,m\,. \tag{14.10}$$

Traffic routing SP constraints

$$t_{ij} = \sum_{s,d} t_{ij}^{(sd)}\,, \quad \forall i,j\,, \tag{14.11}$$

$$t_{ij} \leq b_{ij} C\,, \quad \forall i,j\,, \tag{14.12}$$

$$\sum_{j=0}^{N-1} t_{ij}^{(sd)} - \sum_{j=0}^{N-1} t_{ji}^{(sd)} = \left\{ \begin{array}{ll} t^{(sd)}, & i = s \\ -t^{(sd)}, & i = d \\ 0, & i \neq s,\ i \neq d \end{array} \right\} \quad \forall m,i,j\,. \tag{14.13}$$

Most of the constraints are self-explanatory. The physical topology constraints ensure that lightpaths can only traverse physical fiber links that exist, and that each wavelength on each fiber can only be used by at most one lightpath. Constraint (14.6) is a flow-balance equation of lightpaths on the fiber topology, and (14.7) ensures that no more lightpaths attempt to traverse a fiber link than the number of wavelengths available on that link. (It may be noted that (14.7) is redundant; we discuss this later.) Constraints (14.8), (14.9), and (14.10) together express the wavelength assignment problem; (14.9) enforces that at most one lightpath can be assigned a given wavelength on each fiber link; and (14.8) reconciles these assigned wavelengths with the total number of lightpaths traversing a given physical link on a per-source-destination basis. Constraint (14.10) is in effect the wavelength continuity constraint: at each node m, if this is an intermediate node, the number of lightpaths entering the node on any given wavelength must be balanced by those exiting the node on the same wavelength, on a per-source-destination basis. Finally, (14.11) and (14.12) enforce the capacity constraints, and (14.13) enforces the flow balance constraints, for the routing of subwavelength traffic on the topology of lightpaths.

14.2.3 Variants of the Traffic Grooming Problem

Many variants of the problem may be conceived of that differ in specific details, while falling under the general description of reducing the electronic routing overhead while most gainfully using the optical switching and transport capabilities of the network, given traffic requirements. Different variants may be more representative of the problem under various problem scenarios differing by envisioned switching equipment capabilities, expected network or traffic characteristics, or requirements or expectations posed by end-to-end traffic.

The three alternate objective functions given above in (14.1), (14.2), and (14.3) are examples of differences in representing and envisioning switch architecture and network resource provisioning. Objective (14.1) is simply the number of times a lightpath is terminated (source or sink) in the entire network: this is the number of optoelectronic ports that will be needed in all the switches combined. Some or all of the traffic in each such lightpath may be electronically switched to another lightpath, while the rest is exchanged with the access network, and not all lightpaths may carry a full load of traffic. But the cost of electronic equipment is likely to increase with the number of optoelectronic provisioned ports, whether such a port switches electronic traffic from/to another optical channel or to/from an access channel, and even if such a port is underutilized. Thus (14.1) tries to capture the realistic cost of optoelectronic equipment.

Objective (14.2), on the other hand, takes the position that the exact cost of each optoelectronic port may not be a fundamental quantity; it may change from network to network, and the per-port cost of a 48-port switch may not be the same as that of a four-port switch; so this representation attempts to focus on the unquantized number of bits that were OEO switched over the entire network, considering that all other things being equal, a network that requires less overall OEO switching will be possible to provision with less optoelectronic equipment. Objective (14.2) abstracts the prob-

lem to a higher degree than (14.1), and assumes less about the details of network provisioning and design.

Conversely, (14.3) attempts to assume more about how a network might be realistically provisioned by OEO switches. It attempts to minimize the maximum number of optoelectronic ports that a switch at any network node will need; this is in effect an assumption that all nodes will be provisioned with the same model of switch, with the same OEO switching capability—in this model, minimizing the OEO switching at some or most nodes is useless if this creates a very high burden of OEO switching at some node.

Other differential node capabilities could also be assumed, such as full or partial wavelength conversion capability. We remarked above that constraint (14.7) is redundant; in fact it is implicit in constraints (14.6), (14.8), and (14.9). However, this assumes that none of the nodes have any wavelength conversion capability. If, instead, we assumed full wavelength conversion capability to be available at each node of the network, the entire wavelength assignment SP, i.e., (14.8), (14.9), and (14.10), would be eliminated; in that case (14.7) would become operative in enforcing the number of wavelength channels supported on a fiber link by the transmission system. Similarly, a version of (14.10) (summing over all wavelengths k rather than individually for each) could be written that would make (14.6) redundant as well, but a feature of the formulation as presented is that the wavelength continuity constraint (14.10) can be relaxed on a per-node basis, making it easy to model the more realistic situation whereby some nodes possess wavelength conversion capability while others do not.

14.2.4 Dynamic Traffic Grooming Problem

In the above, we have assumed that all the traffic demands are stable, and can be represented by time-invariant quantities. As we have remarked above, this is a reasonable assumption for the virtual topology problem; aggregated demands at the level of wavelength bandwidths are unlikely to undergo significant changes rapidly. For the grooming problem, it was also initially considered reasonable, since the subwavelength traffic demands were themselves assumed to be aggregated traffic, perhaps at granularities of the OC-3 (about 155 Mb/s) level of SONET—a level that was unlikely to be produced by individual traffic flows subject to rapid variation. However, over time, individual traffic components grew in size, and emerging applications such as online video conferencing introduced traffic demands subject to short-term variations in rate that were comparatively heavy by existing standards. Considering the grooming problem under the model of subwavelength traffic demands that changed significantly over short timescales became a meaningful problem to address.

In terms of the formulation provided above for the static traffic grooming problem, it is reasonable to attempt to model the dynamic grooming problem by allowing the traffic matrix T to represent time-varying traffic. A class of problems can be conceived in which a problem instance is defined by providing a set of specific time epochs $\{t_i\}$, and one version of the traffic matrix T_i that is operative between each successive pair of epochs. Potentially, the set of matrices repeats cyclically, indicating periodic traffic sources—that is, the changes in traffic demand arise from periodic underlying phenomena that produce changes in the traffic demand. Such a problem provides a complete *future history* of the traffic, and requires a plan for the entire period of interest. It is possible to see such a problem as simply a combination of several separate static grooming problems, which can be solved separately, one solution per epoch. However, it is likely that in such a problem, additional constraints (or components of the objective function) will reflect the consequences of the fact that these solutions will need to be realized one after the other on the same network. For example, changing virtual topology typically incurs a *reconfiguration cost*, reflecting energy expenditure as well as latency during which traffic is not serviced. A particular version of the dynamic grooming problem of this class might require jointly minimizing OEO routing and reconfiguration cost—a single virtual topology that is not optimal for either of two traffic matrices in the sequence of input matrices might be the optimal one when the savings in avoiding reconfiguration are factored in.

A different class of dynamic grooming problem is obtained when instead of a certain future history, a random process is specified for some or all of the subwavelength traffic demands. For example, the magnitude of subwavelength traffic demands could be modeled as randomly varying with some specified distribution, and their lifetime could be modeled by a similar process, such as a Poisson process. It is also possible that there is no uncertainty in the actual traffic matrices, and a set of them $\{T_i\}$ is provided as in the previous paragraph, but the epochs $\{t_i\}$ at which traffic changes are random.

The grooming problem could be formulated to require actions from different action spaces. The traditional *call admission control* model could be used, so that the network accepts or rejects each new subwavelength traffic demand, and has to find appropriate routing for accepted demands in the current virtual topology, without rearrangement of currently resident lightpaths or subwavelength traffic. Alternatively, the action space can include actions to rearrange and mod-

ify the virtual topology, or subwavelength traffic routing over the virtual topology. A formulation might even allow elimination of resident traffic, if some traffic demands are deemed more critical than others.

For dynamic grooming problems (especially the latter class, where the nature of the variation is only known statistically rather than exactly), the objective of grooming is likely to be different from that of static grooming (although OEO minimization may continue to be a component of the objective function). Some possible objectives could be to minimize:

(i) The blocking probability, over all traffic demands
(ii) The provisioning time (time to setup a connection for an arrival, traffic delay, etc.)
(iii) The disruption to traffic already being carried
(iv) The unfairness (e.g., traffic demands having different bandwidth requests should have approximately the same blocking probability), or other similar goals.

Further variations in the definition of the problem are possible; *scheduled demands* might specify windows during which they can be satisfied (carried), rather than requiring immediate transport, or demands might change using an increment–decrement model, rather than an arrival–departure model, and so on. A somewhat more detailed discussion, and survey of some relevant literature, can be found in [14.13].

14.2.5 Further Evolution of Grooming

Many other variants were commended to researchers' attention by specific needs or envisioned characteristics of emerging or future network scenarios. Over time, three kinds of evolution can be seen to have occurred in the literature:

1. Continued consideration of the traffic engineering problem under opportunities or limitations posed by evolving switching and transmission systems. As optical communication and switching systems continue to evolve and improve, through advances such as higher speeds, wavebands, elastic or flexible-grid wavelengths, this research area continues to address the challenge of showing how well a network composed of such advanced optical systems can serve in carrying the traffic generated by the edge of the network, which in turn gets further away from the core with the continued growth of the untethered edge of the Internet, mobile devices, and machine-to-machine (M2M) traffic (all of which are even finer grained and even more dynamic than traditional Internet traffic before their introduction).
2. Finer or better traffic engineering, based on various specific requirements or expectations of traffic. This includes considering impairment-aware grooming, survivable grooming (described in more detail in Sect. 14.5), grooming for differentiated *quality of experience* (QoE), and other such topics. It also subsumes the general issue of grooming for time-varying traffic demands we mentioned above. As we mentioned, many different models of time-varying traffic demands exist; for example an incremental model where traffic demands suffer random bulk increase or decrease at random (or predetermined) epochs of time, or a scheduled model where traffic demands can declare their arrival well in advance of proposed network occupancy. Under such different models, the QoE of users can be very different with the same network provisioning of their traffic flows.
3. Better, or more novel traffic engineering, based on design objectives that reflect in whole or in part considerations that were not previously or traditionally considered the purview of traffic engineering, or even networking. Such studies have addressed green grooming (to minimize energy usage of the network), grooming as an economic tool, integration of grooming and software-defined networking, and similar such topics.

However, in general quite a large variety of grooming research has been driven by a basic sequence of considerations that are surprisingly similar to its original inception: (i) The demands of ever more bandwidth spurs innovation in transmission and transport systems for increasingly larger bitrate capabilities. (ii) As soon as such an advance happens, a straightforward scaling up of existing switching methods becomes unscalable and insupportable due to the increased port counts, energy expenditure, latency, or other metrics that would be incurred. (iii) Innovative switching methods are developed to switch traffic at larger bulk, or better efficiency, etc., *but* (iv) switching at the finest granularities continues to be necessary to interface the network with the lower-speed access and tributary networks, requiring the design of innovative approaches to selectively combine bulk and fine-grained switching—in other words, grooming.

A good fraction of recent grooming research has been focused toward grooming to address both the limitations and opportunities represented by the development of waveband transmission and switching systems, and multicore fiber, which fall in the first of the three directions we articulate above. In the next two sections, we review these areas, before concluding the chapter by indicating some other recent directions and horizons of grooming research, although we do not discuss those at similar length.

14.3 Wavelength-Waveband Grooming

As we have seen before, the basic need for grooming arises from the fact that there are multiple natural granularities of traffic and capacity in optical networks, which need to be mapped to each other. In classical grooming, as we have discussed above, the multiple granularities arise from the fact that transports that are appropriate for aggregating electronic traffic at higher layers have a very different (smaller) typical speed than primitives for optical communication (fibers, or even individual wavelength channels). However, multiple granularities can arise from different varieties of optical technology also. We have previously mentioned *colorless* equipment, which behave the same way regardless of the wavelength of the optical signal fed into them (within operating limits). To encode multiple signals into the different wavelengths that can be carried by an optical fiber, a different laser and other transmission equipment is needed for each such wavelength channel. The lasers can thus be considered *colored*, as in operating only on or with optical signals of specific center wavelengths. Similarly, to separate out the signals multiplexed onto the fiber by multiple such lasers, it is necessary to use optical filters and receivers, each of which select only specific wavelength channels. This is the process of wavelength division multiplexing (WDM), described in more detail elsewhere in this handbook.

The idea of wavebands arises from the observation that filters can be designed to select frequency ranges that do not necessarily have to match the precise ranges that were used by the transmission lasers. It is not useful to design filters that have finer granularity than the corresponding transmission lasers, since the resulting output signal will not correspond to any sensible signal injected by the transmitter. However, it is possible to use filters that have somewhat coarser granularity than the transmitting lasers. That is, such a filter would have a broader bandpass than the band of individual transmitting lasers. Such filters would have the effect of *partially demultiplexing* the wavelength channels multiplexed on the fiber. Consider the upper part of Fig. 14.6, which represents a possible spectral use of a fiber. Each peak corresponds to a separate optical signal injected by the transmitter, using 32 different lasers each tuned to a specific center frequency. To extract the signals in the wavelength channel, it would be necessary to employ 32 filters at the receiver that are similarly tuned. However, as the colors show, it is also possible to employ four filters of coarser granularity: that is, tuned to be less frequency selective. The first would filter through all of the wavelengths used by the first (lowest wavelength) eight lasers, the second would filter through the wavelengths for the next eight, and so on. Thus, the four filters partially demultiplex the incoming signal into four subsets of the actual wavelength channels used. Naturally, to extract the actual wavelength channels, it would be subsequently necessary to further demultiplex each such partially demultiplexed signal using finer filters; this process can be generalized into multiple such partial demultiplexing steps, ultimately using filters whose granularity matches the transmission lasers in order to extract the actual transmitted wavelength signals. Different combinations of coarse and fine-grained filters can obviously be used; the bottom half of the same figure shows how the same set of four broader bandpass filters partially demultiplex an optical signal that actually consists of 64 fine-grained wavelength channels.

Fig. 14.6 Wavelength and waveband

It may appear that this is not practically useful, since we are only adding coarse-grained filters, without being able to eliminate any of the fine-grained filters that are necessary to ultimately extract the wavelength channel signals. However, this can pose considerable savings for an optical switch that needs to switch such wavelength channels rather than terminate them, when it so happens (or is achieved by design) that many of the wavelength channels coming in on the same incoming fiber need to be switched onto the same outgoing fiber. In that case, the use of a single coarse filter may suffice to separate out the part of the optical signal that is composed of these channels and no others. The resulting optical signal can then be switched by a colorless optical cross-connect, and multiplexed back into the target outgoing fiber. This not only creates a saving in the number of filters, but also in cross-connect fabric. The collection, or band, of fine-grained wavelength channels that can be separated out by such a coarse filter is called a *waveband*, and this type of switching is called *waveband switching*. Figure 14.7 illustrates waveband switching; Fig. 14.7a shows how some wavebands may be directly switched from an incoming fiber onto an outgoing fiber, while other wavebands have to be terminated so that their component lightpaths may be extracted (for differential routing, or terminating). Figure 14.7b shows the architecture of a switch that can enable systematic waveband and wavelength switching;

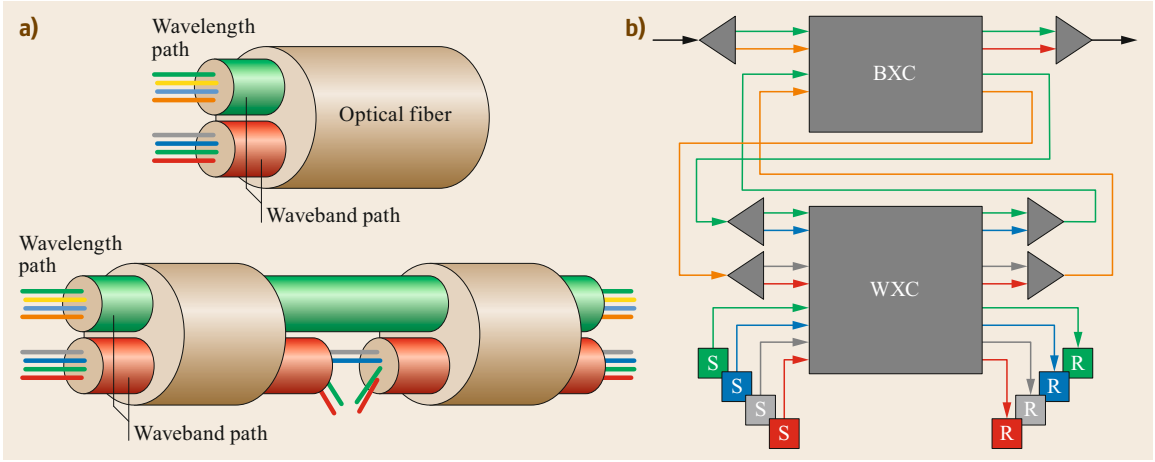


Fig. 14.7 (a) Wavelength, waveband, their lightpaths, and (b) optical node architecture

contrasting this with Fig. 14.5 is instructive, as it shows that the single level of optical cross-connect that we previously considered (composed of multiple identical optical switches for individual wavelengths) is now itself a hierarchy—into one or more (the figure shows only one) levels of waveband cross-connects (BXC), and a final level of wavelength cross-connect (WXC) as before. Together, the BXC and WXC make up a two-level OXC.

A waveband-based approach has become more attractive as optical technology has provided ever narrower wavelength channels. Currently, multiplexing 80 wavelength channels on commercial fiber with 50 GHz interval is very common. Even finer ones are already in use commercially, and likely to become increasingly common in the near future. Waveband switching provides an alternative to the increasing optical switch fabric size that would otherwise be needed, reducing the size of OXCs and the number of maintained lightpaths.

Some earlier optical network research, such as linear lightwave networks (LLNs) [14.14, 15], can be seen as a precursor of wavebands due to their use of a two-layer optical hierarchy. In LLNs, optical components called linear divider/combiners (LDCs) couple and split optical signals. Unlike passive optical networks (PONs), the coupling and splitting are on WDM basis (not TDM as in traditional PONs), and both splitting and combining can occur in a traffic flow. LDCs are collections of passive, colorless optical devices. In order to avoid coupling and splitting all wavelengths, waveband demultiplexers and multiplexers are introduced. Figure 14.8 shows an example with two wavebands, each with two wavelengths.

Waveband switching can be considered a generalization of this concept, with optical switches replacing LDCs. Such an approach, using space-divided optical switches instead of LDCs, and using fiber gratings for waveband demultiplexers, provided one of the early practical demonstrations of waveband-wavelength optical switching [14.16]. Later demonstrations have used arrayed waveguide gratings (AWG), bandpass filters with optical couplers, and wavelength selective switches (WSSs). Waveband conversion (simultaneous conversion of a set of wavelengths) has been demonstrated with field optical fibers in [14.17].

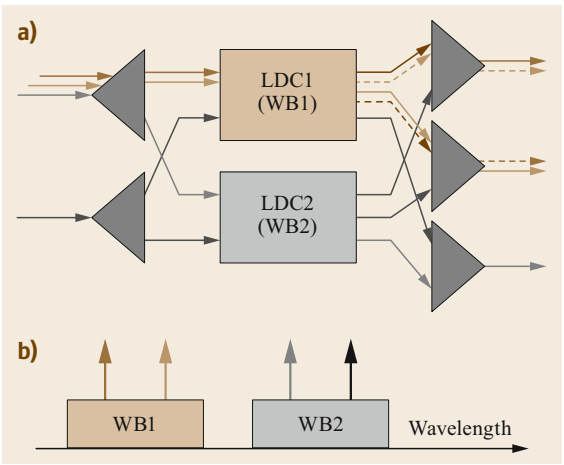


Fig. 14.8 (a) Waveband (de)multiplexing and linear divide/combine, an example with (b) two wavebands with two wavelengths each

The best assignment of wavelengths to wavebands, to attain some engineering or economic objective such as minimum cost of switching, is obviously a design problem similar to the classical grooming problem we considered before, and we refer to it as wavelength-waveband grooming. Research has focused on minimizing network cost, or maximizing carried traffic. Cost

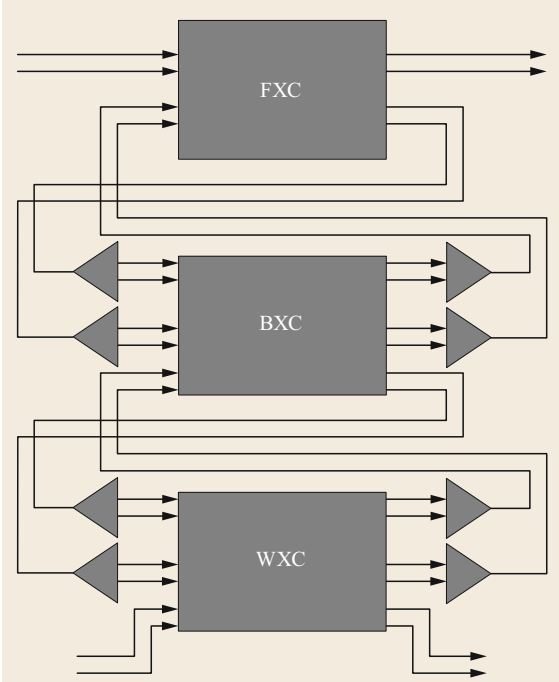


Fig. 14.9 FXC, BXC, and WXC in fiber, waveband, and wavelength layered architecture

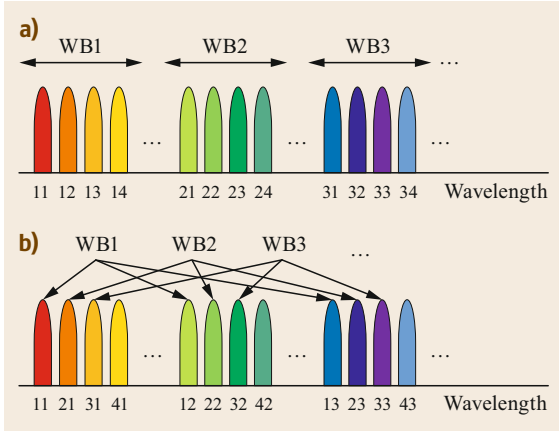


Fig. 14.10a,b Wavelength and waveband allocation: **(a)** continuous, **(b)** interleave/cyclic

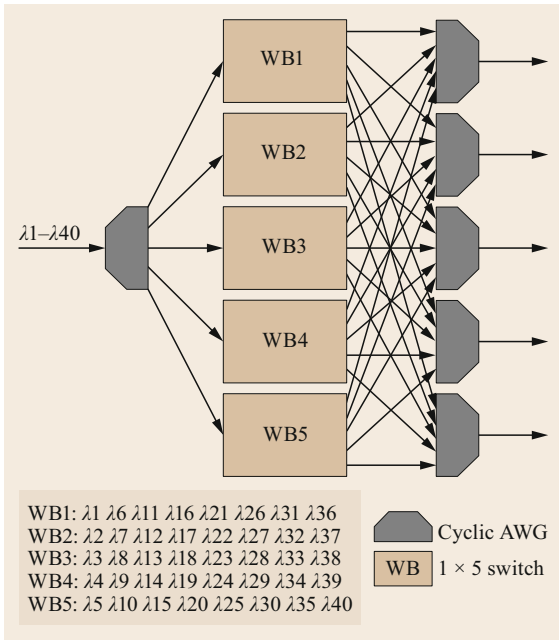


Fig. 14.11 A waveband–wavelength switch architecture with cyclic AWGs

can be defined narrowly, at the scope of each individual switch, by focusing on the number of ports and number of elemental switches. Alternatively, we can adopt a network-wide definition of cost, to include the number or distance of fiber links, the number of optical amplifiers, and so on. The study of [14.18] addresses port minimization by using wavelength conversion to put channels to be routed to the same destination on wavelengths that are suitable for grouping into wavebands. Conversely, [14.19, 20] does not assume wavelength conversion capability, and addresses the cost minimization problem in a three-layer hierarchy (where the fiber is considered a final level of aggregation, with multiple fiber links between node pairs) as in Fig. 14.9. Another interesting switch architecture is a hybrid of waveband switching and TDM (electronic) switching, where the wavelengths in a single waveband are demultiplexed into different digital lines, providing a wide range of granularity, from waveband through subwavelength [14.21].

In [14.22], an alternative waveband scheme and corresponding switch architecture is presented and demonstrated. In this scheme, the wavelength channels that make up any given waveband are not contiguous in frequency, but are interleaved, as shown in the right-hand side of Fig. 14.10. Such wavebands can be implemented by cyclic AWGs, which have a periodic transmission response, and the resulting wavebands have the advantage of highly isolated wavelength channels. Figure 14.11 shows an example of this architecture for five wavebands-per-fiber, eight wavelengths-per-waveband. At the first stage of the cyclic AWG, input wavelengths are divided into separate wavebands (e.g., WB1 = $\lambda 1\ \lambda 6\ \lambda 11\ \lambda 16\ \lambda 21\ \lambda 26\ \lambda 31\ \lambda 36$). At the waveband switches, each waveband can be switched to an output multiplexer (or dropped into a WXC if desired). At the set of output cyclic AWGs, all wavebands are multiplexed onto corresponding outgoing fibers.

14.4 Grooming for SDM Networks

Multicore fibers (MCFs) are a comparatively recent development at this writing. In such fibers, multiple paths exist for light to traverse, that can be used simultaneously by different optical signals; each such signal may be composed of multiple multiplexed wavelength channels [14.23]. In effect, a multicore fiber can be thought of as multiple fibers that are available at a single physical port, and with the same physical routing. Alternatively, we can see a multicore fiber as a fiber with a larger number of wavelength channels, with the added opportunity of wavelength reuse (that is, subsets of the channels can be on the same wavelengths), and can be switched optically to other cores of the same fiber at an intermediate optical switch. Figure 14.12 represents MCFs with seven cores each, three of which are shown as carrying lightpaths on the same set of five different wavelengths. Since the same wavelength can be reused multiple times on the same fiber by injecting different instances into different cores, this is called *space division multiplexing* (SDM), and is discussed at greater length elsewhere in this handbook. Such transmission systems and associated switching have been described in detail elsewhere, such as [14.24, 25]. With the development of multicore optical amplifiers [14.26], multiple wavelength paths in an optical core can be maintained and switched collectively.

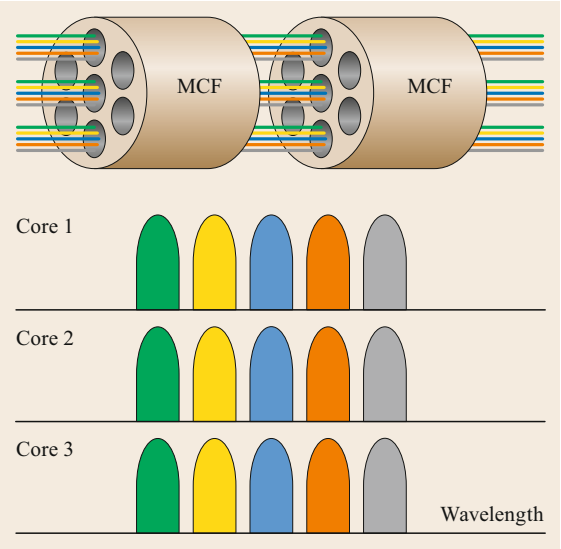


Fig. 14.12 Space division multiplexing and multicore fiber

As with wavebands, grooming in such networks is desirable in order to prevent the huge increase of bandwidth from causing a similar increase in required switching capacity. This can be appreciated by referring to Figures 14.13 and 14.14. The former shows a case where multiple cores are treated effectively as multiple fibers, leading to a very large number of wavelength channels that must be switched by the cross-connect, requiring a very large fabric. In the latter figure, we indicate a joint spatial-optical switch, capable of switching the signals from some cores as a whole to other outgoing cores on outgoing fibers (similar to switching entire wavebands without demultiplexing into component wavelength channels), while extracting individual wavelength channels from other cores. The structure of such a switch would be a further generalization of the switch structures shown in Figs. 14.5 and 14.7.

Another recent development in spectral use in fibers, called *elastic optical networking*, or flexible grid, allows different channels to occupy different amounts of the spectral range of the fiber's transmission capability. This allows provisioning lightpaths of different rates. Although this seems conceptually similar to inverse multiplexing lightpaths on a uniform grid, in fact using some lightpaths of variable capacity allows more flexibility, and being able to provision lightpaths of larger capacity (broader spectrum) allows elimination of wasteful guard bands. Different modulation methods may be chosen appropriately for channels of different bandwidth (or intended to traverse different distances), as shown in Fig. 14.15. The granularity of the lightpaths allowed is determined by the minimum width of the wavelength slot. ITU-T Recommendation G.694.1 specifies a frequency grid anchored to 193.1 THz for dense wavelength division multiplex-

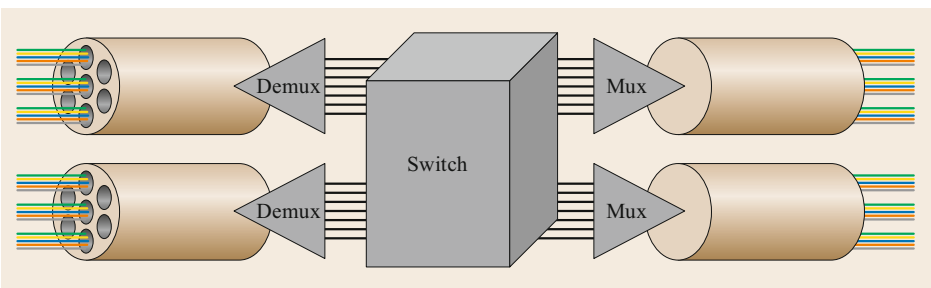


Fig. 14.13 A node in a no-grooming SDM network

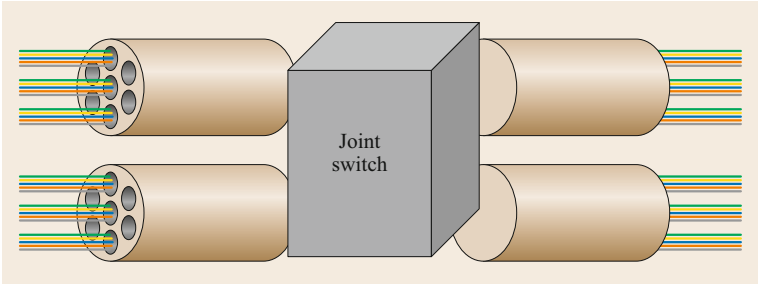


Fig. 14.14 A node in a grooming SDM network

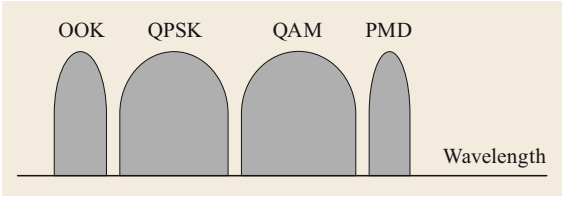


Fig. 14.15 Bit-rate and distance conscious modulation

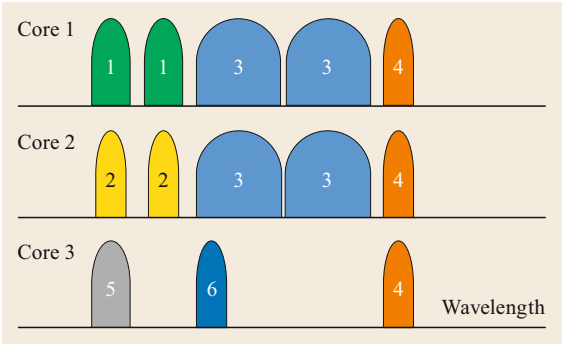


Fig. 14.16 Spatial and spectral super channels

ing (DWDM) applications, and supports a variety of channel spacing ranging from a minimum of 12.5 GHz to 100 GHz and wider. Elastic optical networking has been considered in more detail elsewhere in this handbook.

The idea of elasticity in spectral bands can be combined with multiple cores, as shown in Fig. 14.16. So-called *super channels* are formed by combining the corresponding spectrum in multiple bands and/or cores into the same logical channel. In this example, logical Channels 1 and 2 are each super channels in the spectrum domain, since each consists of two individual wavelength channels. Logical Channel 4 is a spatial super channel, while Channel 3 is a spectral-spatial super channel, in which individual component wavelength channels are themselves channels broader than the minimum bandwidth.

Grooming in such networks can encompass any part or all parts of the design decisions involved in mapping traffic to the available transport primitives, while keeping the switching needs (switch sizes, ports of various types, the total number of logical channel terminations, etc.) within scalable limits. Such problems can be formulated as integer linear programs, more complex versions of the formulations presented in Sect. 14.2. A formulation for the static case, encompassing spatial and spectral super channels, is available in [14.27]. A formulation for minimization of the number of switching modules (with a specific switching architecture) can be found in [14.28].

As before, such precise formulations are even more likely to be difficult to optimally solve in practice, and heuristic approaches using the insights provided by the formulations, and considering the specific network scenarios of the problem, are likely to be more realistic. For example, the study in [14.29] addresses the routing, spectrum, and core assignment (RSCA) problem. (The authors actually refer to *core/mode*, since their formulation can equally well handle multiple spatial cores or the modes of multimode fibers.) In this study, lightpath assignment also takes into account intercore crosstalk. One of their approaches, called predefined core prioritization, selects subsequent cores after the first one by considering adjacency, always selecting the core with least projected interference; they show that this results in better selection than first-fit in rotation order. Figure 14.17 shows an example of this approach (the right-hand side shows the solution that would be produced by first-fit).

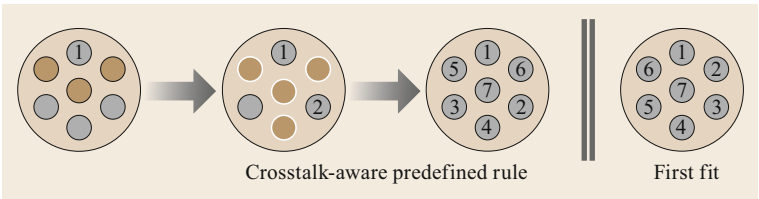


Fig. 14.17 Crosstalk-aware core selection

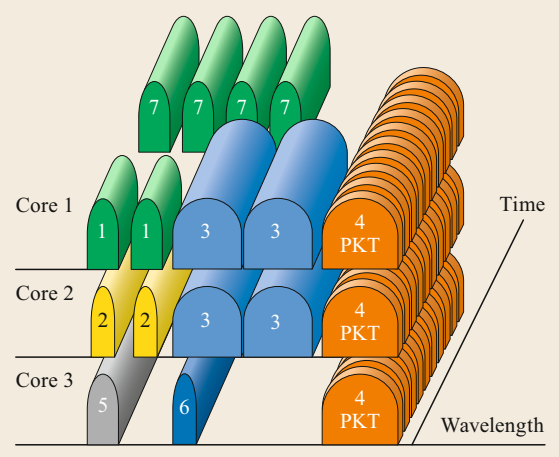


Fig. 14.18 Optical packet extension

Optical packet switching, a comparatively recent technological development that is currently becoming realistic (multiwavelength optical packet switching has been demonstrated at 12.8 Tb/s [14.30]), provides yet another flavor of grooming in SDM networks. Figure 14.18 illustrates grooming packets onto spectral-spatial super channels. Logical Channel 4 carries optical packets. Contrast this with Fig. 14.16; we see the added time dimension allows us to distinguish between long-lasting circuits, and short optical packets (or packet trains). Optical packet switching has been considered in more detail elsewhere in this handbook.

Figure 14.19 summarizes the design alternatives available to the architect of an SDM network. Any or all of these alternatives may be encompassed into the same overall network, and therefore can form the basis of a grooming problem.

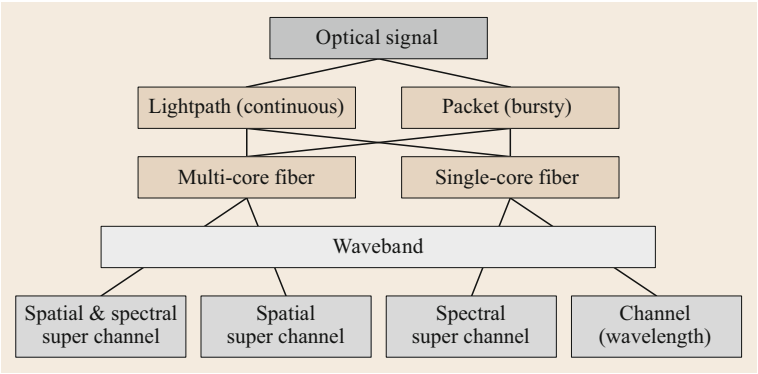


Fig. 14.19 A block diagram for path and packet composition

14.5 Conclusion: Traffic Grooming—Other Arenas

As we remarked before, traffic grooming studies have encompassed issues that have extended the scope of the field into areas that can be broadly considered in three categories. In the previous section, we focused on specific recently active research topics that fall in the first of those categories, namely keeping up grooming techniques with recent advances in optical networking advances.

In this last section, we briefly mention some of these research directions.

14.5.1 Grooming for Specific Traffic Requirements

One of the earliest ways in which grooming techniques were extended was in considering time-varying traffic. Although, as remarked before, the growth of Internet or packet-network traffic was one of the instrumental factors in bringing about the study of grooming, initial work in the grooming area had focused on the traditional paradigm that such traffic, when aggregated for injection into the backbone or transport network, varies with time comparatively slowly. Hence initial studies proceeded with a model of stationary traffic demand matrices, with the assumption that occasional recomputation with fresh traffic matrices would be sufficient to track the slow and gradual change in traffic. However, as the effect noted in Sect. 14.1.4 proceeded to previously unimagined levels, it became clear that such recomputation would be required quite frequently. This provided impetus for research studies that focus not only on near-optimality of a traffic engineering solution for a particular snapshot (matrix) of traffic, but its ease of reconfiguration into new solutions for likely evolution of traffic. We have described the main ideas of dynamic grooming in Sect. 14.2.4.

Beyond this, network application scenarios such as e-science, or high-quality (and more recently immersive) interactive video also motivates further expansion into an even more dynamic model, as the heaviest of subwavelength traffic components, while still subwavelength, become large enough that explicitly considering their arrival and departure models can present an opportunity to significantly improve the quality of the grooming solution. Finally, it is plausible that some traffic demands may be heavy enough in magnitude to merit entire wavelength channels, but retain the transiency of end-to-end traffic rather than the stability of aggregated traffic. Accommodating such demands in the same network would require dynamic resource provisioning algorithms that are a hybrid between the classical grooming algorithms of yesteryears, and the even older user-controlled lightpaths (UCLP) paradigm.

Other specific requirements of traffic that create novel aspects of the grooming problem have to do with finer awareness of factors affecting the quality of service or experience (QoS/QoE) received by a groomed traffic flow. Impairment-aware grooming attempts to take into account known or discovered fiber and transmission impairments in optical links when routing subwavelength traffic, and ensure that traffic demand components with particularly sensitive QoE requirements do not have them violated through being routed over the network of optical links. Similarly the reach of lightpaths may be taken into account.

Significant research attention has gone into *survivable grooming*, ensuring that a grooming solution guarantees given measures of continuity to groomed traffic flows, whether in terms of redundancy, failover time, reliability, or any other metric [14.31]. This is a particularly important current and future thread of traffic grooming, since an increasing fraction of traffic demands expect, or even depend on, service continuity through network disruption events. It is possible that in the near future, every traffic flow will do so. While network survivability has been well addressed in research, there are potential benefits that can be reaped only by jointly considering grooming and survivability aspects. As an example, a popular approach in network survivability is shared protection—designating the same network resources as protection for multiple resident lightpaths that cannot fail together according to the assumed failure model. It is obvious that greater benefit can be obtained by considering such traffic flows at the subwavelength level. More recently, such studies have also branched out into considering advancing optical technology; for example [14.32] considers the joint problem in the light of elastic optical networks.

Finally, some research has focused on additional innovative primitives of transport, beyond the traditional lightpaths and digital switching. Above we have already mentioned optical packet switching. Another novel idea worth mention is the use of the *light-tree* primitive: a one-to-many model of communication, using multicast at the optical signal level [14.33]. Such a light tree can be used for grooming in the same way as broadcast-and-select communication media of an earlier age: the source node designs a light tree to reach a number of destination nodes, and injects traffic bound for all of these nodes into the wavelength channel, and each destination node extracts the subwavelength part that is addressed to it. In such an approach, the source node chooses the combination of subwavelength traffic to groom together, for this reason, it can be considered *source-node grooming*.

14.5.2 Grooming Beyond Traffic Engineering

Finally, some grooming studies have focused on emergent issues that are not typically the purview traffic engineering. They may be considered either the application of traffic grooming techniques to problems other than that of traffic engineering, or grooming for traffic engineering goals as usual but with integrated consideration of other problems. Green grooming, or grooming to minimize energy usage of a network (rather than more traditional network figures of merit) is such an example. Grooming has also been used in network economy research, where grooming can be used as an economic tool by either the provider of the customer of networking services, and the value (OpEx, depreciation cost) of networking resources can be used to decide relative desirability of various network-wide grooming solutions—or to communicate, between provider and customer, such relative desirability for various grooming alternatives for specific traffic components. The newly emerged paradigm of software-defined networking (SDN) has multiple facets, and more than one has intersections with traffic grooming. SDN approaches can be seen purely as realizing a network administration/management channel, and thus a vehicle for coordinated grooming action, especially dynamic grooming. Agility and policy flexibility is another of the hallmarks of SDN approaches, and grooming solutions can be used to complement such agility by seamlessly switching between multiple alternatives for the same resident traffic demand utilizing different grooming solutions, involving different choices of OEO conversion, without any perceived changes for the end-to-end customer.

14.5.3 The Future of Traffic Grooming

Overall, traffic grooming is an important research area that is at the heart of bringing the benefits of optical communication and networking to the rich variety of

evolving traffic needs for new networking applications. By its very nature, grooming continues to be an exciting research area, as the needs and nature of traffic produces new surprises every year, and optical communication and switching techniques advance apace.

Even as new grooming research is published regularly, the core body of original traffic grooming literature continues to be a valuable resource—not just in historical terms, but in providing a wealth of historical techniques that can be called upon for considering new traffic engineering problem afresh as they arise, as well as other network design problems, through developments of technology that were unsuspected when the classical grooming problem was originally formulated.

References

14.1 L.R. Ford, D.R. Fulkerson: *Flows in Networks* (Princeton Univ. Press, Princeton 1962)

14.2 M.R. Garey, D.S. Johnson: *Computers and Intractability. A Guide to the Theory of NP-Completeness* (Freeman, New York 1979)

14.3 W.J. Goralski: *Sonet* (McGraw-Hill, New York 2000)

14.4 M. Pióro, D. Medhi: *Routing, Flow, and Capacity Design in Communication and Computer Networks* (Morgan Kaufman, San Francisco 2004)

14.5 B. Mukherjee, D. Banerjee, S. Ramamurthy, A. Mukherjee: Some principles for designing a wide-area WDM optical network, IEEE ACM Trans. Netw. **4**(5), 684–696 (1996)

14.6 R. Ramaswami, K.N. Sivarajan: Routing and wavelength assignment in all-optical networks, IEEE ACM Trans. Netw. **3**(5), 489–500 (1995)

14.7 E. Modiano: Traffic grooming in WDM networks, IEEE Commun. Mag. **39**(7), 124–129 (2001)

14.8 R. Dutta, G.N. Rouskas: Traffic grooming in WDM networks: Past and future, IEEE Network **16**(6), 46–56 (2002)

14.9 K. Zhu, B. Mukherjee: A review of traffic grooming in WDM optical networks: Architectures and challenges, Opt. Netw. Mag. **4**(2), 55–64 (2003)

14.10 K. Zhu, H. Zhu, B. Mukherjee: *Traffic Grooming in Optical WDM Mesh Networks* (Springer, Boston 2006)

14.11 R. Dutta, A.E. Kamal, G.N. Rouskas (Eds.): *Traffic Grooming for Optical Networks: Foundations, Techniques and Frontiers* (Springer, New York 2008)

14.12 D. Bertsimas, J.N. Tsitsiklis: *Introduction to Linear Optimization*, Athena Scientific Series in Optimization and Neural Computation, Vol. 6 (Athena Scientific, Belmont 1997) pp. 479–530

14.13 S. Huang, R. Dutta: Dynamic traffic grooming: the changing role of traffic grooming, IEEE Commun. Surv. Tutor. **9**(1), 32–50 (2007)

14.14 T.E. Stern: Linear lightwave networks: How far can they go? In: *IEEE Glob. Telecommun. Conf. Exhib.* (1990), https://doi.org/10.1109/GLOCOM.1990.116805

14.15 K. Bala, T.E. Stern, D. Simchi-Levi, K. Bala: Routing in a linear lightwave network, IEEE ACM Trans. Netw. **3**(4), 459–469 (1995)

14.16 H. Harada, K. Shimizu, T. Kudou, T. Ozeki: Hierarchical optical path cross-connect systems for large scale WDM networks. In: *Opt. Fiber Commun. Conf.* (1999), https://doi.org/10.1109/OFC.1999.766439

14.17 J. Yamawaku, E. Yamazaki, A. Takada, T. Morioka: Field trial of virtual-groupedwavelength-path switching with QPM-LN waveband converter and PLC matrix switch in JGN II test bed, Electron. Lett. **41**, 88–89 (2005)

14.18 M. Lee, J. Yu, Y. Kim, C. Kang, J. Park: Design of hierarchical crossconnect WDM networks employing a two-stage multiplex scheme of waveband and wavelength, IEEE J. Sel. Areas Commun. **20**(1), 166–171 (2002)

14.19 L. Noirie, M. Vigoureux, E. Dotaro: Impact of intermediate grouping on the dimensioning of multi-granularity optical networks. In: *Opt. Fiber Commun. Conf. Exhib.* (2001), https://doi.org/10.1109/OFC.2001.927315

14.20 X. Cao, V. Anand, Y. Xiong, C. Qiao: A study of waveband switching with multilayer multigranular optical cross-connects, IEEE J. Sel. Areas Commun. **21**(7), 1081–1095 (2003)

14.21 S. Yao, C. Ou, B. Mukherjee: Design of hybrid optical networks with waveband and electrical TDM switching. In: *IEEE Glob. Telecommun. Conf.* (2003), https://doi.org/10.1109/GLOCOM.2003.1258746

14.22 K. Sato, H. Hasegawa: Optical networking technologies that will create future bandwidth-abundant networks, IEEE/OSA J. Opt. Commun. Netw. **1**(2), A81–A93 (2009)

14.23 J. Sakaguchi, Y. Awaji, N. Wada, A. Kanno, T. Kawanishi, T. Hayashi, T. Taru, T. Kobayashi, M. Watanabe: 109-Tb/s (7x97x172-Gb/s SDM/WDM/PDM) QPSK transmission through 16.8-km homogeneous multi-core fiber. In: *Opt. Fiber Commun. Conf.* (2011), https://doi.org/10.1364/OFC.2011.PDPB6

14.24 D.J. Richardson, J.M. Fini, L.E. Nelson: Space-division multiplexing in optical fibres, Nat. Photon. **7**(5), 354 (2013)

14.25 D. Klonidis, F. Cugini, O. Gerstel, M. Jinno, V. Lopez, E. Palkopoulou, M. Sekiya, D. Siracusa, G. Thouénon, C. Betoule: Spectrally and spatially flexible optical network planning and operations, IEEE Commun. Mag. **53**(2), 69–78 (2015)

14.26 J. Sakaguchi, W. Klaus, B.J. Puttnam, J.M.D. Mendinueta, Y. Awaji, N. Wada, Y. Tsuchida, K. Maeda, M. Tadakuma, K. Imamura, R. Sugizaki, T. Kobayashi, Y. Tottori, M. Watanabe, R.V. Jensen: 19-core MCF transmission system using EDFA with shared core pumping coupled via free-space optics, Opt. Express **22**(1), 90–95 (2014)

14.27 P. Lechowicz, K. Walkowiak, M. Klinkowski: Selection of spectral-spatial channels in SDM flexgrid optical networks. In: *Int. Conf. Opt. Netw.*

Design Model. (2017), https://doi.org/10.23919/ONDM.2017.7958542
14.28 A. Muhammad, G. Zervas, G. Saridis, E.H. Salas, D. Simeonidou, R. Forchheimer: Flexible and synthetic SDM networks with multi-core-fibers implemented by programmable ROADMs. In: *Eur. Conf. Opt. Commun.* (2014), https://doi.org/10.1109/ECOC.2014.6963910
14.29 H. Tode, Y. Hirota: Routing, spectrum, and core and/or mode assignment on space-division multiplexing optical networks, IEEE/OSA J. Opt. Commun. Netw. **9**(1), A99–A113 (2017)
14.30 S. Shinada, J.M.D. Mendinueta, R.S. Luís, N. Wada: Operation of a 12.8 Tbit/s DWDM polarization division multiplexing 16-QAM optical packet switching node after 50-km of fiber transmission. In: *Eur. Conf. Opt. Commun.* (2014), https://doi.org/10.1109/ECOC.2014.6964190
14.31 C. Ou, K. Zhu, H. Zang, L.H. Sahasrabuddhe, B. Mukherjee: Traffic grooming for survivable WDM networks-shared protection, IEEE J. Sel. Areas Commun. **21**(9), 1367–1383 (2003)
14.32 M. Liu, M. Tornatore, B. Mukherjee: Survivable traffic grooming in elastic optical networks – Shared protection, J. Lghtwave Technol. **31**(6), 903–909 (2013)
14.33 L.H. Sahasrabuddhe, B. Mukherjee: Light-trees: Optical multicasting for improved performance in wavelength-routed networks, IEEE Commun. Mag. **37**(2), 67–73 (1999)

Rudra Dutta

Dept. of Computer Science
North Carolina State University
Raleigh, NC, USA
rdutta@ncsu.edu

Rudra Dutta has been employed on the faculty at Computer Science, NCSU since 2001 and has been Full Professor there since 2013. He is a Senior Member of IEEE and a Distinguished Member (Distinguished Engineer) of ACM. His current research interests focus on design and performance optimization of large networking systems, Internet architecture, optical networking, Internet of Things, and software-defined networking.

Hiroaki Harai

National Institute of Information and Communications Technology
Tokyo, Japan
harai@nict.go.jp

Hiroaki Harai received his PhD in Information and Computer Science from Osaka University, Japan in 1998. After this he led R&D on innovative network architecture technologies. In 2018, he was appointed as Director General at the National Institute of Information and Communications Technology, Tokyo, Japan, where he currently leads R&D testbed construction and stable operation of networks and the cloud.

15. Dynamic Control of Optical Networks

Mohit Chamania, Admela Jukan

Significant developments in technologies such as distributed/cloud computing, mobile applications and the Internet of things (IoT) are driving demands for lower costs and higher throughput from the underlying network infrastructure. Optical networks have evolved accordingly (flex-grid, all-optical switching, etc.) to accommodate increasing demand and reduce capital expenditure (CAPEX). In order to further lower costs, automation of control and management processes is essential to reduce human intervention over the service lifecycle. The need for automation has driven the development of optical control planes, which attempt to automate operations associated with provisioning and recovery from failures (service restoration). This chapter introduces the core functions involved in the control of optical networks, including network discovery, wavelength routing, dynamic service provisioning and seamless fault recovery. We chart the evolution from the inception of fully distributed control plane architectures to the more recent software-defined networking architectures and outline the already visible future trends in the dynamic control of optical networks, including intent-based networking and applications of artificial intelligence (AI).

This chapter is structured as follows. Section 15.1 introduces the background of optical control and management planes in legacy optical networks, and introduces control plane standards. Section 15.2 illustrates one of the first software-defined control and management plane artifacts, called path computation element (PCE). Section 15.3 analyses the applications of software-defined networking (SDN) in optical networks. Section 15.4 presents some emerging trends in optical control planes, such as intent-based networking and applications of AI. Section 15.5 gives a summary and a few final remarks.

Network control is a subset of network management operations associated with the provisioning and lifecycle management of services. In networks, the term *control plane* (often abbreviated as CP) is generally used to describe a set of (typically distributed) control processes that can automate functions associated with network control. The term *plane* specifically originates from voice networks, for it was often illustrated by a plane above the physical devices over which specific control functions are located to indicate its distributed nature and independence from the underlying physical network. Today, almost all telecommunication networks, including the Internet, cellular networks and content distribution networks, use the notion of a control plane

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_15

to refer to the infrastructure created to carry signaling traffic such as protocols for routing, resource reservation and failure detection, as well as the infrastructure that hosts the key networking functions, such as path computation—all of which are associated with the dynamic control of these networks.

Configuration of optical services in particular requires orchestration of numerous operations on devices, such as configuration of optical cross-connects in ROADMs (reconfigurable optical add-drop multiplexers), reconfiguration of gains from optical amplifiers, and power equalization of existing wavelengths that span the device to minimize induced noise. As operations on any individual device can have an effect on the end-to-end service, manual orchestration of these operations is known to be complex and error-prone. Furthermore, the mechanisms associated with these operations on individual devices are mostly vendor-specific. Vendor-specific environments make it harder to deploy services that span multiple vendor devices and introduce additional configuration and human-resource related challenges, as network administrators need to be trained to master control on different vendor systems, which is expensive and time consuming. Control planes have therefore been designed to abstract and automate multiple device-level operations in an effort to simplify distributed service provisioning, and potentially deliver both multivendor and multicarrier interoperability.

Control planes host critical network functions and operations so that the infrastructure remains resilient to failures. Traditional network management systems (NMS) were the primary source of error monitoring and correction, and were designed as a centralized entity responsible for delivering FCAPS (fault identification, configuration, accounting, performance, security) operations. However, a centralized network management system also presented a single point of failure. As a result, specific functions such as service monitoring and restoration were moved to the control plane. Control planes were designed as a distributed architecture, which allowed the control plane to automatically recover from failures on one or more sites. However, the distributed architecture is inherently more complex to implement, and also presents scalability challenges for managing networks with a large number of nodes, devices and/or services. These restrictions, and trade-offs between the centralized and distributed control and management architectures, coupled with recent advances in high-availability software architectures, have driven the conceptualization and design of innovative optical control planes, primarily in the context of SDN.

SDN proposes the separation of control functions, with node-specific control operations such as forwarding functions implemented on individual devices, while global functions such as topology discovery and path computation are implemented on a *logically centralized* SDN controller. In order to orchestrate network-wide operations such as service provisioning, the central controller can compute and push rules for individual devices using standardized protocols. Efforts to integrate SDN-based control in optical networks are still ongoing, but it has already emerged as a preferred standard for the joint management of multilayer (Internet protocol (IP)/optical) networks.

The development of SDN architectures and standards have now been integrated within a larger push to deliver concepts such as data analytics, machine learning, intent-based networking, and network function virtualization (NFV) that are essential for *intelligent automation* of optical networks. Control frameworks for optical networks are undoubtedly witnessing a rapid evolution alongside developments in AI and computing in general. All these trends in combination are soon expected to yield a true innovation in future optical networks.

15.1 Background and Evolution of Control Planes

Before all-optical networks became the technology of choice for transport of high-bandwidth services in the physical layer, end-to-end connections were multiplexed and switched at the electronic layer using SONET/SDH (synchronous optical network/synchronous digital hierarchy), which in turn operated over point-to-point WDM (wavelength division multiplexing) links. Switching at the electrical layer, while simpler, also required optical-electrical conversions at each node and was not cost efficient. The introduction of all-optical transport in telecommunication networks required accompanying control functions that could provide the basic feature set for managing services, namely service provisioning and automatic recovery from failures.

The common consensus to achieve these goals was to have a distributed control function that would be present on all devices participating in the optical network. Two standardization bodies, the International Telecommunication Union-Telecommunications (ITU-T) and Internet Engineering Task Force (IETF) came up with complimentary proposals for the same goals.

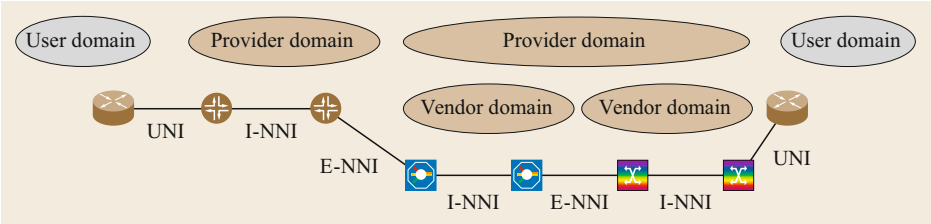


Fig. 15.1 Network representation in the ASTN/ASON standards

The ITU-T proposed the automatic switched transport network (ASTN) or automatic switched optical networks (ASON) that defined the requirements on the interfaces for communication between the different actors participating in the control plane. The generalized multiprotocol label switching (GMPLS) framework, proposed by the IETF, was a collection of protocols that could fulfill the functional requirements to setup and monitor end-to-end optical connections as outlined in ASON [15.1].

This chapter presents a brief overview of the ASON and GMPLS protocols.

15.1.1 Automatic Switched Optical Networks (ASON)

The ASTN/ASON standards proposed a reference network presentation as shown in Fig. 15.1. As shown in the figure, control domains indicate a distributed infrastructure to manage the underlying physical network infrastructure. Control domains are further classified based on the entity owning the underlying infrastructure. As seen in the figure, user domains are established for the infrastructure managed by the end-user, and a service can be requested from the transport network to interconnect entities in different user domains. A provider domain represents a *logical* control domain for the transport network infrastructure. As seen in the figure, a provider domain can either directly control the optical transport network infrastructure, or can be composed of multiple distinct control domains. The hierarchical control infrastructure is typical for large networks, where lower-level control domains can be organized to manage a subset of the network classified (for example) by the vendor or by the geographical area.

Using this reference network presentation, the ASTN/ASON standards identified three primary functions essential for controlling optical networks, namely:

- Topology discovery [15.2]
- Path control [15.3]
- Call and connection control [15.4].

In order to facilitate distributed network control, the control plane must have the capability to discover the actual network topology. The topology discovery process outlines a mechanism for devices to advertise link adjacencies in a heterogeneous multilayer, multivendor network in order to generate the complete network topology.

Path control functions focus on dissemination of network state information, and combined with the topology information can calculate routes for new service requests or can be used to calculate alternate routes for existing requests in case of failures in the network.

Finally, the call and connection control function is responsible for validating the authenticity and integrity of an incoming request, and for the setup of one or more optical connections to serve the request. The function is also responsible for communicating the status of a connection to the user, and provide detailed information in case of failures during connection provisioning or operation.

The ASTN/ASON standards address important issues faced by telecommunications (telecom) providers operating across multiple administrative domains and highlight the restrictions on each of these functions while operating across control domain boundaries. For example, topology discovery functions must restrict network visibility so that the user domain is not aware of the detailed provider topology, as this can be an inherent security risk. While some topology information must be exchanged between different provider control domains, the amount of information must be limited in order to improve the scalability of the control infrastructure. Topology information exchanged between different control domains must also hide specifics of the underlying topology, especially when exchanging information between domains controlled by competing providers.

In order to address these requirements, the ASTN/ASON standards identify three different types of interfaces between different control domains, namely:

- User-to-network interface (UNI)
- Internal network-to-network interface (I-NNI)
- External network-to-network interface (E-NNI).

The UNI is responsible for signaling operations between an end-user and a telecom provider domain. The

UNI interface allows an ASON user/client to initiate signaling to create, modify, and terminate a connection. The I-NNI defines the interface between adjacent optical nodes within the same control domain, which can share information with minimal restrictions to perform full topology discovery, routing and signaling (path setup) inside a control domain. Finally, the E-NNI demarcates administrative domain operations, either inside a single service provider or between different service providers. The E-NNI allows the exchange of network reachability information between different domains for routing, but limits the amount of information disseminated as compared to the I-NNI to preserve domain confidentiality.

As previously mentioned, the ASTN/ASON control plane does not specify a particular set of protocols, which means that any protocol that satisfies the requirements and operations specified in the ASTN/ASON architecture (defined in ITU-T G.8080 [15.5]) could be a candidate protocol. The IETF GMPLS protocol suite defines a set of protocols to address the requirements as defined in ASTN/ASON, and is discussed in the next section.

15.1.2 Generalized Multiprotocol Label Switching (GMPLS)

The popularity of the Internet and the rapid growth of capacity and network speed motivated the Internet community to start an IETF group in 1996 to focus on developing a new switching protocol standard. Traditional IP routers employ lookup tables (or a list of forwarding rules) that classify and forward packets based on the destination IP address and subnet. Subnet matching is an expensive operation, and destination address-based forwarding forces traffic to the same destination to follow the same path, inherently restricting the capability to distribute traffic across different potential paths in the network.

The IETF multiprotocol label switching (MPLS) group was established to standardize routing and signaling protocols that could operate over different underlying network infrastructures, with the primary objective of reducing the lookup complexity in order to support faster switching and facilitate efficient routing of traffic across the network, commonly known as traffic engineering (TE) [15.6, 7]. In contrast to the destination-based IP forwarding, MPLS was based on the so-called label switching. In an MPLS network, a data packet is *tagged* with a label at the ingress of the network, and subsequent forwarding of the packet is based on the label assigned to the packet. At each intermediate hop, a router inspects the packet and uses a label-based lookup table to forward the packet to the next hop. MPLS labels have a fixed size (20 bits) and use exact label matches, which makes it significantly faster than subnet-based lookups. MPLS also supports label swapping at intermediate hops, which means that labels can be reused. This functionality reduces the number of labels employed in the lookup tables, further reducing switching complexity and increasing the size of networks that can be controlled via MPLS.

MPLS employs the label distribution protocol (LDP) [15.8, 9], which is a distributed protocol to establish label-switching rules along multiple routers to create label-switched paths (LSP) for an application flow. The capability to dynamically establish a label-switched path on a per-application basis also made MPLS an excellent candidate to perform traffic engineering on the network.

With the emergence of all-optical networking, it was envisioned that label switching could be extended to also support switching across multiple transport network layers. The generalized MPLS (GMPLS) was an extension of the MPLS control plane architecture for different technology domains. This extension would not only enable optical networks to employ a more efficient control plane, but also would enable the creation of a unified control plane for Internet (packet) and optical (circuit-switched) networks [15.10]. GMPLS consisted of a suite of protocols extending the main MPLS signaling protocols to provide functions for topology discovery, service provisioning and automatic restoration in accordance with the ASON/ASTN specifications.

Topology discovery in GMPLS is performed using a combination of two protocols. The link-management protocol (LMP) runs between a pair of nodes or label-switched routers (LSRs) and is used to manage and maintain connectivity (or TE link information) between adjacent (peer) optical nodes [15.11]. LMP is used to maintain control-channel connectivity, verify the physical connectivity of data channels, and localize link failures for protection/restoration purposes in networks that include both optical and packet switching technologies. Information about LSRs and TE links is then distributed in the network using routing protocols (OSPF-TE (open shortest path first with traffic engineering) [15.12], ISIS-TE (intermediate system with traffic engineering) [15.13]) with extensions to to accommodate the circuit-switched characteristics of SDH/SONET and optical networks. Finally, protocols like the resource reservation with traffic engineering (RSVP-TE) [15.14] and constraint-based routing LDP (CR-LDP) [15.15] are used to make per-hop routing decisions to establish label-switched paths at different technology layers in order to create an end-to-end service.

GMPLS was developed to control both the Internet layer and optical layer in a unified fashion. As such, it is designed to setup an LSP (connection) across packet-switched (e.g., IP networks) and circuit-switched net-

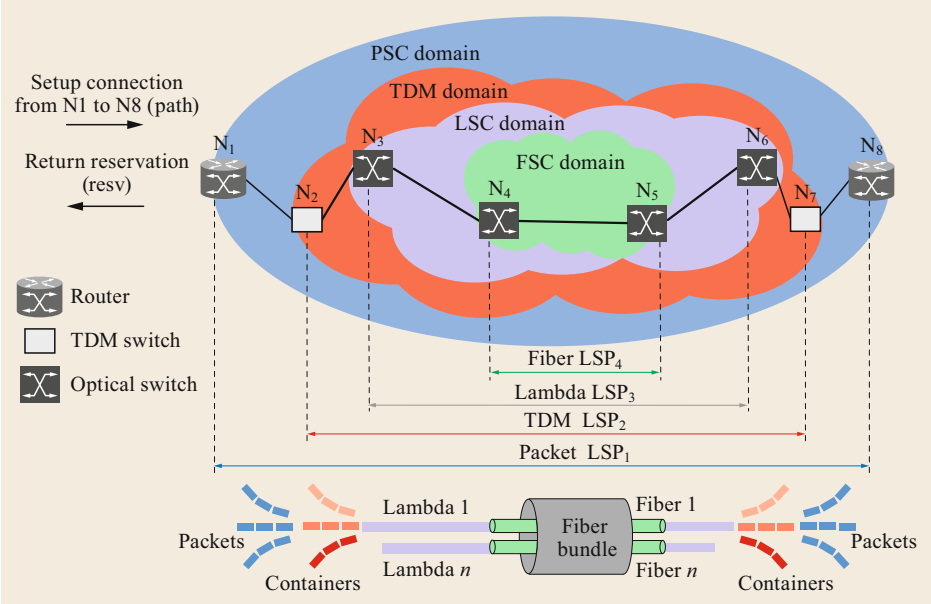


Fig. 15.2 Hierarchical LSP setup in a unified packet and circuit-switched control plane

works (e.g., SONET/SDH, wavelength-routed optical network). To this end, there are five interfaces standardized for a GMPLS-capable LSR: packet-switch capable (PSC), layer-2-switch capable (L2SC), time-division multiplex capable (TDM), lambda-switch capable (LSC) and fiber-switch capable (FSC). Packet-switch capable (PSC) interfaces have the capability to recognize packet boundaries and switch data based on (IP) packet header and MPLS labels. Layer-2-switch capable (L2SC) interfaces recognize frame (e.g., Ethernet) or cell (ATM (asynchronous transfer mode)) boundaries and can switch data based on the content of frame/cell header. Time-division multiplex capable (TDM) interfaces can switch data based on (SONET/SDH/OTN (optical transfer protocol)) assigned time slots. Lambda-switch capable (LSC) interfaces can switch data based on the received wavelength. The last one, fiber-switch capable (FSC) interfaces switch data based on their position in the physical fibers or ports. In each case, the unique resource used by a service was represented as a *label*, which was then used in the signaling processes, similar to MPLS.

Figure 15.2 illustrates various switching domains in a heterogeneous (packet and circuit-switched) network based on [15.16]. For routing, signaling and link management, the GMPLS control plane is installed in each node (routers, TDM switches and optical switches). Once the topology-related information is disseminated by TE capable IP routing protocols (OSPF-TE, ISIS-TE), a LSP can be established across various switching domains. Consider the setup of a connection (LSP) from node 1 to 8. It is important to note that only a single LSP can be established between the same interfaces, e.g., LSC. Thus, in a heterogeneous environment, the RSVP-TE signaling mechanism allows the provisioning of LSP hierarchies (PSC as the lowest level, and FSC as the highest level). In Fig. 15.2, node 1 sends a *Path* message to node 2 to set up *LSP1*, which triggers a setup of *LSP2* at node 2. Note that *LSP2* acts as a link for *LSP1* at node 2, which is an ingress node for higher-order LSPs. Similarly, higher-order LSPs are set up sequentially between node pairs (4–5), (3–6), which is only possible if the higher-order LSPs (here *LSP3* and *LSP4*) have sufficient bandwidth, as requested by the end node. When node 2 receives a *Resv* message (confirmation for an *LSP3* creation) from node 7, then a *Path* message is sent from node 2 to node 8 to establish *LSP1* between nodes 1 and 8. Finally, node 1 receives a *Resv* message from node 8, completing the provisioning of *LSP1*.

The GMPLS protocol suite can facilitate automated fault management in multilayer networks. In case of failures in the control or data plane (such as fiber cut) or due to software exceptions in a LSR, protocols in GMPLS can automatically recover from failures. For instance, loss of light can be used for fault detection, and fault localization can be achieved via monitoring loss of LMP hello messages. RSVP-TE can be used to communicate fault notifications, and protection and restoration mechanisms can be used to recover the connection in question. Finally, GMPLS protocols can maintain the original state of the service, and can revert the connection to the normal path from recovery LSP. These salient features of GMPLS constitute the foundation of the optical network control plane used in telecommunication networks today.

15.2 Path Computation Element (PCE)

Optimal path computation is a key component to enable dynamic provisioning of services in optical networks. As described in the previous section, distributed routing protocols in conjunction with reservation protocols like RSVP-TE can make per-hop routing decisions and perform path computation and reservation in one step. This mechanism however only allows for a single end-to-end path in the network. Applications requesting connections can have varied demands from the network and require specialized path computation. Furthermore, the setup of a new service can also affect the quality of service (QoS) of existing services sharing the same physical resources (links, wavelength cross-connects).

The IETF path computation element (PCE) [15.17] was envisioned as a centralized third-party management entity to perform constrained path computation. As shown in Fig. 15.3, entities or path computation clients (PCCs) can request a PCE server to perform constrained path computation using the PCEP [15.18] protocol. The *logical* separation of the path computation functions from the provisioning functions empowers operators to employ different path computation algorithms with no impact on the other control plane functions in the network. Constrained path computation, especially in optical networks, is computationally expensive and this logical separation also enables operators to move path computation logic away from the devices where computation resources are limited. Finally, the PCE enables network operators to have common algorithms governing path computation in multivendor networks, which is nontrivial in legacy systems.

The PCE architecture distinguishes between the PCE server and the traffic engineering database (TED). The PCE was envisioned as a complimentary control plane entity, with standardized interfaces for communication with other devices in the control plane. The TED is responsible for maintaining the global network topology and traffic engineering state, as is used by the PCE for path computation. The TED could monitor the traffic engineering information disseminated by control plane routing protocols. However, the PCE architecture can theoretically be used for path computation in any network, and the population and maintenance of the TED is intentionally abstract to support operations in networks operating without and underlying control plane.

15.2.1 Optimal Path Computation in Multidomain Networks

Under the ASON architecture, domain boundaries in networks represent technology, vendor or administrative separation, and topology visibility is restricted across domain boundaries. In order to perform optimal path computation in multidomain networks, PCEs in different domains share limited information with each other to compute paths spanning multiple domains. PCE-based routing architectures used in multidomain networks can be classified into two major groups: (a) peer-to-peer and (b) hierarchical.

In a peer-to-peer model, PCEs of neighboring domains create peering relationships and interact with each other to compute optimal multidomain paths.

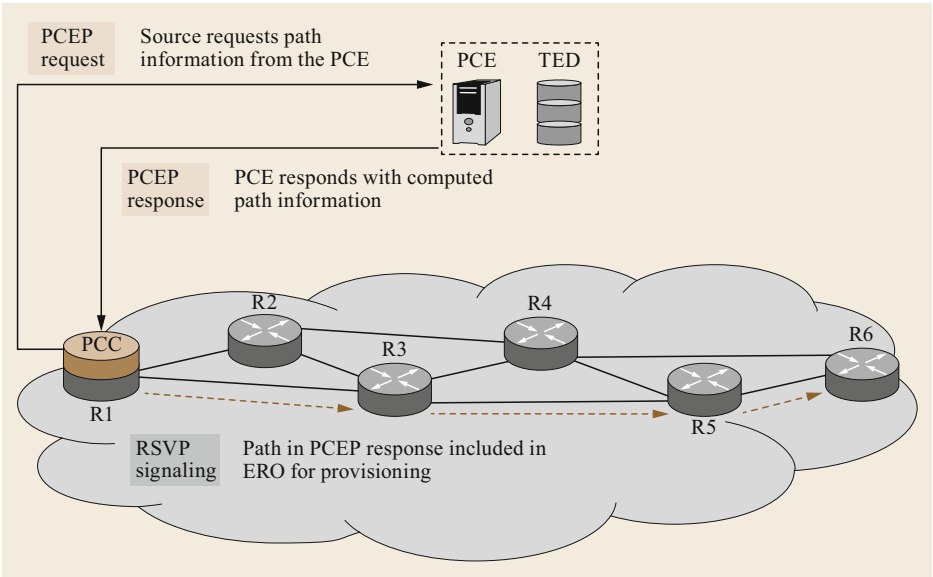


Fig. 15.3 Path computation element architecture

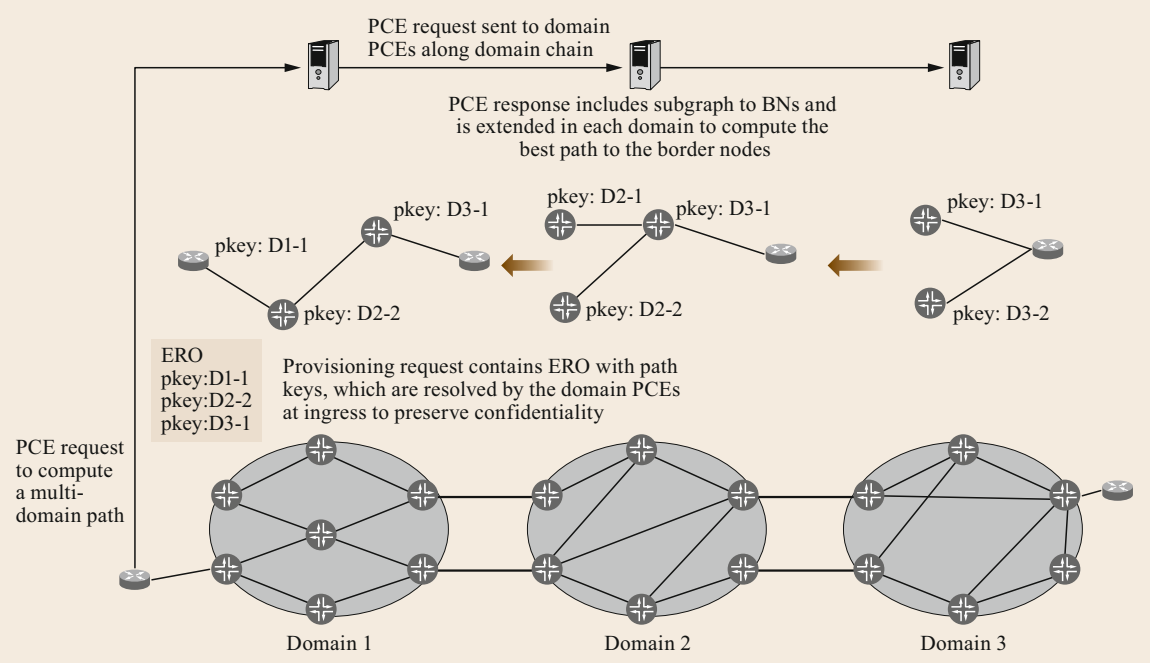


Fig. 15.4 Optimal multidomain path computation in a peering PCE model

A reference example is the backward recursive path computation (BRPC) [15.19], which, as shown in Fig. 15.4, can compute an optimal multidomain path along a prespecified sequence of domains between the source and destination. In this mechanism, a service request with endpoints in different domains is routed along a predetermined sequence of provider domains (or a domain chain). In order to compute the multidomain path, the PCEP protocol is used to contact domain PCEs in a sequential manner from the source to the destination, as determined by the domain chain. On reaching the destination domain, the destination domain PCE returns a set of path segments from the destination to all *border* nodes connecting the destination domain to the previous domain in the domain chain. Each domain extends these path segments within their domain boundaries to the border nodes connected to the previous domain in the domain chain, and sends it to the preceding domain PCE. On reaching the source domain, the source domain PCE uses the domain topology information and the path segments as computed by the domains along the domain chain to compute the optimal end-to-end path for the given domain sequence.

In the hierarchical PCE [15.20] architecture, domain PCEs for each domain interact with a centralized global PCE as shown in Fig. 15.5. The central PCE can be owned by the operator when operating across multiple control domains belonging to the same operator, or can be managed by a third party in the case of a multioperator system. Domain PCEs provide aggregate topology information to the central PCE, which uses this information to generate an abstract global topology. Interdomain path computation requests are forwarded to the central PCE, which computes a path on its abstract multidomain network view. Based on this path, the central PCE then communicates with each of the PCEs along the end-to-end path to compute the optimal path inside the domain boundaries to generate an optimal end-to-end path.

When operating in a multidomain environment, it is critical to preserve topology confidentiality. To this end, path-key-based mechanisms have been proposed in [15.21] to ensure that optimal path metrics can be transmitted to other domains without exposing the internals of the domain topology. In this mechanism, the path computed by a domain PCE for a multidomain path computation request is represented by a confidential path segment (CPS) and is identified by a *path key*. Messages exchanged between PCEs of different domains identify intradomain path segments using the domain border router pairs and the path key. Once path computation is completed, this path key is then included in the RSVP-TE explicit route object (ERO) during service provisioning. During provisioning, the domain ingress LSR extracts the path key associated with the domain, and queries the PCE to identify the associated CPS in order to route the provisioning request inside the domain, thereby supporting optimal multidomain path

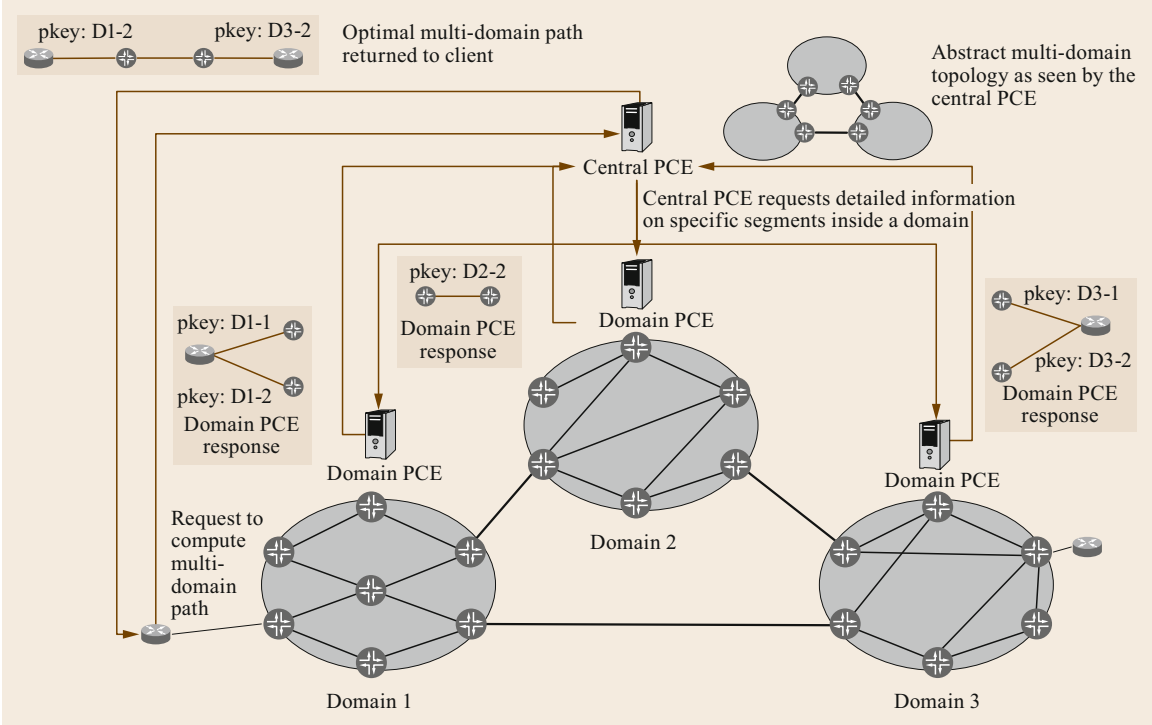


Fig. 15.5 Optimal multidomain path computation in a hierarchical PCE model

computation and provisioning, while ensuring topology confidentiality.

15.2.2 Multilayer PCE Architectures

The ability of the PCE architecture to perform path computation in multilayer networks is especially important in the current technological environment. Architectures proposed to perform multilayer PCE-based path computation [15.22, 23] differ in the scope of topology information available across the technological domain boundaries.

The *integrated multilayer PCE* solution, as the name suggests, has full multilayer topology visibility (a common TED) and uses this information to compute optimal paths. There are obvious advantages to having a complete view of the network for path computation, and the approach is quite popular with upcoming SDN frameworks. However, there are limitations to centralizing path computation, especially when operating with optical networks. The analog nature of optical transmission makes it very sensitive to the components used in a transport system, and optical vendors invest heavily in engineering their systems to achieve high optical transmission performance. As a result, multivendor optical networks still consist of vendor-specific domain islands, with interconnections between the domains typically established at the electrical (OTN/Ethernet) layer. As a result, there is little incentive for operators to implement a centralized PCE, and most commercial solutions to date defer path computation in the optical domain to vendor-specific PCE implementations.

In a coordinated PCE approach, each network layer is treated as a separate control domain with its own PCE. To serve a request, the client (IP/MPLS) network requests resources from the server (WDM) network. In its simplest form, a configuration like that assumes no direct communication between the PCEs. Instead, the IP/MPLS-PCE identifies a pair of entry points (ingress and egress routers) between the Internet and the optical network, and computes the path assuming that an optical circuit can be always be established between them. After selecting the abstract entry points, the client PCE communicates with the server PCE to compute the detailed path in the server layer to extract the complete multilayer path. This mechanism however has no information about the cost associated with establishing a circuit in the optical network, and can therefore be inefficient.

Exchange of information between the PCEs prior to path computation exposes additional information related to cost of establishing connections between a pair of endpoints, which enables the client IP/MPLS PCE to choose a pair of entry points to the optical network. To this end, extensions have been proposed to

GMPLS standards [15.24] and their SDN-centric implementation [15.25] to advertise *virtual* links to higher (client) network layers, which represent potential paths between a pair of LSRs in the lower (server) network layers. These links are used by the client PCE to compute an end-to-end path. In the case where a virtual link is used during service provisioning, the ingress LSR queries the server PCE to resolve the path associated with the virtual link, and initiates signaling to establish the associated circuit in the server network layer, before proceeding with the establishment of the LSP in the client network layer. Such mechanisms can significantly improve the quality of the computed path, but do not scale very efficiently to large networks due to the large number of virtual links that must be managed and exposed to the client network's PCEs.

15.3 Software-defined Networking (SDN)

The growing size, complexity and heterogeneity of networks requires novel approaches for network control and management, which has led to innovation in control plane architectures and protocols in all networks [15.26]. One of these recent innovations was SDN, a recent Internet control plane concept. The novelty of SDN as compared to the traditional Internet can be summed up with three major principles:

1. The separation of control plane and data plane (which are traditionally integrated in IP networks)
2. The logical centralization of control plane functions (which are traditionally distributed)
3. Programmability of network functions (which are traditionally nonprogrammable and static).

The three principles are related; for instance, the separation of control and data planes enables the network control to become programmable, and the network control is logically centralized, thus enabling a consistent view of that separation. A logically centralized control system also inherently improves programmability. In such a system, new functions are tested and deployed on a centralized location, which is inherently simpler than developing, testing and deploying distributed control functions, potentially across different vendor platforms, in a coordinated fashion.

Such a conceptual setting enables dynamic, automated SDN applications that are independent from the underlying hardware and can be written by any network operator based on their requirements. The latter is of significant business importance, since it allows any network and its resources to quickly adapt to changing business and market demands.

The Open Networking Foundation (ONF) consortium was one of the first forums to propose a reference architecture for SDN, and this architecture, with a number of variations, was widely adopted [15.27]. In a nutshell, the SDN architecture includes three planes: the data, control and application planes. As illustrated in Fig. 15.6, these planes are also connected to the management plane for configurations. As is commonly the

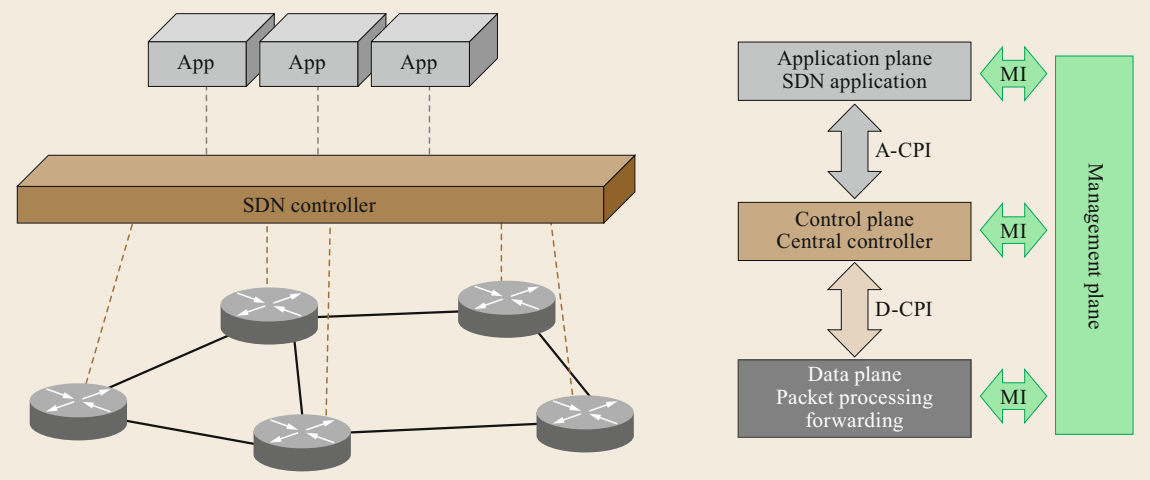


Fig. 15.6 Simplified SDN architecture proposed by ONF, with northbound interfaces (NBIs) to applications, southbound interfaces (SBIs) to devices and management interfaces (MIs) at different layers

case, the data plane processes and forwards the packets, based on the rules created by the control plane. The control plane uses the southbound data-controller plane interface (D-CPI) to interact and configure the elements in the data plane. The first protocol used for D-CPI was OpenFlow [15.28], though also other protocols, such as PCEP and transport application program interface (T-API) [15.29] can also be used. The control plane consists of the so-called *SDN controllers*. SDN controllers are software applications with the capability to store and maintain network-wide network topology and state information, and intelligence to define and configure rules to configure the data plane functions on network elements. The control plane also connects to the application plane, through the application-controller plane interface (A-CPI), also called northbound interfaces. The A-CPIs allow applications to be developed and tailored to the actual needs of the network systems, realizing advanced networking functions ranging from specialized routing to data analytics and special security mechanisms.

The salient feature of this architecture is its ability to separate the innovation, development and deployment in various planes and hide the details of vendor-specific hardware and software solutions. These features ensure that various applications can be implemented by software developers without a deep knowledge of network hardware. An example of such an application is the security application or the application to interconnect data centers. A security specialist, for instance, can write a security application on top of the control plane without deep understanding of the routing hardware, which can speed up innovation and development cycles. Similarly, a data center specialist can request extra capacity in a remote data center infrastructure without knowing the details of intercenter networking. Based on the application characteristics, the SDN controller translates the requirements and makes decisions that can be further translated into specific commands to the router. This capability to orchestrate (via software) network-wide operations in a logically centralized vendor-agnostic fashion is also commonly referred as *network programmability*.

As shown in Fig. 15.6, the management plane configures and manages the entire ecosystem of these three different planes through what is simply defined as the management interface (MI). There is a rather broad choice of protocols that can be used to implement the MI, and protocols like the Network Configuration (NETCONF) Protocol, the Open vSwitch Database (OVSDB) management protocol and the OpenFlow Configuration (OF-CONFIG) management protocol are some of the popular choices [15.26, 30, 31]. Management plane is a centralized component used for basic configurations and initialization of the system, such as configuring associations between SDN controllers and the data plane devices. For applications, the management plane can configure important parameters of user authorization or the so-called service-level agreements (SLAs), which are basically contractual obligations for the network to provide resources for applications, expressed in bandwidth, delay or other service-related parameters [15.32]. The architecture in Fig. 15.6 illustrates only one of the many variations of SDN architectures, and related efforts such as by the Software-Defined Networking Research Group (SDNRG) created by the Internet Engineering Task Force (IETF) [15.30, 33] define some variations of the functions and layers in the SDN architecture.

Given these developments in the Internet and considering that the optical control plane has been traditionally separated from the data plane, and as such had adopted the SDN philosophy much before the term was invented, the question is what the new paradigm of SDN today can do for the optical control plane. Fortunately, SDN can help optical networks solve a number of important problems, including:

i) Multilayer dynamic service provisioning
ii) Dynamic multilayer service restoration
iii) Dynamic traffic aggregation with optimizations applied jointly in multiple layers
iv) Network automation and intelligence, including enabling concepts of artificial intelligence [15.34–36].

Most of these features rely on the ability of the SDN-enabled Internet layer and the SDN-enabled optical layer to seamlessly interoperate by developing coordination functions in a centralized multilayer SDN controller [15.37, 38]. With SDN, services such as dynamic bandwidth-on-demand and restoration can now make use of capacity in multiple layers and are easy to implement and deploy in large and diverse production networks.

The introduction of SDN also paves the way for *service orchestration*, which in this context is a process of creating a service over multiple domains, layers and technologies. One way to orchestrate services to this end is to create a centralized SDN controller, and control all domains, technologies and layers in a centralized fashion [15.39]. Another way is to operate multiple controllers, and have them communicate either in a peer-to-peer or hierarchical fashion. At the heart of any solution is the orchestration process that is able to deal with a variety of technologies, network architectures and ownership models [15.40]. As of today, there is no consensus on the best solution, as all

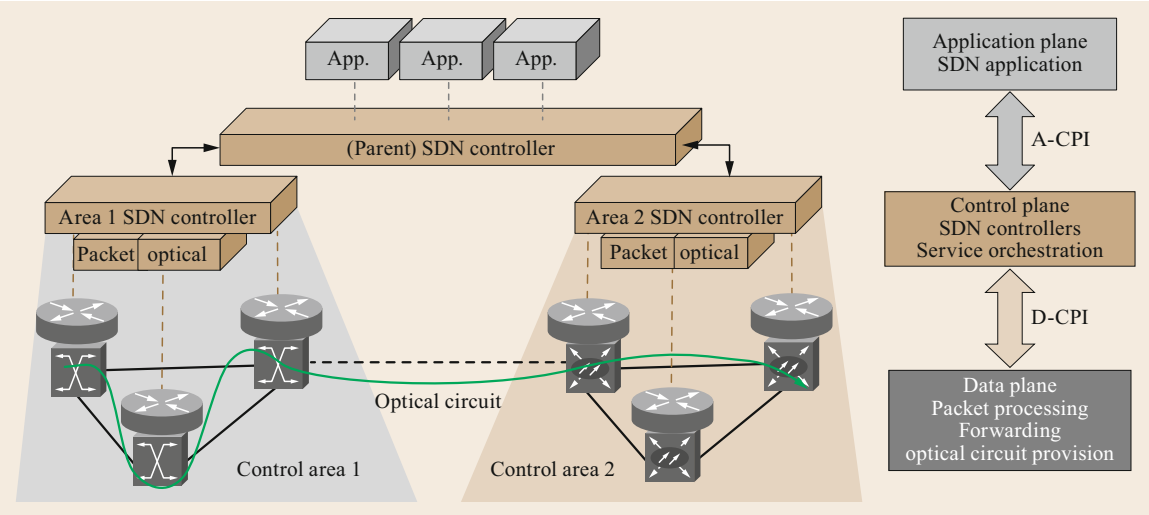


Fig. 15.7 Hierarchical SDN control plane example for multilayer and multidomain service orchestration

of them have significant advantages and disadvantages. As in the case with path computation in the context of the PCE, a centralized architecture is simpler to manage and can efficiently orchestrate operations inside the domain, while an architecture with multiple controllers provides administrative separation and is more scalable.

Figure 15.7 illustrates an example of a hierarchical SDN architecture over multiple control areas based on the previous concept of application, control and data planes. The optical network (data plane) is clustered into two control areas, whereby every area is a multilayer domain itself, as is practically always the case. The control plane is arranged hierarchically, but the SDN controllers are logically centralized. Individual area controllers (children, in the hierarchy parlance) are responsible for controlling the devices/resources within their corresponding segments and they also implement multilayer operation with SDN. A higher-level (parent) SDN controller includes the orchestration function, which not only gathers all the data relevant to service provisioning from all network domains, but it also actually optimizes the service path computation and guarantees its quality [15.41].

15.3.1 SDN Deployment Trends in Optical Transport Networks

The primary challenge in introducing SDN control for optical networks stems from the system complexity. Optical transport networks today involve control of photonic (fixed/flex-grid) media, multiple packet technologies (Ethernet, MPLS and MPLS-TP (multiprotocol label switching/transport profile)), as well as circuit-switching networking technologies (SDH/SONET, OTN). Packet technologies already support SDN control and the integration of SDN to manage OTN and its predecessor SONET/SDH appears feasible, as they involve electronic switching. However, implementing SDN in networks composed of fixed and flex-grid optical wavelength switching technologies is still a major challenge. Optical switching systems typically have constraints on switching capabilities, and have to take into account numerous physical layer impairments that can affect the end-to-end path. As a result, SDN-centric systems will need to come up with enhancements for devices to expose their switching restrictions and quantify *penalties*. This complexity is further increased when taking into account different switching granularities that can be seen in flex-grid optical network systems.

The next generation of SDN interfaces have different capabilities and control granularities [15.42], and are evolving around three distinct philosophies, namely *aggregated*, *partially disaggregated*, and *fully disaggregated* SDN control. Figure 15.8 outlines the scope of individual SDN control domains in each of these philosophies. Under the aggregated control philosophy, the complete optical network is presented as a single SDN domain, and only exposes UNI ports (client interfaces on transponders) and limited domain topology information. Provisioning requests to the SDN controller involve the creation of end-to-end circuits between two UNI ports, without specific control instructions for the other equipment in the optical domain, and the SDN controller is responsible for orchestrating operations across all nodes in the optical network. This control philosophy is similar to the existing control plane philosophy, and is preferred for its simplicity.

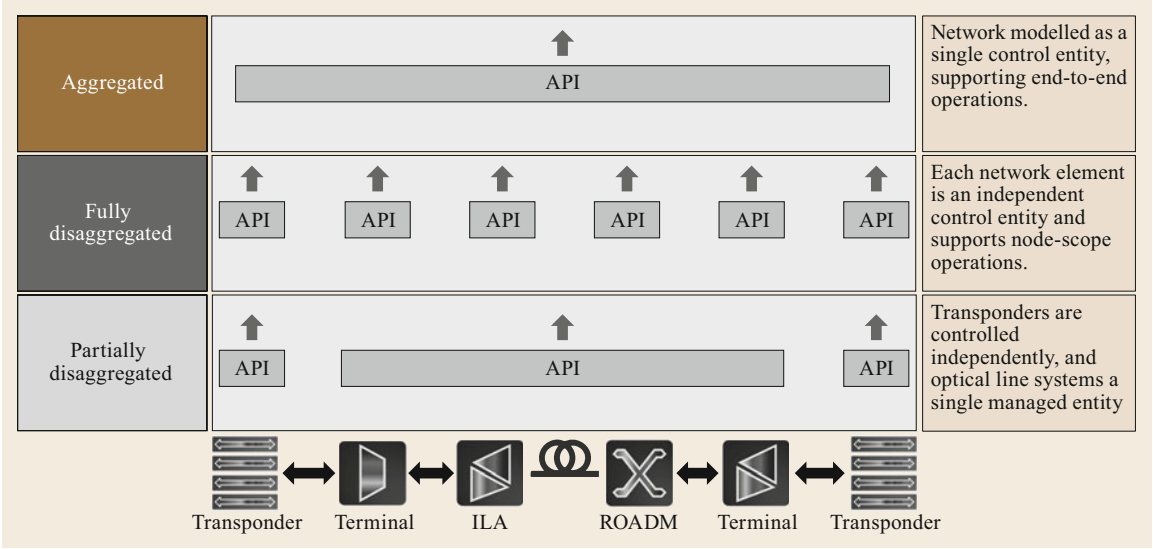


Fig. 15.8 Classification of next-generation optical SDN interfaces based on the control granularity

However, control domains in these paradigms are still vendor-specific, and operators do not have the option to mix and match vendor infrastructures in the same control domain.

In the fully disaggregated model, each individual optical component, such as transponders, optical reconfigurable add-drop multiplexers (ROADMs), and amplifiers are represented as a distinct network control entity. Each of these entities can be controlled individually via the SDN interface, and a central third-party orchestrator is responsible for orchestrating network-wide configuration operations. This approach has the clear advantage of reducing vendor lock-in, as any component can be replaced by a variant from a different vendor as long as the appropriate SDN interfaces are supported. However, such a system would require that measurement and qualification of penalties and performance of individual components be standardized across vendors. Additionally, the complexity associated with orchestrating network-wide operations such as provisioning is nontrivial and third-party systems need to be developed to deliver these functions in a multivendor environment. The associated complexity makes the wide adoption of fully disaggregated optical systems unlikely in the short term.

An achievable compromise, especially in the case of a network with brown-field deployments are the so-called partially disaggregated optical systems. Here, only the transponders are represented as individual control entities, while the rest of the optical network is exposed as a single control domain. The rationale behind this strategy is the significantly long life of the installed optical systems. Transponders, on the other hand, are upgraded frequently and have significant innovation and application customization associated with them. As a result, operators would like to employ transponders from different vendors while reusing the installed optical infrastructure, also called the open line system (OLS), in the context of disaggregated optical networks. In such deployments, all optical equipment would belong to the same vendor, who can use their engineering rules to estimate penalties effectively in the optical domain. Transponders, on the other hand, can provide feedback on the performance of the optical system, and in the case of coherent/flex-grid systems, can also be configured to operate according to the available optical performance in the network. The OLS provides an evolutionary path to disaggregated control of optical networks, especially in brown-field deployments, but suffers from operational issues such as topology discovery (transponder OLS interconnects), which need to be addressed, especially when operating in multivendor environments.

Irrespective of the approach adopted, implementing SDN in optical networks will require upgrades of the network equipment as well as the management software [15.39]. Network components designed for a more static environment will need to be replaced or enhanced with software that is SDN enabled [15.26, 27]. While the vertical (interlayer) interoperability can be addressed with the introduction of the SDN paradigm in the Internet, Ethernet and OTN, WDM architectures need to evolve further to support horizontal (intertechnology, interdomain) service provisioning. Optical SDN adoption not only requires seamless control across various technology or ownership

domains [15.43], but also the integration of a SDN-enabled network with non-SDN enabled (legacy) optical networks [15.40], which is still an open issue.

An evolutionary approach to introducing SDN control in optical networks involves the introduction of an *adaptation* or *middleware* subsystem to expose the network as a SDN capable domain while utilizing underlying control plane functions. For instance, an optical network can run a proprietary version of GMPLS and implement an abstraction layer that *translates* network state and configuration requests between the GMPLS control plane and the SDN controller. This approach provides an evolutionary path to prototype and evaluate SDN interfaces, and has been used extensively to prototype the first versions of SDN control interfaces. The use of the middleware approach is also important to bring legacy network infrastructure under SDN control, which is essential for deployments in brown-field networks. This approach however, can not be the de facto mechanism for future carrier-grade installations, as it is limited by the same scalability and timing challenges associated with the traditional distributed control planes. In order to scale SDN control for large optical transport network installations (5000+ nodes), vertical signaling or direct communication with the central controller is necessary to reduce overheads in computing and in maintaining the TED. Furthermore, the use of a control plane as the provisioning system also implies that SDN interfaces can only be used to make network-wide configurations, and cannot deliver fine-grained node-level control that may be expected by some SDN interfaces. As the carriers move to adopt SDN control, we can expect new network installations to be controlled by carrier-grade SDN controllers and legacy optical network domains to be integrated into the SDN control over time using the middleware approach.

15.3.2 PCE in the Context of Software-defined Networks

During the emergence of SDN control architectures, proposals were made to extend the PCE architecture to provide the capability to provision and maintain services over their lifetimes [15.44]. However, the stateful PCE specifications were significantly dependent on either the existence of a control plane in the network, or the presence of a entity capable of provisioning and maintaining connections in the network, which were replaced by SDN controllers. While the PCE protocol itself is not extensively used as is, the concepts developed in the PCE architecture have been incorporated into the current generation of SDN controllers, which have clean separation of functions for path computation and provisioning, and support the delegation of path computation to a third party or vendor-specific implementations.

15.4 Emerging Trends for Optical Network Operation

Control plane architectures, and recently SDN-based solutions, have been very successful in automating provisioning operations, driving down costs across all network technologies. As telecommunication operators converge on an IP-optical network architecture, numerous synergies in terms of common features (e.g., protection/restoration) as well as dependencies between the two network layers provide an opportunity for minimizing redundancies and optimizing operations across the multiple layers by introducing multilayer network control. However, coordination of operations across multiple network technologies is nontrivial. The fundamental separation between routing (IP) and transport networking has led to the organizational separation and fragmentation of technical competencies within an operator's ecosystem. Operations in carrier network ecosystems also involve numerous operations such as inventory management, reporting and alerting, in conjunction with the actual control and monitoring of the network infrastructure. These operations are tightly coupled to the tools and processes used in various organizations, and can differ significantly even within the same carrier ecosystem. The emergence of new applications with demanding requirements from the network, as well as architectures where network functions are virtualized and moved into cloud infrastructures, also imply that control operations are tightly coupled with applications/virtualized infrastructures associated with it.

Numerous extensions to control plane architectures have been proposed to address the issue of operating multilayer networks, which typically build upon the multilayer PCE [15.22] and application-based network operations (ABNO) [15.45] architectures. These architectures however limit their scope to control operations, and operators are looking towards solutions that have the capability of automating operations beyond traditional network control boundaries. The next generation of SDN architectures are designed with the inherent capability to *program* operations to suit the operator requirements, and this section discusses some of the notable trends in this area.

15.4.1 Software-defined Networks as Enablers of Network Function Virtualization (NFV)

The increasing cost pressure in managing networks has led to a *softwarization* and *cloudification* of network functions in general, especially in packet-based mobile networks. Network operators already deploy entities in clouds, such as the mobility management entity (MME), the IP multimedia subsystem (IMS) and the home subscriber server (HSS). This evolution will no doubt also happen in dynamic control of optical networks, since operations remain a significant portion of carrier network control expenses. To this end, SDN can be an enabler for network functions virtualization (NFV) and eventually be used to control network clouds. NFV is a remarkable trend in networking as it enables basic network functions to run on commodity hardware, and provides easy and extensible implementations of network functions through software-based solutions. NFV provides a replacement of the network middleboxes and dedicated network hardware, such as load balancers, and implements the same functions in the commodity hardware and clouds. Another important feature of NFV is the concept of a service function chain (SFC), a concatenated series of network functions to provide a dynamic network control service, such as combined security, fault tolerance and routing. In combination with programmable optics, the usage of optical technologies in an NFV setting will prove particularly promising in the future [15.46].

15.4.2 Workflow Orchestration and Intent-based Networking

Workflow orchestration is a standard technique that was primarily used to automate processes involving multiple third-party systems. In an orchestration flow, a central executable process is defined to automate a task, and interactions with other subsystems are based on message exchanges. Workflow orchestration, in the form of business process management (BPM), already has an extensive footprint and mature tooling for orchestrating operations that span multiple divisions, and involve integration between multiple software systems [15.47]. Workflow orchestration has also been extensively used for automating scientific workflows, operations inside data centers (virtual machine and container management), and is generally a key component of service-oriented software architectures.

Workflow orchestration involves automating interactions, or message flows, between different atomic operations in a complex process. As a result, it facilitates incremental introduction of automation in a complex process. Consider the process of provisioning an IP link. In a traditional network, the IP network and optical network would be managed by different entities, and the creation of an IP link would require multiple interactions between human operators managing the IP and optical networks. Operations in the manual provisioning process would include steps such as:

1. Identifying available IP ports and the corresponding transponders in the optical network
2. Planning of the underlying optical circuit (taking into account constraints on metrics such as bandwidth, delay and availability)
3. Identifying the IP interface configuration, including IP addresses and metrics for participation in routing protocols
4. Provisioning of the optical circuits in the optical network
5. Provisioning of the IP ports.

When introducing workflow orchestration, the integration between the various entities involved is automated as the first step, even while the individual atomic operations or *services* are manual operations. Manual operations are migrated to automated services over time, which enables seamless transition in the context of the workflow execution.

Workflow orchestration has gained significant traction in the context of network control with the emergence of SDN controllers and frameworks, which have simplified the integration of existing tooling for workflow orchestration with the systems that control the network infrastructure. Open-source frameworks such as the Open Network Operating System (ONOS) [15.38] and OpenDaylight [15.37] are designed in a fashion where all core features are treated as microservices and provide specific well-defined functions. Such architectures facilitate the definition and orchestration workflows, which can then be introduced by operators to perform multilayer operations. Research and open-source efforts [15.48–50] as well as commercial solutions are exploring orchestration as a means to achieve end-to-end automation in the network.

Intent-based Networking

The advances in network infrastructures have also driven the development of new network applications that expect specific capabilities from the network. Capabilities may include constraints on traditional network metrics such as delay, minimum bandwidth, jitter, high-level requirements in terms of availability and service downtimes, as well as specialized requirements such as in-flight encryption of data along the service path.

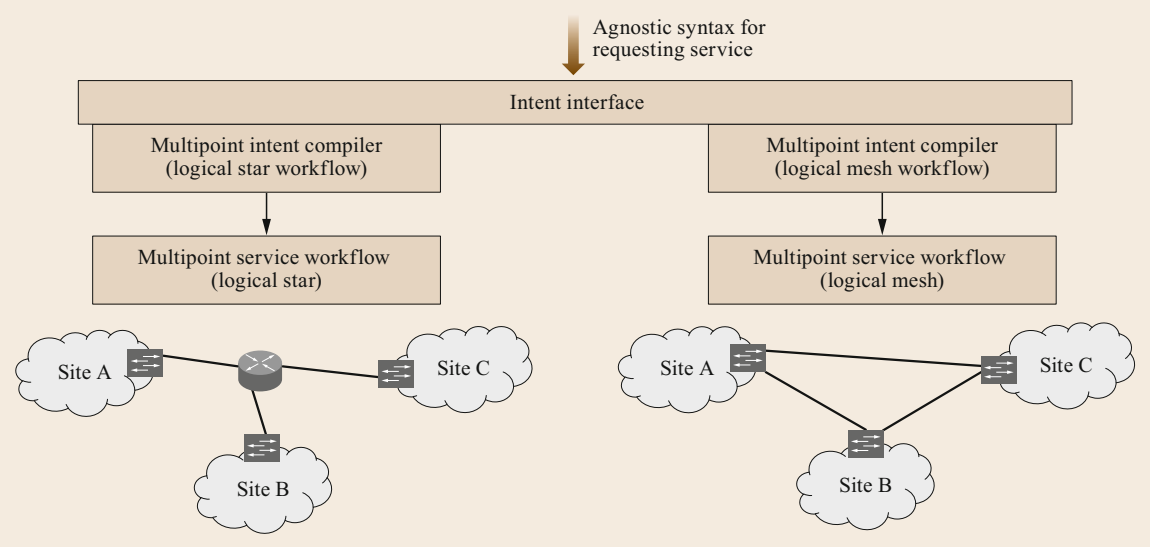


Fig. 15.9 Abstracting specific application workflows using intents

As described earlier, workflows define *how* to automate a desired sequence of operations to achieve a specific objective. Consequently, there can be multiple workflows that satisfy the constraints but optimize different network metrics for a complex service orchestration request. As a reference example, consider the request to interconnect multiple sites (e.g., a virtual private network), as shown in Fig. 15.9. The request can be potentially served by:

- The creation of a logical *star* topology, where each site is connected to a single node and traffic is always routed over this node
- The creation of a logical *mesh* topology, where each site is connected to each other directly.

Both solutions can potentially serve the requirements of the requesting application, but optimize different metrics, with the first solution optimizing on cost and used resources, with the second solution optimizing on throughput and redundancy. In a system where multiple workflows exist to serve a request, the user must identify the preferred workflow for their requirements and execute the operations accordingly, which can become cumbersome with the introduction of new workflows.

Intent-based interfaces have emerged as the preferred northbound interfaces for multidomain multilayer SDN controllers (also known as network operating systems) and provide a necessary abstraction between the application requirements and the underlying workflows that can serve these applications. Northbound intent interfaces provide clients (users, other applications) with a syntax to define *what* is desired from the network, which can be agnostic of the underlying technology or the specific mechanism/algorithm to fulfill a request. With the use of intents, the applications can treat the underlying network technology as a black box, and only be responsible for defining the application requirements.

As shown in Fig. 15.9, operators can expose abstract technology-agnostic syntax for the clients to request a service. Network operators can then implement capabilities (or intent compilers) to identify one or more workflows that can serve the abstract requirements expressed in the intent. In the case of multiple workflows, the network operators can seamlessly introduce policies to automatically choose one of the possible workflows, or provide *offers* (potential solutions) from each of the workflows to the user, who can select the one best fitting its requirements.

A reference example to demonstrate this capability is an application requesting a interconnection between two sites, with in-flight encryption support [15.51]. In a multilayer IP-over-optical network, encryption of data can be performed at the IP (IPSec (IP security)), L2 (MACSec (media access control security)) or directly at the optical layer. Provisioning encrypted connections at each layer involves a different workflow to orchestrate operations on the specific devices. However, from the applications perspective, the choice of layer at which the encryption is performed is not critical, but the encryption algorithm and key exchange mechanism used might be critical for compliance with regulatory authorities. As a result, the application would define the requirements on the encryption mechanism in their intent request, and the network operator would map these

requests to a specific workflow. This abstraction would also help the operator introduce encryption technologies and associated workflows in a seamless fashion from the application's perspective.

15.4.3 The Role of AI in Network Control

AI is a key emerging technology with the potential to significantly transform control processes in optical networks [15.52]. In the context of network control, AI techniques can be used to solve complex optimization problems in near real time. Problems such as the routing and wavelength assignment (RWA), routing and spectrum assignment (RSA), and multilayer network reoptimization have been investigated in detail, but the complexity of the associated optimization problem means that the problems cannot be solved in real time. AI-inspired heuristic techniques such as genetic algorithms, simulated annealing, and swarm intelligence-based algorithms have been used to solve these problems in theory. These results, in combination with the dynamic control now available with SDN-based frameworks, are inspiring solutions that can incorporate these resource optimizations as a part of the network control infrastructure. It is envisioned that such a feature could observe the network state and compute an optimal state as well as a migration path, which would then be used by the provisioning infrastructure in the network to migrate to the optimal network state.

In [15.52], it was shown that machine learning and other artificial intelligence methods have been used for various optical parameter motoring, resource (re)allocation, and network reconfiguration applications in response to a security attack. Since optical networks are vulnerable to different types of attacks, including jamming, eavesdropping, and physical infrastructure attacks [15.53], machine learning has been considered to address physical layer security in optical networks. While other attack prevention mechanisms, such as data encryption, quantum key distribution, and chaotic scrambling techniques [15.54, 55] remain important to apply, and are hard to break, attackers can still try to launch signal insertion and splitting attacks without being noticed. In such scenarios, machine learning (ML) has recently gained significant attention as a statistical method that is a viable solution to attack detection, addressing the issues of attack detection and prevention in real time and accurately [15.56].

Machine learning-based techniques can also be used classify, and potentially predict, failure conditions in the network. In the context of optical networks, this is an especially important capability due to the complexity of the associated infrastructure. Consider a typical scenario for a human operator that involves identifying the root cause of a failure along a service path. Optical network services traverse over numerous active/passive entities such as WSSs (wavelength selectable switches), amplifiers, AWGs (arrayed wave guides) etc., and any failure, including component failures, laser failures, bad interconnections or fiber cuts can result in multiple alarm indications from entities in the service path. Historically, experienced human operators typically analyze individual failure signatures to identify the root cause, but such analysis reduces the scale (and increases operational expenditure [OPEX]) associated with fault management in optical networks. Machine learning-based applications have already shown the capability to analyze and classify control information coming to the network control and management systems in order to improve the speed and scale of classification of failures in the network, and are now being investigated for application in commercial networks. AI is also being employed to monitor and classify problems with the physical layer in the network, which in turn can affect the control functions, through actions like rerouting optical services away from specific links, or changing the modulation format used for the service.

In conclusion, the combination of intent-based networking and AI-based techniques is a promising combination that can significantly increase automation in the network while maintaining the capability for operators to intervene or approve operations when required.

15.5 Summary and Final Remarks

The evolution of control frameworks, from the distributed control plane avatars to the centralized SDN architectures, has highlighted the need to automate operations for controlling optical networks. As networking technology evolves, increasing complexity and the dynamic nature of applications requesting resources will drive further automation of network operations. A key factor for introducing new technologies will be the capability to integrate with other existing and future subsystems, preferably in a technology-agnostic manner, and the ability to integrate humans-in-the-loop as/when required, to give the operator a chance to approve or veto actions in the network.

References

15.1 S. Tomic, A. Jukan: GMPLS-based exchange points: architecture and functionality. In: *Emerging Optical Network Technologies, Architecture, Protocols and Performance* (Springer, Berlin, Heidelberg 2004)

15.2 ITU-T Std. Rec. G.7714: Generalized automatic discovery for transport entities (2005)

15.3 ITU-T Std. Rec. G.7715: Architecture and requirements for routing in the automatically switched optical networks (2002)

15.4 ITU-T Std. Rec. G.7713: Distributed call and connection management (DCM) (2009)

15.5 ITU-T Std. Rec. G.8080: Architecture for the automatically switched optical network (2012)

15.6 E. Rosen, A. Viswanathan, R. Callon: Multiprotocol Label Switching Architecture, RFC 3031 (2001)

15.7 D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, J. McManus: Requirements for Traffic Engineering Over MPLS, RFC 2702 (1999)

15.8 L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas: LDP Specification, RFC 3036 (2001)

15.9 L. Andersson, I. Minei, B. Thomas: LDP Specification, RFC 5036 (2007)

15.10 E. Mannie: Generalized Multi-Protocol Label Switching (GMPLS) Architecture, RFC 3945 (2004)

15.11 J. Lang: Link Management Protocol (LMP), RFC 4204 (2005)

15.12 D. Katz, K. Kompella, D. Yeung: Traffic Engineering (TE) Extensions to OSPF Version 2, RFC 3630 (2003)

15.13 T. Li, H. Smit: IS-IS Extensions for Traffic Engineering, RFC 5305 (2008)

15.14 D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow: RSVP-TE: Extensions to RSVP for LSP Tunnels, RFC 3209 (2001)

15.15 B. Jamoussi, L. Andersson, R. Callon, R. Dantu, L. Wu, P. Doolan, T. Worster, N. Feldman, A. Fredette, M. Girish, E. Gray, J. Heinanen, T. Kilty, A. Malis: Constraint-Based LSP Setup Using LDP, RFC 3212 (2002)

15.16 A. Banerjee, L. Drake, L. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, Y. Rekhter: Generalized multiprotocol label switching: an overview of signaling enhancements and recovery techniques, IEEE Commun. Mag. **39**(7), 144–151 (2001), https://doi.org/10.1109/35.933450

15.17 A. Farrel, J.-P. Vasseur, J. Ash: A Path Computation Element (PCE)-Based Architecture, RFC 4655 (2006)

15.18 J. Vasseur, J. L. Roux: Path Computation Element (PCE) Communication Protocol (PCEP), RFC 5440 (2009)

15.19 J. Vasseur, R. Zhang, N. Bitar, J. L. Roux: A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths, RFC 5441 (2009)

15.20 D. King, A. Farrel: The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS, RFC 6805 (2012)

15.21 R. Bradford, J. Vasseur, A. Farrel: Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism, RFC 5520 (2009)

15.22 E. Oki, T. Takeda, J. L. Roux, A. Farrel: Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering, RFC 5623 (2009)

15.23 O. de Dios, V. Lopez, M. Cuaresma, F. Munoz, M. Chamania, A. Jukan: Coordinated computation and setup of multi-layer paths via inter-layer PCE communication: standards, interoperability and deployment, IEEE Commun. Mag. **51**(12), 144–154 (2013), https://doi.org/10.1109/MCOM.2013.6685770

15.24 D. Papadimitriou, M. Vigoureux, K. Shiomoto, D. Brungard, J. L. Roux: Generalized MPLS (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN), RFC 6001 (2010)

15.25 T. Saad, R. Gandhi, X. Liu, V. Beeram, H. Shah, I. Bryskin: A YANG Data Model for Traffic Engineering Tunnels and Interfaces, working draft draft-ietf-teas-yang-te-16 (2018)

15.26 W. Xia, Y. Wen, C.H. Foh, D. Niyato, H. Xie: A survey on software-defined networking, IEEE Commun. Surv. Tutor. **17**(1), 27–51 (2015), https://doi.org/10.1109/COMST.2014.2330903

15.27 M. Vissers, I. Busi, M. Betts, L. Ong, G. Zhang: *SDN Architecture for Transport Networks* (Open Networking Foundation, Palo Alto 2016), White Paper

15.28 OpenFlow: http://www.openflowswitch.org/

15.29 C. Qiaogang, E. Segev, E. Varma, G. Zhang, H. Ding, I. Busi, J. He, K. Sethuraman, L. Ong, N. Davis, R. Vialta, S. Bellotti, V. Lopez: *Functional Requirements for Transport API ONF TR-527* (Open Networking Foundation, Palo Alto, CA 2016), Version No.01

15.30 A. Mendiola, J. Astorga, E. Jacob, M. Higuero: A survey on the contributions of software-defined networking to traffic engineering, IEEE Commun. Surv. Tutor. **19**(2), 918–953 (2017), https://doi.org/10.1109/COMST.2016.2633579

15.31 D.B. Rawat, S.R. Reddy: Software defined networking architecture, security and energy efficiency: a survey, IEEE Commun. Surv. Tutor. **19**(1), 325–346 (2017), https://doi.org/10.1109/COMST.2016.2618874

15.32 Open Networking Foundation: *SDN Architecture 1.0 Overview* (Open Networking Foundation, Palo Alto 2014), White Paper

15.33 E. Haleplidis, K. Pentikousis, S. Denazis, J. H. Salim, D. Meyer, O. Koufopavlou: Software-Defined Networking (SDN): Layers and Architecture Terminology, RFC 7426 (2015)

15.34 ECI The Elastic Network: Optical Software Defined Networks. Bringing Optical Networks into the Modern Age, White Paper (2017)

15.35 A. Jukan, M. Chamania: Evolution towards smart optical networking: where artificial intelligence (AI) meets the world of photonics. In: *Adv. Photonics 2017 (IPR, NOMA, Sensors, Networks, SPPCom, PS)* (Optical Society of America, Washington, DC 2017), https://doi.org/10.1364/NETWORKS.2017.NeM2B.4, paper NeM2B.4

15.36 M. Chamania, A. Jukan: A comparative analysis of the effects of dynamic optical circuit provisioning

on IP routing, IEEE/ACM Trans. Netw. **22**(2), 429–442 (2014), https://doi.org/10.1109/TNET.2013.2251897
15.37 The Linux Foundation: OpenDaylight, https://www.opendaylight.org/ (2017)
15.38 Open Network Operating System (ONOS): http://onosproject.org/
15.39 R. Alvizu, G. Maier, N. Kukreja, A. Pattavina, R. Morro, A. Capello, C. Cavazzoni: Comprehensive survey on T-SDN: software-defined networking for transport networks, IEEE Commun. Surv. Tutor. **19**(4), 2232–2283 (2017), https://doi.org/10.1109/COMST.2017.2715220
15.40 A.S. Thyagaturu, A. Mercian, M.P. McGarry, M. Reisslein, W. Kellerer: Software defined optical networks (SDONs): a comprehensive survey, IEEE Commun. Surv. Tutor. **18**(4), 2738–2786 (2016), https://doi.org/10.1109/COMST.2016.2586999
15.41 A. Sadasivarao, D. Naik, C. Liou, S. Syed, A. Sharma: Demystifying SDN for optical transport networks: real-world deployments and insights. In: *2016 IEEE Glob. Commun. Conf. (GLOBECOM)* (2016) pp. 1–7, https://doi.org/10.1109/GLOCOM.2016.7841727
15.42 T. Szyrkowiec, A. Autenrieth, W. Kellerer: Optical network models and their application to software-defined network management, Int. J. Opt. (2017), https://doi.org/10.1155/2017/5150219
15.43 M. Chamania, A. Jukan: A survey of inter-domain peering and provisioning solutions for the next generation optical networks, IEEE Commun. Surv. Tutor. **11**(1), 33–51 (2009), https://doi.org/10.1109/SURV.2009.090104
15.44 E. Crabbe, I. Minei, J. Medved, R. Varga: Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE, RFC 8231 (2017)
15.45 D. King, A. Farrel: A PCE-Based Architecture for Application-Based Network Operations, RFC 7491 (2015)
15.46 A. Gumaste, A. Kushwaha, A. Jukan: On the benefits of programmable optics for post-failure VM migrations in data-centers. In: *Opt. Fiber Commun. Conf.* (Optical Society of America, Washington, DC 2018), https://doi.org/10.1364/OFC.2018.Th2A.33, p. Th2A.33
15.47 M. Weske: Business process management architectures. In: *Business Process Management: Concepts, Languages, Architectures*, ed. by M. Weske (Springer, Berlin, Heidelberg 2012) pp. 333–371
15.48 M. Chamania, E. Demirbilek, A. Jukan, X. Masip-Bruin, M. Yannuzzi: Using BPEL workflow processing for cross-layer orchestrations in IP-over-optical networks: a proof of concept. In: *Netw. Oper. Manag. Symp. (NOMS), 2012 IEEE* (2012) pp. 1191–1194, https://doi.org/10.1109/NOMS.2012.6212049
15.49 Linux Foundation: Harmonizing open source and standards in the telecom world, https://www.linuxfoundation.org/publications/2017/05/new-networking-harmonization/ (2017)
15.50 Energy Sciences Network: On-demand secure circuits and advance reservation system (OSCARS), https://www.es.net/engineering-services/oscars/ (2017)
15.51 T. Szyrkowiec, M. Santuari, M. Chamania, D. Siracusa, A. Autenrieth, V. Lopez: First demonstration of an automatic multilayer intent-based secure service creation by an open source SDN orchestrator. In: *Proc. ECOC 2016 42nd Eur. Conf. Opt. Commun.* (2016) pp. 1–3
15.52 J. Mata, I. de Miguel, R.J. Durán, N. Merayo, S.K. Singh, A. Jukan, M. Chamania: Artificial intelligence (AI) methods in optical networks: a comprehensive survey, Opt. Switch. Netw. **28**, 43–57 (2018), doi:https://doi.org/10.1016/j.osn.2017.12.006
15.53 M. Medard, S.R. Chinn, P. Saengudomlert: Attack detection in all-optical networks. In: *Opt. Fiber Commun. Conf.*, 1998 OSA Technical Digest Series, Vol. 2 (Optical Society of America, Washington, DC 1998), paper ThD4
15.54 P.R. Prucnal, M.P. Fok, Y. Deng, Z. Wang: Physical layer security in fiber-optic networks using optical signal processing. In: *Proc. Commun. Photonics Conf. (ACP)* (2009), https://doi.org/10.1117/12.852911
15.55 X. Hu, X. Yang, Z. Shen, H. He, W. Hu, C. Bai: Chaos-based partial transmit sequence technique for physical layer security in OFDM-PON, IEEE Photonics Technol. Lett. **27**(23), 2429–2432 (2015)
15.56 Y. Li, N. Hua, Y. Yu, Q. Luo, X. Zheng: Light source and trail recognition via optical spectrum feature analysis for optical network security, IEEE Commun. Lett. **22**(5), 982–985 (2018)

Mohit Chamania
Dept. of Research and Development
ADVA Optical Networking SE
Berlin, Germany
mchamania@advaoptical.com

Mohit Chamania received his PhD in Electrical Engineering from the Institute of Computer and Network Engineering, TU Braunschweig, in 2013. He is currently working at ADVA Optical Networking, focusing on the development of the next generation of network management systems. His areas of interest include automated control and optimization of multilayer transport networks.

Admela Jukan
ECE
TU Braunschweig
Braunschweig, Germany
ajukan@me.com

Admela Jukan is Chair Professor in Communication Networks at TU Braunschweig. Her current research interests include optical networking, metrology for THz communications, and high-performance networking with applications in cloud and edge computing. She received her BSc degree from the University of Zagreb, Croatia, her MSc degree from Politecnico di Milano, Italy, and her PhD degree from TU Wien, Austria.

16. Cross-Layer Design

Suresh Subramaniam, Koteswararao Kondepu, Andrea Marotta

The chapter is organized as follows. Sections 16.1 to 16.3 constitute the first part of the chapter, namely, physical network layer design. In Sect. 16.1, the physical impairments that affect long-haul optical fiber networks are described. Models for predicting and measuring the quality of transmission (QoT) of network connections are detailed. In Sect. 16.2 various routing and wavelength assignment (RWA) algorithms are presented, some of which only guarantee QoT requirements in the presence of PLIs, while others optimize network performance while taking into account the PLIs. The latter, called PLI-aware algorithms, can significantly improve the performance of the network, i.e., lower the blocking probability. Sample numerical results illustrating the efficacy of cross-layer methods are presented. Section 16.3 addresses the design of protection and restoration techniques for physically impaired optical networks. The survivability of these networks to link failures is greatly enhanced by including information about the PLIs directly within the protection or restoration algorithms. The second part of the chapter, focusing on application-network-layer design, consists of Sects. 16.4 to 16.6. An application-aware metro-access programmable architecture is presented in Sect. 16.4. Resource allocation and path protection based on software-defined networking (SDN) is presented in Sect. 16.5, and Sect. 16.6 presents application-aware converged wireless-access resource scheduling. The chapter is concluded in Sect. 16.7.

This chapter is divided into two parts – cross-layer design aspects considering the interactions between the physical/optical and network layers, and design aspects based on the interactions between the network and application layers. The first part of the chapter deals with physical network layer design aspects. As optical signals traverse large distances, they are distorted and attenuated by a variety of phenomena, collectively called physical layer impairments (PLIs). This may lead to unacceptable lightpath performance unless judicious cross-layer methods are employed. Such cross-layer methods design the network layer using the underlying knowledge of the physical network, as well as allocate network layer

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_16

resources utilizing the knowledge of the current physical layer state. The various PLIs that affect optical signals are first introduced and techniques for modeling them are presented. Following this, several widely used cross-layer design and resource allocation approaches are outlined. The second part of the chapter describes application-aware networking paradigms. The application-aware network can dynamically respond to meet application requirements, to enhance user experience and traffic demands. The various application-aware software-defined architectures such as metro-access architecture, resource allocation and path restoration, and wireless-access converged architecture are outlined.

Today's transport networks can span thousands of kilometers and have information capacities on the order of several Tbps per network link. Network nodes consist of optical routers and switches, while links are formed by fiber optic cables. Optical fibers are able to carry this vast amount of information with remarkable quality of transmission (QoT) and reliability compared with other technologies. Yet, at these throughput rates and distances, some degradation in the quality of the transmitted signals does occur. These so-called physical layer impairments (PLIs) limit the distance that the signals can travel before requiring electronic regeneration, at which point the signal quality can be reset.

PLIs limit the optical signal-to-noise ratio (OSNR), which in turn limits the QoT, as measured by the bit error rate (BER). The BER must be kept below a specified value, say, 10^{-9}, so that the data remains useful to the user. PLIs originate from various sources, both internal and external to the link. Long fiber links require either periodic or distributed optical amplifiers to boost the power level along the line; these amplifiers add undesired random noises to the signal that corrupt the information. Fiber channels introduce both linear and nonlinear distortions on the communication signal. Signals traveling together on the same fiber or across the same node can create crosstalk that further degrade the quality of transmission. These PLIs have been the subject of many years of research by numerous research groups. Many signal manipulation techniques have also been introduced in the last decade to minimize these effects, including signal processing approaches and novel optical devices, yet residual impairments remain that must be accounted for in networks where QoT must be guaranteed.

Network designs that take into account the PLIs within the implementation, deployment, utilization, or management algorithms and protocols are termed *cross-layer*, as they cross boundaries between the attributes of the physical layer and decisions made at the higher layers. In this chapter, we describe mechanisms for the cross-layer design of large-scale optical networks so that these can be utilized more efficiently and reliably while providing the users high-quality signals. Cross-layer approaches to network design became popular for wireless RF (radio frequency) wireless networks in the late 1990s. For optical networks, the issues are quite different due to the guided nature of transmission as well as the relatively static nonrandom nature of the impairments. The discussion contained in this chapter applies solely to optical fiber networks. Cross-layer effects in optical networks began to be examined in the late 1990s, but the incorporation of physical layer effects into network design and resource allocation only became popular in the mid 2000s. The topic has been widely researched since then.

16.1 Physical Layer Impairment Sources and Models

Optical networks transmit communication signals through one or more links of optical fiber that can affect the signal quality. This section describes the most important PLIs such as dispersion, nonlinearity, noise, and crosstalk. Models to quantify and measure the quality of transmission (QoT) that can be provided by optical networks operating over these degraded links are then presented. There are two distinct goals addressed here: to predict the performance of the system under various network conditions so that cross-layer network design and management decisions can be made, and to monitor and evaluate the system performance to confirm the efficiency and QoT of the network.

16.1.1 Physical Layer Impairments

The term *physical layer impairment* is used to describe any phenomenon that diminishes the QoT of a lightpath. From the point of view of the optical network, PLIs must be understood and measured to facilitate the cross-layer design, and not for the purpose of directly diminishing their effect. Here, we describe PLIs and provide models that can be used by network designers to account for these effects. Some impairments, such as amplified spontaneous emission noise, are entirely independent of the current network state, making them easy to model, while other impairments, especially

nonlinearity, are inextricably tied to the instantaneous channel usage. The most critical impairments are reviewed below; these are classified as linear impairments and nonlinear impairments.

Linear Impairments

Power Losses. An optical signal loses power as it traverses a lightpath from source to destination. The power loss consists of intrinsic and extrinsic losses [16.1]. Intrinsic losses are caused by attenuation, absorption, reflections, refractions, Rayleigh scattering, optical component insertion losses, etc. The power loss is simply modeled as follows: if P_{in} is the power launched at the input of a fiber of length L, the output power P_{out} is given by $P_{out} = P_{in}e^{-\alpha L}$, where α is the fiber attenuation coefficient. The loss introduced by the insertion of optical components, such as couplers, filters, multiplexers/demultiplexers, and switches, into the optical communications system is called insertion loss and is usually independent of wavelength. Conversely, extrinsic losses are caused by micro- and macrobending losses. Additional losses occur due to the combined effects of dispersion resulting from intersymbol interference (ISI), mode-partition noise, and laser chirp as discussed later.

Chromatic Dispersion (CD). Dispersion is the widening of optical pulses as they traverse a fiber. In chromatic dispersion, the broadening occurs due to the various spectral components propagating at different velocities, and causes signals in adjacent bit periods to overlap, resulting in ISI. CD is dependent on bit-rate, modulation format, and fiber type. Dispersion accumulates in a linear manner over the fiber, and is quantified by the dispersion parameter D_f which is expressed in ps/(nm km), i.e., the pulse broadens D_f picoseconds per nanometer of spectral width of the pulse for every kilometer of fiber traversed. CD may be compensated or canceled by using dispersion-compensating fiber (DCF) which has a negative dispersion parameter. For example, the dispersion on a fiber of length L_f and dispersion parameter D_f can be compensated by using a spool of DCF of length L_c and dispersion parameter D_c such that the dispersion at the end of the fiber is close to zero and satisfies $D_f L_f + D_c L_c = 0$. However, because it is not possible to perfectly match the dispersion slopes of CD and DCF over the entire WDM (wavelength division multiplexing) spectrum, some wavelengths may be overcompensated and some others may be undercompensated. Receivers may be able to tolerate a certain amount of residual CD depending on the type of modulation and bit rate used.

Polarization Mode Dispersion (PMD). PMD is the result of the two different polarizations of light traveling at different velocities because of asymmetries and other random imperfections in the fiber. PMD is characterized by a PMD fiber dispersion parameter, denoted D_{PMD} and measured in ps/$\sqrt{\text{km}}$. The pulse spread (also called differential group delay (DGD)) over a fiber of length L is given by $\Delta\tau = D_{PMD}\sqrt{L}$. Typical PMD parameter values range from 0.1 to 2 ps/$\sqrt{\text{km}}$ [16.2]. PMD is a serious issue at bit rates higher than 40 Gb/s, and is quite difficult to compensate because of its random and time-varying nature. PMD-induced problems can be mitigated by decreasing the optical transmission distance at the expense of deploying costly optical-electrical-optical (OEO) regenerators. Alternatively, dispersion-compensation modules (DCMs) that compensate for the accumulated PMD can be deployed at reconfigurable optical add/drop multiplexers (ROADMs), optical cross-connects (OXCs), or amplifier sites. As PMD effects are random and time-varying, an adaptive PMD compensator that uses feedback may be required.

Polarization-Dependent Loss (PDL). This is another impairment induced by polarization, and is particularly important when polarization multiplexing is used for transmission. PDL is the difference in the loss of power for the best- and worst-case polarization states. It is measured in decibels and is given by $PDL_{dB} = 10\log(P_{max}/P_{min})$, where P_{max} and P_{min} are the maximum and minimum power for different polarizations, respectively. PDL typically occurs in passive components such as couplers and splitters.

Amplified Spontaneous Emission (ASE) Noise. ASE noise originates from optical amplifiers. Optical amplifiers are used periodically over long-haul links to increase the power of the signal. To preserve the signal quality, they must be used every 60–100 km, depending on the modulation. ASE noise is well modeled as an additive white Gaussian noise with power spectral density $S_{ASE} = n_{sp}hf_c(G-1)$, where n_{sp} is the spontaneous emission factor, h is Planck's constant, and f_c is the carrier frequency of the signal being amplified. Assuming that all amplifiers are operated with the same gain G, the total ASE power for a chain of N amplifiers through a linear fiber is approximated by $P_{ASE} = 2NS_{ASE}$, where the factor of two takes into account the unpolarized nature of ASE noise [16.3]. Amplification reduces the signal-to-noise-ratio (SNR) of signals, and this is quantified by a metric called the noise figure (NF) which is the ratio of the input SNR to the output SNR, expressed in dB.

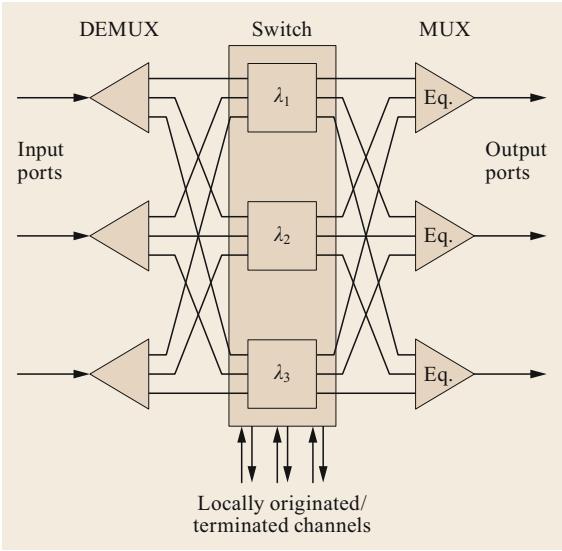


Fig. 16.1 OXC functional blocks

Crosstalk. Linear crosstalk, as opposed to nonlinear crosstalk described later, is caused by the imperfect isolation of optical signals as they propagate through various system components such as multiplexers/demultiplexers, ROADMs, and OXCs. As WDM signals at different wavelengths pass through these components, the signal is impacted by crosstalk wavelengths that are leaked from adjacent channels or ports. The crosstalk can be either in-band or coherent or out-of-band or incoherent, depending on the source of the crosstalk. Crosstalk can occur in many different ways and an exact calculation would depend on the particular system under consideration, but the types of crosstalk in optical networks have been broadly classified by several researchers. One such classification is presented below. Consider a typical OXC model, shown in Fig. 16.1, in which input WDM signals are demultiplexed and each wavelength is switched by a wavelength-plane space switch (such as a MEMS (micro-electro-mechanical systems) switch).

Three different types of crosstalk can then be identified [16.4]. In *cowavelength crosstalk*, shown in Fig. 16.2a, a signal on a particular wavelength, say λ_i, is affected by crosstalk from another lightpath on the same wavelength because of imperfect isolation in the switch. This is incoherent crosstalk, as the crosstalk comes from another lightpath. Conversely, the *self crosstalk*, shown in Fig. 16.2b, is coherent crosstalk because it is caused by interference from the same lightpath passing through a different path. The third type of crosstalk is called *neighbor-port crosstalk*. As shown in Fig. 16.2c, this crosstalk comes from another lightpath on the same wavelength.

Other Effects. There are two other main effects that are also important considerations in PLI-aware networking, namely, filter concatenation and amplifier tilt effects. The former refers to the narrowing of the signal's bandwidth as it passes through a series of filters from its source to its destination. The latter refers to the unequal gain that an erbium-doped fiber amplifier (EDFA) provides to the various channels of a WDM system. Techniques such as gain-flattening are available to mitigate the tilt effect.

Nonlinear Impairments

As the power of the optical signal increases, which is increasingly common now due to the large number of wavelengths that are multiplexed on a fiber and the increased signal power to increase reach, optical fiber starts exhibiting nonlinear effects. The nonlinearity can be classified as arising due to two different effects. In the first, known as the Kerr effect, the refractive index of the fiber depends on the optical fiber. This dependence causes three types of nonlinearities, namely, self-phase modulation, cross-phase modulation, and four-wave mixing. The second type of nonlinearity is due to scattering effects in the fiber; the two impairments of this type are stimulated Brillouin scattering and stimulated Raman scattering.

Self-Phase Modulation (SPM). The nonlinear phase modulation of an optical pulse caused by its own intensity is called SPM. SPM occurs due to temporal refractive index changes and manifests itself as chirp in the optical signal. Chirp refers to the change in the frequency of a pulse with time. Chirping leads to a change in the pulse width and, in most cases, causes dispersion. SPM is the most severe nonlinear impairment for WDM systems with relatively large wavelength spacings, e.g., 100 GHz.

Cross-Phase Modulation (CPM). The nonlinearity of refractive index is not just experienced by pulses on a single wavelength; all copropagating wavelengths experience the effects of the nonlinearity. In WDM systems, all copropagating wavelengths undergo phase modulation even if the individual wavelengths have moderate powers, as the aggregate power may be large enough to cause severe nonlinear effects. This effect is called cross-phase modulation. CPM is negligible in systems with large wavelength spacings, e.g., 100 GHz, but becomes significant when bit rates exceed 10 Gb/s and wavelength spacings become 50 or 25 GHz. CPM is also enhanced in dispersion-shifted fiber, which is a special fiber in which the zero of the linear dispersion is shifted to the 1550 nm wavelength from the 1300 nm wavelength, where it occurs naturally.

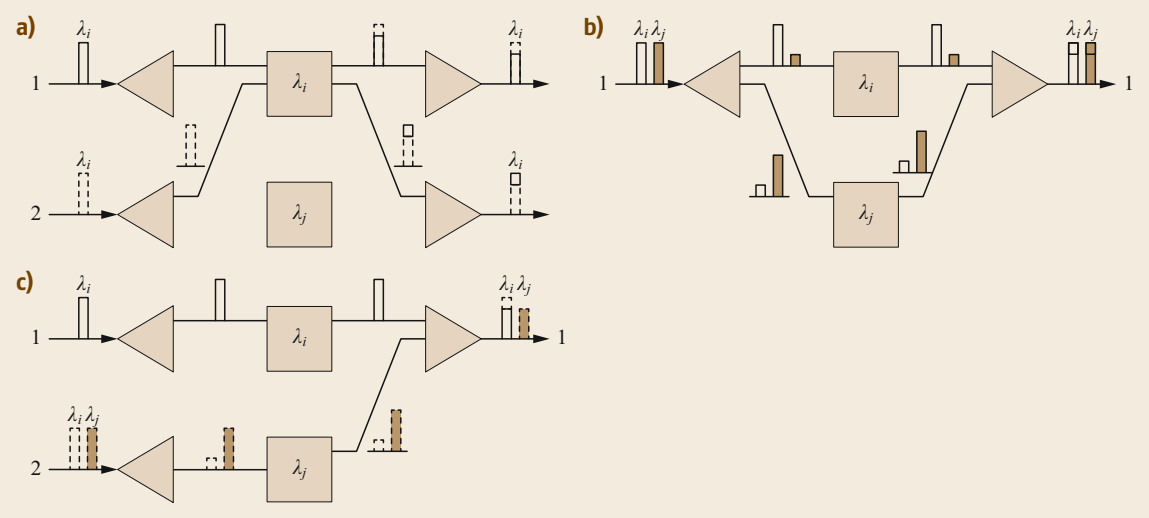


Fig. 16.2a–c Three types of in-band crosstalk (after [16.4])

Four-Wave Mixing (FWM). In a WDM system using carrier frequencies $f_1, f_2, \ldots, f_W$, the refractive index dependence on intensity not only induces phase shifts but also new frequency components. These components occur at frequencies $f_i \pm f_j \pm f_k$, i.e., any three frequencies combine to produce a fourth frequency, hence its name. The number of such frequency components is $W^2/2(W-1)$. FWM can have a severe impact on WDM systems if the generated frequencies coincide with or are close to the system frequencies. The effects of FWM are worse when dispersion is zero, unfortunately. Therefore, practical systems have small amounts of dispersion to mitigate FWM effects.

Stimulated Brillouin Scattering (SBS). SBS occurs when an optical signal interacts with acoustic phonons in the optical fiber. The resulting scatter effects cause a strong reflection of the transmitted signal, which has two implications. First, this places a limit on the optical power that can be launched from a transmitter, and second, any reflections must be blocked by an isolator placed in front of the transmitter.

Stimulated Raman Scattering (SRS). SRS causes the transfer of power from lower wavelength signals to higher wavelength signals in a WDM system. This can reduce the SNR of the lower wavelength signals and introduce crosstalk in the higher wavelength signals. SRS is a broadband effect compared to SBS. Moreover, SRS needs much higher power thresholds (of the order of watts) to occur than SBS. The effects of SRS can be reduced by keeping the powers low and by reducing wavelength spacings.

16.1.2 QoT Modeling and Measurement

Network providers guarantee a given QoT to their customers. This measure is typically given as the worst-case or average BER that the lightpath will experience over the duration of the connection. The designer of the network needs to have a model by which to predict the BER of the lightpath so that the best routing and wavelength assignment (RWA) decision can be made. Typical requirements are a BER $< 10^{-3}$ before error control coding, which would yield an error rate of $< 10^{-9}$ or better after error correction. The BER depends on all the impairments described above, and so depends on the actual state of the network: what connections are active, what data rates and modulations they use, etc. In this section, we provide various models by which to use this information to derive estimates of the BER. We then discuss how the lightpath performance can be directly measured.

The simplest model to incorporate PLIs into the QoT is through a limit on the length of transmission possible before the PLIs degrade the signals too much for the error control system to be able to recover. The so-called *transmission reach* is defined as the maximum length $L_{\max}$ of the lightpath such that the BER will reliably remain below a specified threshold under normal network operation. It takes into account the possible presence of linear and nonlinear effects by being conservative and considering the worst case. Many researchers have relied on a simple transmission reach constraint, e.g., [16.5].

A popular technique to estimate the performance of fiber-optic communications systems is to use a sim-

ple Gaussian noise approximation on both the effects of physical impairments and on the ASE noise. The BER of a system affected by Gaussian noise can be estimated given the first- and second-order statistics of the received decision variable. In fiber systems, these statistics are collected after the photodetection circuit, and depend on the type of modulation used. For on-off-keying (OOK) modulation, the decision variable is the sampled filtered output of the photodetector. For differential phase-shift keying (DPSK), a differential balanced detector is used, details of which can be found in [16.1]. For either case, the BER can be approximated using $\mathrm{BER} = 1/2\mathrm{erfc}(Q/2)$, where the metric Q is referred to as the Q-factor, and is defined as

$$Q = \frac{\mu_1 - \mu_0}{\sigma_0 + \sqrt{\sigma_\mathrm{i}^2 + \sigma_\mathrm{n}^2 + \sigma_\mathrm{nlx}^2 + \sigma_\mathrm{nx}^2}} \,. \tag{16.1}$$

Here, μ_1 and μ_0 are the mean photodetector output for a 1 bit and 0 bit, respectively; and σ_i^2, σ_n^2, σ_nlx^2 and σ_nx^2 are the variances due to intersymbol interference (ISI), amplifier noise, nonlinear crosstalk, and node crosstalk, respectively [16.6]; σ_0 is dominated by the ISI term because there is no power transmitted for a 0 bit. Dispersion effects are usually assumed to be canceled by dispersion compensators placed in the network. Higher order modulation techniques, such as differential quadrature phase-shift keying (DQPSK), have similar Q-factor formulations and are also primarily affected by the same four impairments. ISI, SPM, concatenated filtering, and amplifier noise are single-channel effects independent of the network state, and can therefore be predetermined given a network topology and physical layer parameters. Nonlinear crosstalk, occurring when signals copropagate in the fiber, is a network state-dependent impairment that cannot be precomputed. In this chapter, the variance due to nonlinearity is calculated as in [16.7]. Node crosstalk originates from signal leaks in the optical switches (in OXCs) and from the imperfect wavelength demultiplexing, and is therefore also network-state dependent. The formulation in [16.4, 8] is used to estimate the added variance due to node crosstalk.

More detailed analytical expressions for the decision sample variance and the probability distribution of the nonlinear phase noise caused by the interaction of ASE with fiber nonlinearity have been discussed in recent literature. In [16.9] the authors compare the effects of phase noise on DPSK and DQPSK modulation systems, while [16.10] models the optical fiber channel in the presence of nonlinear phase noise and presents BER results for a DPSK system. In [16.11], the authors calculate the variance of the nonlinear phase noise in an orthogonal-frequency-division-modulated (OFDM) system. The authors of [16.12] also derive the nonlinear phase noise variance, but for a PSK (phase shift keying) system, and analytical expressions for the linear and nonlinear phase variance due to SPM are derived. In [16.13], the authors propose and verify, with simulation and experiments, an empirical phase noise channel model for a long-haul optical system. The results agree with the data in the case of QPSK (quadrature phase shift keying) transmission. Optical networks transitioning to higher (100 Gb/s +) data rates often must operate using multiple line-rates, requiring special consideration; 10 Gb/s links spectrally adjacent to higher data rates cause the strongest degradation, as modeled in [16.14].

New and improved models for physical layer impairments continue to be developed. An alternative approach to modeling is to employ a Monte Carlo simulation, which can be used to effectively model all significant PLIs. The computational complexity is significant, and increases with the inverse of the expected BER. Estimating an error-probability below, say, 10^{-9}, requires too many trials to predict accurately.

16.2 Cross-Layer Routing and Wavelength Assignment

An important resource allocation problem in wavelength-routed optical networks is to find a route and wavelength(s) for a lightpath request. (The terms lightpath, call, and connection are usually used interchangeably in the literature.) In general, it is assumed that connections arrive one after another and their durations and arrival times are generally not known. This problem is the dynamic routing and wavelength assignment (RWA) problem (which is also discussed in detail in another chapter in this Springer Handbook, but without PLI considerations). In this section, we review a few cross-layer RWA algorithms and examine their performance.

16.2.1 Early Studies

One of the earliest studies on the effects of PLIs was published in [16.15]. The authors of [16.15] presented a detailed model for ASE noise and the calculation of BER. Using these models, the signal, crosstalk, and noise powers of a lightpath can be computed for an arbitrary wavelength on a route in a given network

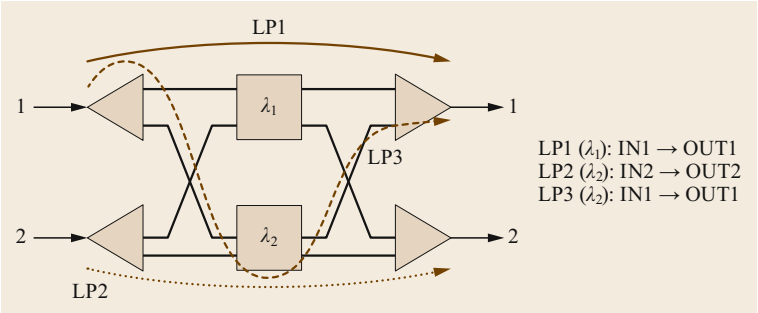


Fig. 16.3 Lightpath configuration in an example OXC

state. The paper then investigated the blocking performance of two common WA (wavelength assignment) algorithms in the literature, namely, random-pick (RP), which picks an available wavelength on the route randomly, and first-fit (FF), which picks the first available wavelength on the route. Shortest path routing was assumed in that work. An interesting observation that was made was that RP actually outperformed FF when PLIs were taken into account. This surprising observation was contrary to numerous prior studies that had shown the superior performance of FF over RP for networks with no impairments. The explanation for this counter-intuitive observation was that FF tends to assign lightpath wavelengths close to each other in the spectrum, causing larger crosstalk components, as compared to RP, which tends to spread wavelengths on a fiber, thereby mitigating the effects of crosstalk.

Another early paper that went a step further and incorporated the calculation of impairments due to crosstalk in the RWA process was [16.4]. In [16.4], four crosstalk-aware WA algorithm extensions of standard WA algorithms were presented. Besides the two standard algorithms RP and FF, two other algorithms were proposed: Most-used (MU), which picks the wavelength that is available on the route and is used on most other fibers in the network, and least-used (LU), which selects the wavelength least used in the rest of the network. Studies [16.16, 17] have shown that the general performance order of these algorithms is: MU > FF > RP > LU. However, as these algorithms ignore PLIs, the actual BER performance of the lightpaths may be unacceptably poor, leading to what is called BER *blocking*, distinct from the blocking due to wavelength unavailability, which is called *wavelength blocking*.

The algorithms in [16.4] employ a simple crosstalk-counting algorithm prior to wavelength selection. Here is how it works.

Consider Fig. 16.3. Let $n_{i,(x)}^{\mathrm{OXC}}$ denote the number of type (x) crosstalk components imposed on LPi within an OXC. Recall that type (a), (b), and (c) represents *cowavelength*, *self-*, and *neighbor-port crosstalk*, respectively, as described in Sect. 16.1. For example, in the LP (lightpath) configuration shown in Fig. 16.3, for *cowavelength*, since LP1 does not share the switching module of λ_1 with any other LP, while LP2 and LP3 use the same switching module of λ_2, we have $n_{1,(\mathrm{a})}^{\mathrm{OXC}} = 0$, and $n_{2,(\mathrm{a})}^{\mathrm{OXC}} = n_{3,(\mathrm{a})}^{\mathrm{OXC}} = 1$. For *self-crosstalk*, which is generated when multiple LPs traverse the OXC from the same input port to the same output port, it only occurs for the pair of LP1 and LP3 in this example. Hence, $n_{1,(\mathrm{b})}^{\mathrm{OXC}} = n_{3,(\mathrm{b})}^{\mathrm{OXC}} = 1$, and $n_{2,(\mathrm{b})}^{\mathrm{OXC}} = 0$. *Neighbor-port crosstalk* is applied when the signal LP enters the OXC with an LP on a wavelength and exits the OXC with another LP on the same wavelength. Obviously, such a scenario does not exist in the example. Hence, $n_{1,(\mathrm{c})}^{\mathrm{OXC}} = n_{2,(\mathrm{c})}^{\mathrm{OXC}} = n_{3,(\mathrm{c})}^{\mathrm{OXC}} = 0$. As the intensities of the various types of crosstalk may differ considerably, these numbers must be weighted. For instance, if the port isolation of the switching module is 40 dB, and the adjacent-channel isolation of the demultiplexer filters is 30 dB, then $w_{m,(\mathrm{b})} = w_{m,(\mathrm{c})} = 10^4 w_{m,(\mathrm{a})}/10^3$.

For a given lightpath i and its given wavelength in the network, the overall weighted crosstalk number is computed by

$$N_i^{\mathrm{LP}} = \sum_{m=1}^{M} \left\{ w_{m,(\mathrm{a})} n_{i,(\mathrm{a})}^{\mathrm{OXC}_m} + w_{m,(\mathrm{b})} n_{i,(\mathrm{b})}^{\mathrm{OXC}_m} + w_{m,(\mathrm{c})} n_{i,(\mathrm{c})}^{\mathrm{OXC}_m} \right\}, \tag{16.2}$$

where M is the total number of OXCs on the route of LPi.

The WA algorithm first chooses the wavelength(s) with the smallest N_i^{LP}, and uses one of the four standard selection methods (RP, FF, MU, or LU) if there are multiple such wavelengths. Using this approach, it was shown that blocking probabilities could be improved by a factor of 5 to 10 for typical network configurations over the standard PLI-agnostic WA algorithms.

16.2.2 Improved Cross-Layer RWA Algorithms

There have been numerous subsequent proposals for cross-layer RWA algorithms since the above early work. These proposals primarily differ in the following aspects: (a) the PLIs that are considered, (b) the signal transmission quality metric (Q-factor, BER, PMD, etc.), and (c) whether the algorithms are centralized or distributed. In this section, one such proposal [16.6] is described. In [16.6], four centralized cross-layer RWA algorithms, called SP2, HQ, MmQ, and MmQ2, are proposed and compared with a baseline algorithm, shortest path (SP). This work considers BER as the QoT metric and presents adaptive approaches that dynamically consider the QoT of existing and new connections when a lightpath request arrives before making a decision on whether to admit or block the arriving request, and what resources (route and wavelength) to allocate to the request if it is to be admitted. A distinguishing feature of this work is that it considers fairness. It is well known that long lightpaths suffer from higher blocking rates because of the extra links on which wavelengths are needed. This is exacerbated in optical networks in two ways – first, the wavelength continuity constraint adversely affects longer paths more than shorter paths, and second, PLIs similarly impact longer paths more than shorter paths. One of the algorithms in this work attempts to ensure that longer paths do not suffer unduly.

Fair and Adaptive RWA

Figure 16.4 shows a model of a typical transmission path assumed in this work. The lightpath traverses several nodes and fiber spans, and there are several amplifiers along the path. Four different PLIs are considered in this work:

- Intersymbol interference (ISI), which results from chromatic dispersion and is a single-channel effect
- Amplifier spontaneous emission (ASE) noise, which is independent of other lightpaths and is introduced by the optical amplifiers
- Nonlinear fiber crosstalk (due to XPM (cross-phase modulation) and FWM) and
- Node crosstalk due to leakages from the various node components [16.4].

The node and nonlinear crosstalks are dependent on the global network status, i.e., what other lightpaths are active, and therefore must be computed online when a request arrives. The Q-factor of a potential lightpath can be computed assuming that these four PLIs contribute to Gaussian noise terms. Typically, there is a threshold for the Q factor [16.6] that results in a maximum BER that must be satisfied by every lightpath.

The algorithms work according to the following general method. When a lightpath request arrives, for each wavelength λ_i, the links on which the wavelength is not available are first removed from the network graph. Then, the shortest path on this modified graph for wavelength λ_i is determined. Let this be SP(λ_i). If the set of paths in $\{\text{SP}(\lambda_1), \text{SP}(\lambda_2), \ldots, \text{SP}(\lambda_W)\}$ is empty (W is the number of wavelengths per link), then the call is blocked. Otherwise, one of the lightpaths is selected according to the four proposed policies, or the baseline policy SP. The five policies, including SP, work as follows:

- *SP* (shortest path) selects the wavelength that corresponds to the physically shortest path among the candidates.
- *SP2* is similar to SP but with *protecting threshold*. The protecting threshold technique [16.18] enhances fairness of longer connections (which have much higher blocking rates as explained earlier), by setting a threshold for available wavelengths that must be met for shortest paths to be admitted. In this work, a single-hop (i.e., a lightpath with just one link) path is admitted only if there is more than one wavelength available.
- *HQ* (highest Q factor) selects the candidate lightpath with the highest Q factor.
- *MmQ* (max–min Q factor) considers the Q factors of not only this lightpath but of other impacted

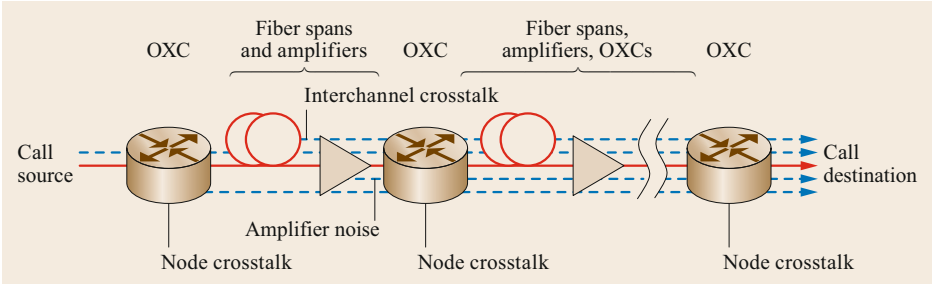


Fig. 16.4 A typical transmission path (after [16.6])

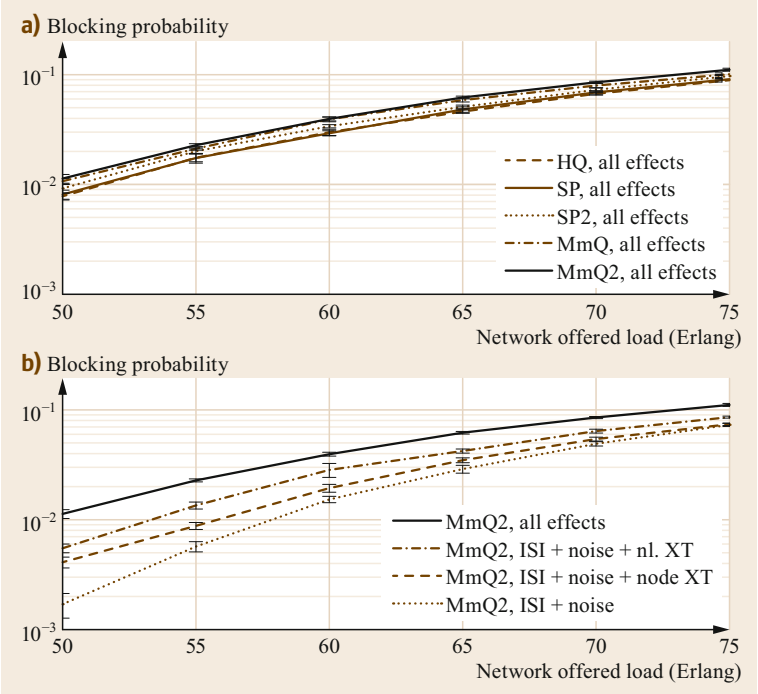


Fig. 16.5a,b Blocking performance of various algorithms (after [16.6])

lightpaths as well. Admitting a lightpath increases crosstalk in the network, possibly bringing the Q factor of a previously established lightpath close to the threshold or even violating it. MmQ maximizes the margin of QoT operation in the network by selecting the lightpath that maximizes (over the at most W candidates) the minimum Q factor (over the potential lightpath itself and all lightpaths previously established in the network that the potential lightpath crosses). MmQ is designed to decrease the worst-case network BER.

- *MmQ2* is MmQ augmented with the protecting threshold technique to enhance blocking fairness while maintaining good BER.

For each policy, ties are broken by choosing the first lightpath in the list of candidates.

Numerical Results

A few sample results are presented here to demonstrate the effectiveness of the proposed algorithms and to illustrate the relative impact that each PLI has on the performance. First, the blocking probability of calls for the proposed algorithms is presented in Fig. 16.5 for the standard NSFNET (national science foundation network) network topology (with a BER threshold of 10^{-9}). Figure 16.5a shows the performance of all algorithms; SP performs the best while MmQ2 is the worst, but all algorithms perform similarly. Figure 16.5b shows the blocking probability for MmQ2 for various PLIs. Network-state-dependent PLIs (i.e., crosstalks) can cause up to an order of magnitude difference in blocking probability.

The BER of admitted calls is shown in Fig. 16.6. Figure 16.6a shows the BER for all algorithms including all PLI effects. The benefit of HQ is clearly seen here, as it results in a much lower BER than the other algorithms. Figure 16.6b shows the effect of BER due to various PLIs on the MmQ2 algorithm. As in the case of blocking probability, the BER can increase by as much as an order of magnitude when all PLI effects are present as opposed to just the topology-dependent PLIs.

The fairness is quantified using Jain's fairness index [16.19], which is a number between 0 and 1 and is defined as follows. Suppose n users share a resource X. If X_i is the share of X received by user i, the fairness index is given by

$$f_X = \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{\sum_{i=1}^{n} X_i^2}\,; \tag{16.3}$$

f_X is 1 when the resource is equally shared and is $1/n$ when one user consumes all of X. The plots below show the fairness in terms of both blocking probability and

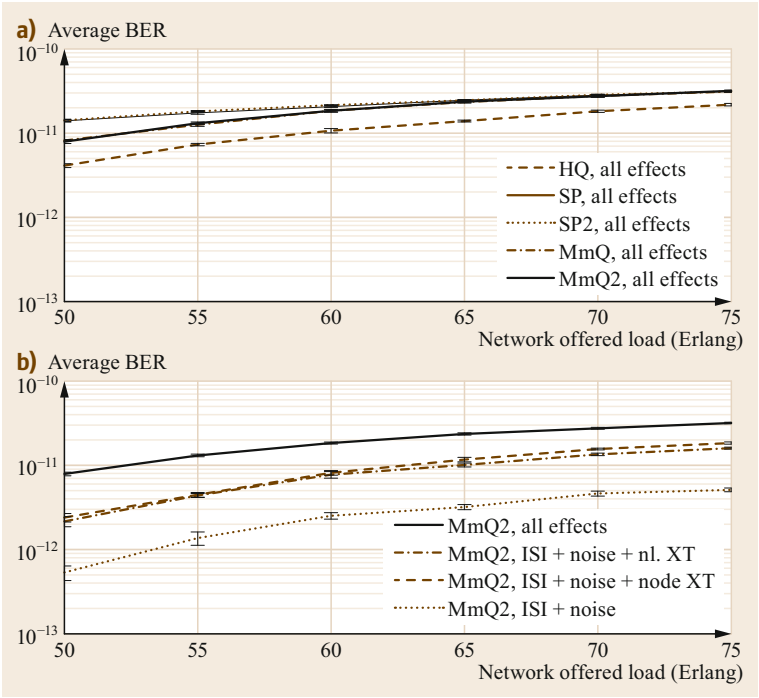


Fig. 16.6a,b BER performance of various algorithms (after [16.6])

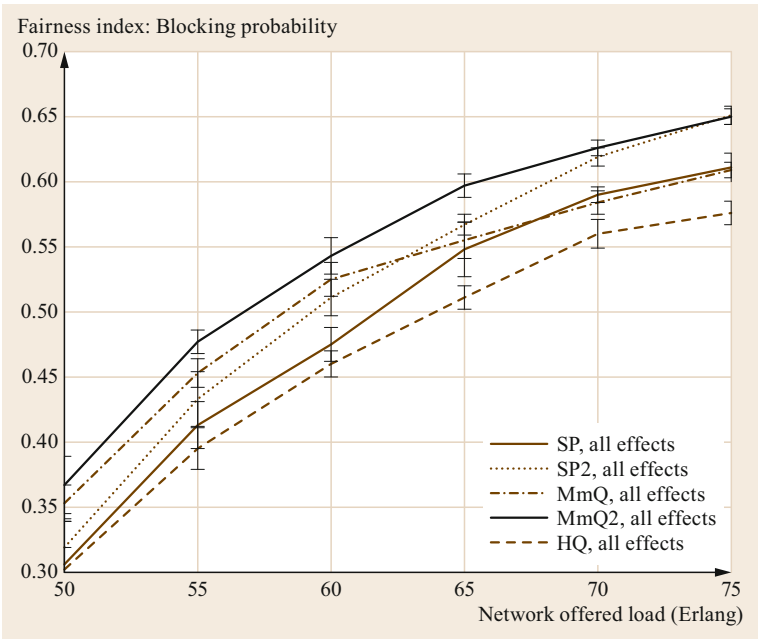


Fig. 16.7 Blocking probability fairness (after [16.6])

BER for the proposed algorithms. The MmQ2 algorithm outperforms the other algorithms in terms of both blocking probability fairness and BER fairness. It can be observed that HQ is much less fair even though it provides the best BER performance.

There has been an enormous amount of work on this topic, and this chapter has given a flavor of some of the work. The interested reader is referred to a host of survey papers on the topic (e.g., [16.20, 21]) for more information.

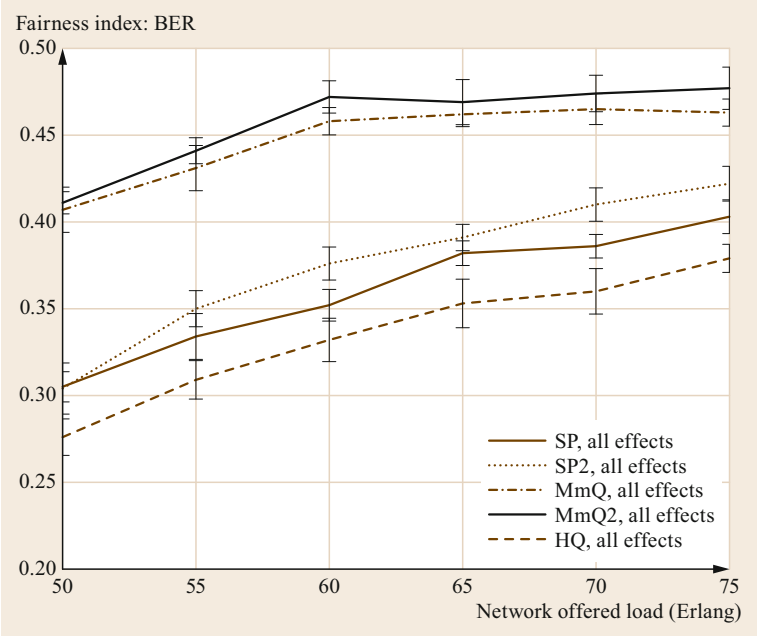


Fig. 16.8 BER fairness (after [16.6])

16.3 PLI-Aware Survivability

The last section focused on RWA in networks with PLIs. In such networks, the survivability of a network, i.e., the ability of a network to withstand and recover from failures, is also affected by the RWA used. While the blocking probability captures the performance of an RWA algorithm, another metric is needed to capture the survivability performance, because a connection may not survive a failure even if it was not blocked. The restorability of a connection, defined as the probability that a connection can be restored successfully after a failure event, is one such metric. Another metric that was specifically proposed for measuring the survivability performance of networks with PLIs is called the *vulnerability ratio* (VR), defined as follows [16.22]. The VR is defined as the probability that a randomly picked connection cannot be restored because of unacceptable QoT, if a single random failure event occurs at a random point of time during the operation of the network. Since the vulnerability of a connection can only change with the network state, VR can be calculated by averaging over all network states S. For a failure event j in network state i, the probability that a random ongoing connection fails is

$$P_i^j = \frac{D_i^j}{T_i}\,, \tag{16.4}$$

where D_i^j is the number of the connections that cannot be restored due to unacceptable QoT, and T_i is the total number of active connections in state i. Then,

$$P_i = \sum_{j=1}^{F} p_j P_i^j = \frac{1}{L}\sum_{j=1}^{L}\frac{D_i^j}{T_i}\,, \tag{16.5}$$

where p_j is the probability of failure event j, and F is the number of possible failure events. Then, the VR $\mathcal{V}$ is computed by averaging P_i over all the states

$$\begin{aligned}\mathcal{V} = E[P_i] &= \frac{1}{\sum_{i=1}^{S}\tau_i}\sum_{i=1}^{S} P_i\tau_i \\ &= \frac{1}{\sum_{i=1}^{S}\tau_i}\sum_{i=1}^{S}\sum_{j=1}^{F} p_j\frac{D_i^j\tau_i}{T_i}\,,\end{aligned} \tag{16.6}$$

where τ_i is the duration of state i.

Protection, wherein redundant resources are reserved at the time a connection is established, is a well-known approach for survivability in optical networks. Protection can be implemented at the path level or link level. Further, protection resources may be dedicated to a path/link or may be shared. Conversely, restoration, another approach for survivability, does not reserve resources beforehand; instead, any available resources are

discovered and used to recover from a failure after the failure happens [16.23]. This section examines the performance of protection and restoration algorithms in networks with PLIs [16.22]

16.3.1 Path Protection and Restoration

In (1+1) path protection, two link-disjoint lightpaths are established for every connection, a *primary* path and a *backup* path. Typically, data is transmitted on both paths simultaneously, and the receiver automatically selects the signal with the better SNR. (This is the case in dedicated protection; in shared protection, the backup path is established and its links possibly shared with other backup paths, but not used until a failure occurs.) This ensures quick recovery from the failure of a link on one of the paths. In networks with PLIs, making both primary and backup paths active (i.e., lit) at the same time may adversely impact the QoT of all the lightpaths due to crosstalk. Conversely, keeping the backup path inactive (i.e., dark) may increase failure recovery time.

For the case of the lit backup path, the RWA algorithm is used to determine two link-disjoint paths, and the connection is blocked if they cannot be found. If at least one of the two lightpaths has a satisfactory BER, and if activating both lightpaths does not violate the BER requirement of *both* the lightpaths of any other connection, then the connection is accepted.

When the backup path is not activated, the BER for all primary paths of every ongoing connection must be satisfactory. If the BER constraint is violated with one path, then the same procedure is repeated with the second path found for the arriving connection. If both fail, the connection is blocked. Three RWA algorithms, namely, SP, BF (best fit), and HQ are considered. In SP, the shortest path and the first fit wavelength is chosen for each of the two paths. In BF, the shortest path on each wavelength is computed, and the shortest among these is selected. HQ, as described earlier, finds the shortest path on each wavelength and picks the one with the highest Q factor.

When a link fails, each affected primary lightpath must be switched over to its corresponding backup. In the lit backup scheme, the backup path must be checked for satisfactory BER. Recovery is successful if the BER is adequate; otherwise, not. In the dark backup scheme, the backup path must be activated upon failure, and the BER of this path as well as the primary lightpath of every ongoing connection must be checked for adequate BER. Recovery fails if any of these BERs are too high.

In the case of path restoration, an RWA algorithm is used to establish a path for an arriving connection. When a link fails, a restoration path between the source and destination are determined by using the same or different RWA algorithm. In the nomenclature used in the results later on, the RWA for the primary path and that for the restoration path are both used; for example FF–FF path restoration uses FF for both the primary and the restoration lightpaths. Three schemes are compared: in FF–FF, SP routing with FF WA is used for selecting both the primary and the restoration paths; in HQ–HQ, the HQ algorithm is used for both paths; the BF RWA algorithm is used for both paths in BF–BF. Results for other combinations can be found in [16.24].

16.3.2 Link Protection and Restoration

The protection entity in link protection is a single link – for each link, a backup path to detour the traffic if that link fails is computed offline. These backup paths can be computed in a number of ways. The following approach is used in this chapter.

The backup paths for links are first computed using the following algorithm [16.25]. A 2-connected directed subgraph, called the *blue* digraph (directed graph) is first determined. Another subgraph, which is the same as the blue digraph but with the edge directions reversed, called the *red* digraph, is then generated. The set of wavelengths on each fiber is divided into two sets, Λ_1 and Λ_2. Λ_1 is reserved for use by primary paths on the blue digraph and by protection paths on the red digraph; similarly, Λ_2 is reserved for primary paths on the red digraph and for protection paths on the blue digraph. When a connection request arrives, a primary path is found on either of the two digraphs by using the appropriate wavelength set. SP routing with FF WA (FF) is used to find the primary path. Note that the shortest path may be either in the blue digraph or in the red digraph. When a link fails, the connections traversing the link on the blue digraph are rerouted on the backup path on the red digraph, thus avoiding the failed link. This can be done because Λ_1 is reserved for protection in the red digraph and, therefore, available for all the affected connections. A similar method is followed for rerouting the connections traversing the failed link on the red digraph.

In contrast to link protection described above, backup paths are not precomputed in the case of link restoration. An RWA algorithm (e.g., FF) is used for an arriving connection request. When a link fails, the shortest available path around that link on the same wavelength is determined for each affected connection. Note that different connections may use different restoration paths in this case.

16.3.3 Protection and Restoration Performance Evaluation

Figs. 16.9 and 16.10 show the blocking probability and the vulnerability ratio of both the lit and dark backup methods. It can be seen that the dark backup scheme has a much better blocking probability than the lit backup scheme, especially at lower loads, because of the high crosstalk introduced by the latter method. This comes at the expense of longer traffic recovery times because the back up path needs to be activated and the QoT of all ongoing connections checked for adequacy, after the failure occurs. HQ, which is a QoT-aware algorithm, significantly improves the performance of both schemes. As the loads increase, the performance of all algorithms is limited by wavelength blocking and QoT-awareness cannot help with improving the performance.

When the network experiences a link failure, some backup paths that become primary paths may not have sufficiently high QoT to be used. This is captured by the vulnerability ratio (VR), plotted in Fig. 16.10. The HQ algorithm has much better VR at low offered loads, especially for the lit scheme. When the backup is dark, HQ cannot predict the performance of the path when it is activated. When the load is very high, HQ actually performs worse than SP; this is because HQ admits more connections than SP, which means that there are more connections to recover upon failure.

The performance of link and path protection and restoration algorithms are compared in Figs. 16.11 and 16.12. From Fig. 16.11, one can see that link protection has high blocking probability due to severe degradation of QoT. The reason for this is that the path for a connection may be much longer than the shortest path because it is restricted to lie on one of the two digraphs. Longer path lengths mean more noise and crosstalk in more intermediate nodes and ultimately lower QoT. We see from Fig. 16.11 that the BF restoration algorithm performs much worse HQ. The HQ–HQ path restoration algorithm performs significantly better than other algorithms in the presence of PLIs, which supports the idea of using cross-layer approaches.

Link protection algorithms also have high VRs (Fig. 16.12), again because of the long paths used by connections. When the protection path of a link is activated, the QoT of a large number of ongoing connections may be affected, thereby making the connection unrecoverable. It must be noted that if we had a perfect physical layer, the VR would be zero in 16.12 for all schemes except the restoration schemes, which do not reserve resources to handle failure events.

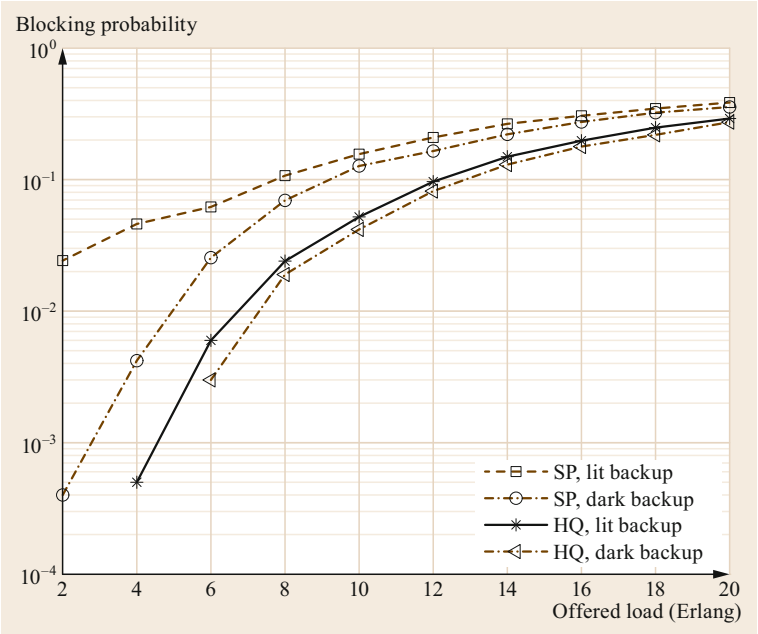


Fig. 16.9 Blocking probability versus traffic load for dedicated path protection (after [16.22])

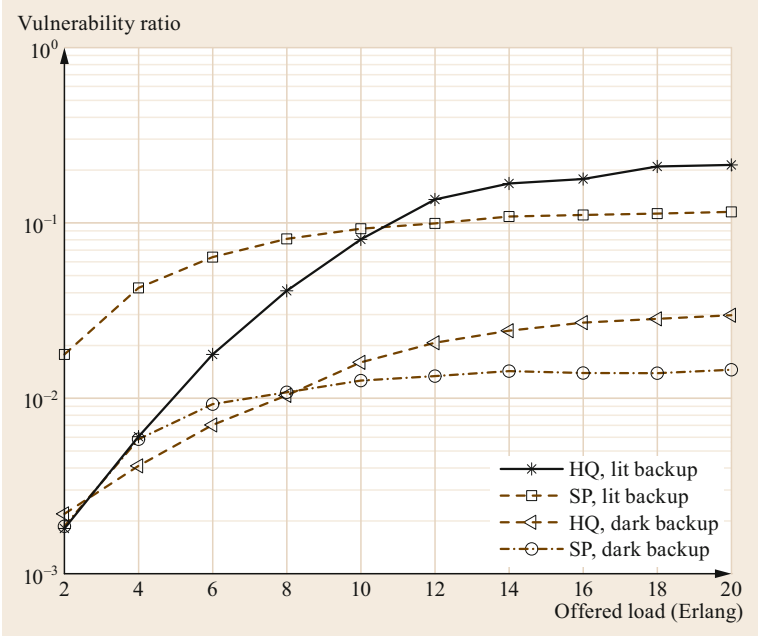


Fig. 16.10 Vulnerability ratio versus traffic load for dedicated path protection (after [16.22])

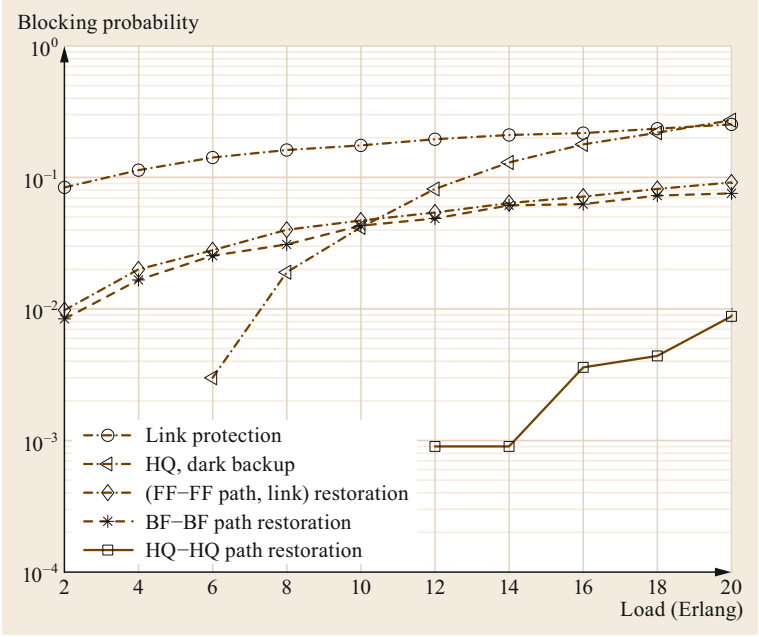


Fig. 16.11 Blocking probability versus traffic load for link and path protection and restoration (after [16.22])

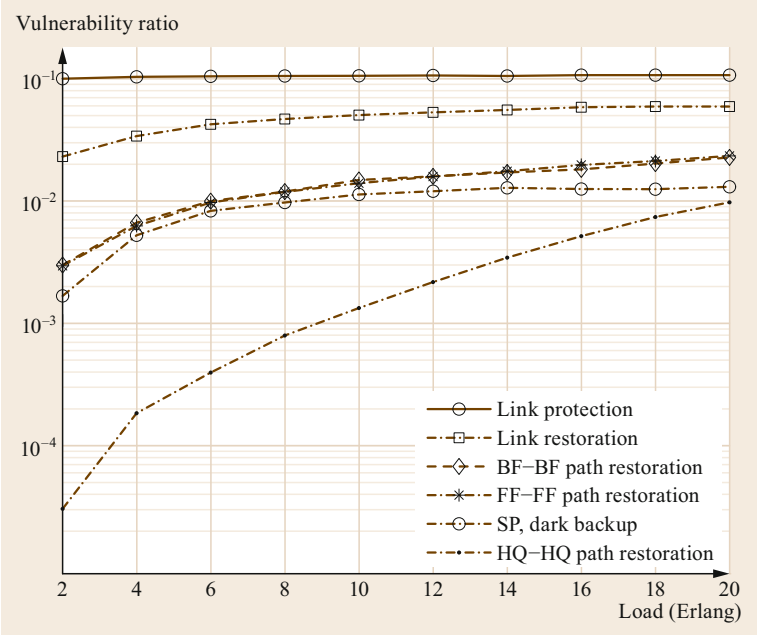


Fig. 16.12 Vulnerability ratio versus traffic load for link and path protection and restoration (after [16.22])

16.4 Application-Aware Metro-Access Programmable Architecture

The network is becoming progressively content-centric as users are mainly interested in accessing data and services, irrespective of their physical locations [16.26]. In contrast to network connectivity, content connectivity (i.e., reachability of content from any point of a network) is becoming increasingly important [16.27] to ensure users access to data even in case of failures. Internet usage has evolved with the development of smartphones and apps which is shaping today's user behaviors and their expectations [16.26]. A decade ago, connectivity requirements were limited to certain types of devices and locations; now, with proliferation of various kinds of hand-held devices, network users expect seamless connectivity to cloud, data, apps, IoT, etc. from any location. Devices are getting smarter to support more intelligent apps with the goal to provide personalized services.

As apps are now an indispensable part of daily lives, consistent and reliable access to applications is expected anytime, anywhere, on any device. Users require prompt online services or else they leave or choose alternate sites/apps. In 2008, the typical acceptable response time was 4 s and this reduced to 2 s in 2016 [16.28]. However, current users demand instant access to high-quality content with uninterrupted viewing despite mobile device type or even congestion in the cellular network. Users are no longer patient enough to tolerate a degraded level of experience (low-quality picture and occasional buffering) which can have a significant impact in terms of functionality. Hence, apps need to be optimized to perform regardless of device and network limitations – to provide a more *user-oriented* service.

With increasing expectations of network users, driven by the growth of devices and applications, the network's role has itself become updated from merely providing connectivity to efficiently delivering applications [16.26]. The interaction between the network and the applications, i.e., application-centric networking, is made possible with the help of virtualization and programmability [16.29] of networking devices. As a result, network capabilities have enhanced to track important information about the applications that are running on it to optimize its performance [16.30].

This requires a different level of management, beginning with the application and exploiting more than just traditional networking information (in contrast to managing individual servers and routers as in the past) [16.26]. It can benefit from software-defined networking (SDN) [16.32], which enables global management of a network by abstracting its lower level

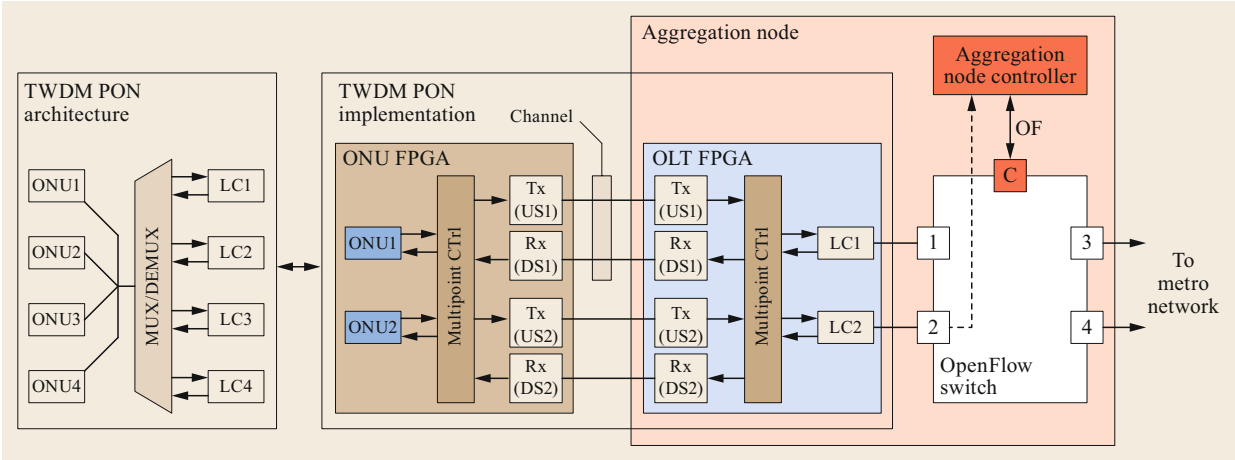


Fig. 16.13 System architecture (after [16.31])

functions [16.33] and decoupling the control plane from the data plane. In this way, an application-centric network can directly correlate the resource-allocation strategies in the network with quality of experience of clients [16.34]. SDN has been applied to numerous network scenarios and demonstrated by a large number of studies, not only by academia but also in the industry [16.35, 36]. In [16.37] and [16.38], SDN has been proposed for managing mobile backhauling. Moreover, an SDN-based approach for provisioning multitechnology multitenant connections has been proposed in [16.39].

In application-centric networking, success is determined by the network user's experience [16.26]. In this part, we analyze novel methods for three network-to-application layer designs. First, SDN is exploited to preserve the user experience in an integrated metro-access network. A major metric to measure a network user's experience can be application responsiveness or uptime [16.26]. In the second use case, we review a restoration mechanism that minimizes mobile user downtime by leveraging multilayer design that considers optical metro and mobile radio access networks. In the third use case, we show how to utilize application information from clients, namely the target delay, to tune resource allocation in optical and mobile access network leading to better quality of experience.

Considering the concept of SDN, several solutions have been proposed with multiple networking scenarios as in [16.40, 41]. This section introduces studies related to the optical metro-access network. For example, [16.41] describes the SDN-based passive optical network (PON) scenario, which highlights the advantages enabled by SDN for business-to-business (B2B) applications. Considering the SDN literature, so far, SDN-based architectures for a single network scenario are proposed to replace existing control and management solutions.

This section describes an SDN-based architecture applied to multiple network segments (e.g., access, metro); in particular, an implementation of a latency-aware controller to effectively control heterogeneous technologies. Moreover, the section also shows how to preserve the end-to-end delay of the connection across the metro and access segments of the network. The architecture consists of a time- and wavelength-division multiplexed (TWDM)-PON access and the layer 2 switch of an aggregation network. The scheduling priority of the frames in the metro segment is dynamically changed based on the number of active line cards (LCs) in the access network. Thus, the overall service level requirements [16.42] can be met even when additional latency is introduced by the sleep operation (for energy savings) in the TWDM PON [16.43].

16.4.1 System Architecture

Figure 16.13 depicts the considered SDN-based architecture. The architecture consists of optical line terminal (OLT) LCs, each transmitting on a different wavelength. The LCs are connected to the Ethernet OpenFlow switch of the aggregation node, and the aggregation node is connected to the metro network. On the other end, at the user premises side, optical network units (ONUs) are assumed to be equipped with tunable transmitters, thus, the ONUs can flexibly transmit users' traffic to any LCs and receive users' traffic from any LCs.

The aggregation node controller is implemented (as a light version of an SDN controller) to control the OpenFlow switch. The aggregation node controller is responsible for the following activities:

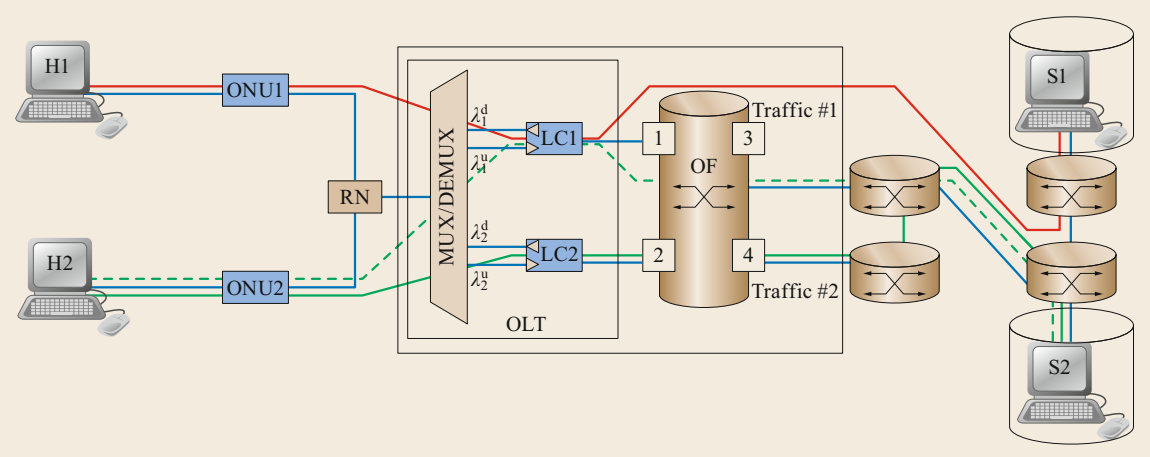


Fig. 16.14 Implemented scenario architecture (after [16.31])

(i) Configuring the flow entries in the OpenFlow switch, to forward the flows to the appropriate LC by considering both normal and sleep-mode operations of the LCs
(ii) Monitoring the status of the aggregation node ports, which includes the internal ports between LCs and the OpenFlow switch and
(iii) Installing or modifying the connections through the metro network.

16.4.2 Implementation Scenario

Figure 16.14 shows the implemented scenario, where two ONUs and LCs of a 10-Gb/s Ethernet-based TWDM-PON are considered. The OLT is comprised of four bidirectional channels, and two of the channels are connected to ONU1 and ONU2. The ONUs can transmit/receive traffic on the corresponding wavelength pairs at the LC for upstream (US)/downstream (DS) traffic. Moreover, both ONUs can transmit traffic on a single wavelength pair by electronically multiplexing the traffic in a time-division multiplexing (TDM) fashion. The OLT receives the US traffic generated at the ONU. Upon receiving the traffic at the OLT, it forwards the traffic towards the metro network by utilizing the OpenFlow switch. A 5-port server with OpenVSwitch (OVS) (version 2.61) is used to implement the OpenFlow switch. Note that the two ONUs and two LCs are implemented in an Altera Stratix® IV field-programmable gate array (FPGA).

16.4.3 TWDM-PON and OpenFlow Implementation

The TWDM-PON features a dynamic wavelength and bandwidth allocation (DWBA) scheme as depicted in Fig. 16.15. The general approach to implement the DWBA is to utilize two-phase scheduling schemes such as TWDM and TDM. In the TWDM phase, each ONU is assigned to different LC (e.g., ONU2 is assigned to LC2 and ONU1 is assigned to LC1). But, in TDM, both ONUs share a single LC. The OLT sends a `tuning GATE` notification to the ONU (e.g., ONU2) that needs to change its wavelength to communicate with another LC (i.e., LC1). At the same time, the OLT sends a reconfiguration request to the SDN controller. The following actions are performed upon receiving the reconfiguration request at the SDN controller:

(i) Reconfiguration of the OpenFlow switch scheduling priorities
(ii) Traffic balancing towards the metro network and
(iii) Interaction with the metro network controller for possible flow reconfiguration.

As shown in Fig. 16.15, the tuning time is defined as the time required to tune both the ONU transmitter and receiver to a different wavelength. During this process, both DS and US traffics are buffered and the corresponding LC (i.e., LC2) enters into sleep mode. When the reconfiguration process successfully completes, the immediate timeslot is assigned to the tuned ONU (i.e., ONU2 in this case). The successive timeslot is assigned to ONU1 after the ONU2 timeslot expires. In the TDM phase, the single wavelength is shared among both ONUs till the next reconfiguration is triggered.

The aggregation node controller is implemented as a simple, light version of the SDN controller as presented in [16.31]. The Python-based openflow/SDN controller (POX) controller [16.44] was used to implement the aggregation node controller by adding several

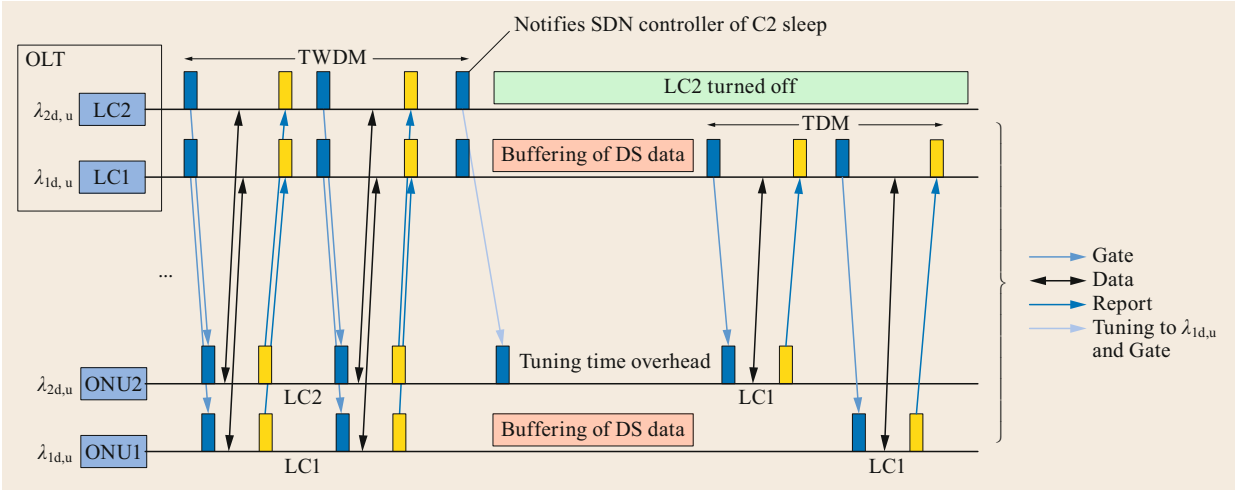


Fig. 16.15 Dynamic wavelength and bandwidth allocation (DWBA) scheme

ad hoc functionalities. The Linux server (Ubuntu 13.04) is used to implement the OpenFlow switch. The flows from the OLT are aggregated by utilizing the OpenFlow switch, and the flows are differentiated by using virtual local area network (VLAN) IDs. For example, as shown in Fig. 16.14, initially, the traffic path between host 1 (H1) and server 1 (S1) is indicated by a traffic path traffic 1, and the path between host 2 (H2) and server 2 (S2) is indicated by a traffic path traffic 2. However, when the sleep mode is enabled in the OLT and the reconfiguration is triggered, port 1 receives both traffics (i.e., traffic 1 and traffic 2) with different VLAN IDs. Table 16.1 shows possible match and corresponding output actions for the sleep mode and the normal mode.

16.4.4 Performance Evaluation and Results

Table 16.1 shows the considered experiment performance parameters. The US frame arrival process follows a Poisson distribution with a fixed frame size of 1250 B. In the considered experiment, the value of round trip time (RTT) is negligible and the traffic of US1 and US2 is set to 64 Mb/s. The time period T_{rec} is used to generate reconfigurations periodically at TWDM-PON. Tuning time T_t is emulated by using the tuning timer. The considered performance parameter is the average frame delay, which is defined as the average time interval between the arrival of a frame into the buffer and its departure from the buffer. The experiment duration is set to 4 s for all transmissions between the OLT and the ONUs.

Figure 16.16 shows the average US frame delays in the TWDM-PON as a function of reconfiguration time T_{rec} with different tuning time values (i.e., 5 and 10 ms). All the delays are obtained by setting the timeslot T_{slot}

Table 16.1 Flow table, sleep and normal mode

Match (in_port, Vlan_id)	Actions	Traffic	Sleep_mode
(1, 11)	Output: 3	Traffic 1	Disabled
(3, 11)	Output: 1	Traffic 1	Disabled
(2, 12)	Output: 4	Traffic 2	Disabled
(4, 12)	Output: 2	Traffic 2	Disabled
(1, 21)	Enqueue: 3:2	Traffic 1	Enabled
(3, 21)	Output: 1	Traffic 1	Enabled
(1, 22)	Enqueue: 4:2	Traffic 2	Enabled
(4, 22)	Output: 1	Traffic 2	Enabled

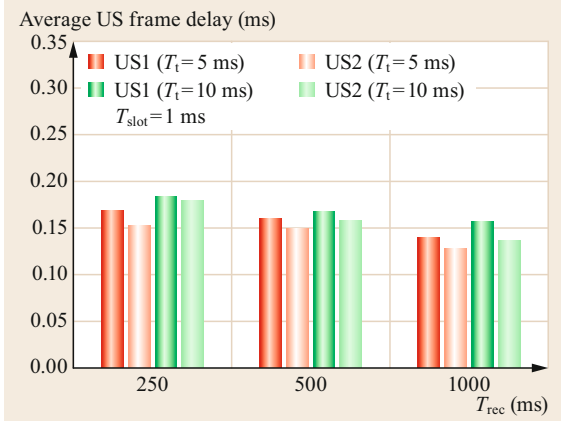


Fig. 16.16 Average US frame delay as a function of reconfiguration time

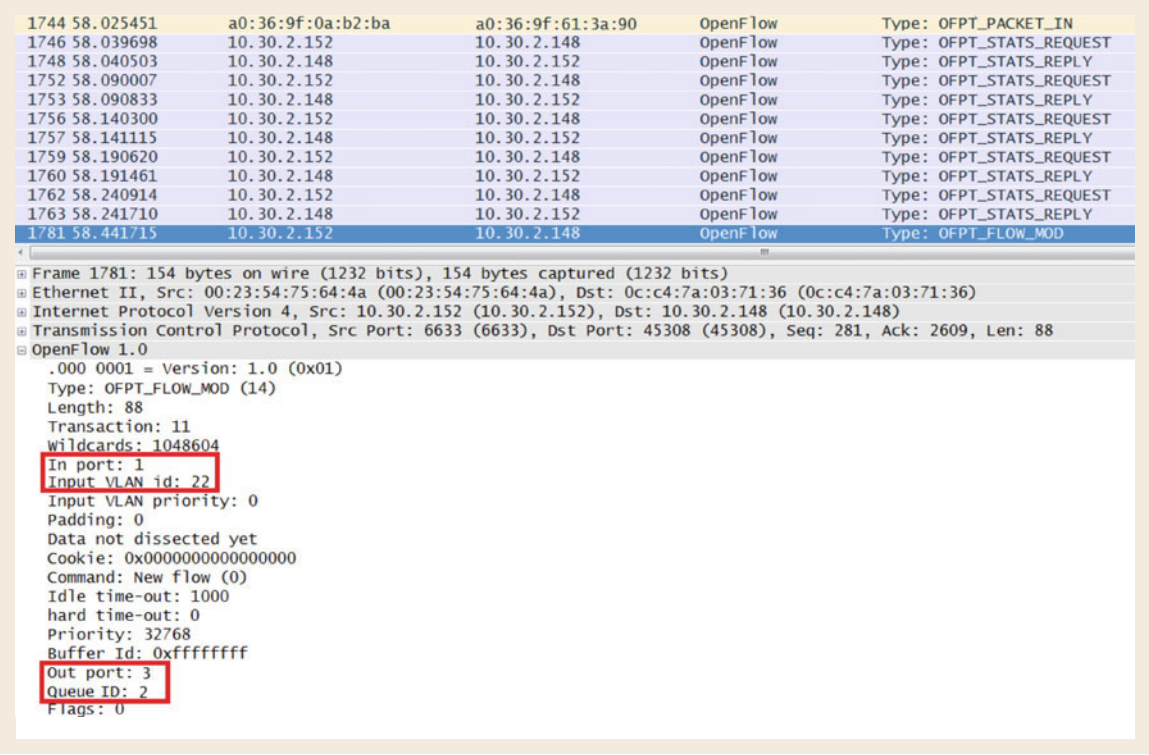


Fig. 16.17 Control traffic capture (after [16.31])

to 1 ms, and different fixed reconfiguration times (i.e., 250, 500, and 1000 ms). It can be observed that the average US frame delay decreases with an increase of T_{rec} because of less reconfigurations during the fixed experimentation time (i.e., 4 s). Moreover, the average US frame delay increases when the tuning time T_t increases (i.e., from 5 to 10 ms), because the number of frames buffered during the tuning process increases. This behavior can be observed even for the larger reconfiguration periods (i.e., 1000 ms).

The contribution of increased average US frame delay can be compensated with the node controller by communicating with the OpenFlow switch to adapt the implemented DWBA scheduling behavior. Thus, the adaptable scheduling provides high priority to packets that experience the additional latency due to the sleep operation. The aggregation node controller configures the OpenFlow switch with two flow entries (i.e., one for normal operation, and one for sleep mode).

Figure 16.17 shows control traffic capture between the OpenFlow switch and the aggregation node controller. When the reconfiguration is performed at the TWDM-PON, the traffic directed to the ONUs are assigned a high priority class of service to compensate for the increased average US frame delay. In order to change the operational status (i.e., normal operation or sleep mode) of the aggregation node controller, the LC2 sends a specific control packet. Here, the OpenFlow switch sends the OFPT_PACKET_IN message to the controller. The aggregation node controller then monitors the status of the high-priority queues by requesting the OpenFlow switch with the OFPT_STATS_REQUEST message, and receives the queue status response in OFPT_STATS_REPLY. On the other end, when the aggregation node controller detects that load at the ports is low, it modifies new flow using the OFPT_FLOW_MOD message. As shown in Fig. 16.17, the in_port 1 with VLAN_ID 22 is sent through out_port 3 with Queue_ID 2.

Note that the tuning time at the OLT must consider the time required to perform the OpenFlow switch reconfiguration. For example, here, the minimum required tuning time during the experiment is set to 5 ms because the successful reconfiguration takes around 4 ms.

The next section presents the SDN and packet-based protection scheme for a seamless user experience when network failures occur.

16.5 SDN-Based Resource Allocation and Path Protection

The 3GPP TR 38.801 [16.45] defines several functional split options where the next-generation NodeB (gNB) functions are separated and hosted at two distinct entities; namely a central unit (CU), deployed in a centralized location, and a distributed unit (DU), deployed near the antenna. The connection between the CU and DU is known as the fronthaul, whose role is to facilitate communication without adversely affecting the radio performance. DU, CU, and fronthaul, all together form the 5G next-generation radio access network (NG-RAN), also known as the centralized/cloud RAN (C-RAN).

TR 38.801 mentions only the specific requirements about one-way delay and capacity for the fronthaul, however the requirements regarding the reliability of the fronthaul and packet loss are not defined. Available literature show several methods evidently proposed to recover the C-RAN from different types of failures (i.e., hardware and software). For example, resilient schemes are described to recover from C-RAN failure based on the concept of the access cloud network (ACN) in [16.46], and some schemes are proposed to recover from the failing of a virtualized mobile network in the case of a grid [16.47] and cloud network [16.48, 49].

Fronthaul network failures might also exploit the underlying network recovery schemes [16.51]. Recently, an interesting scheme put forth a two-step recovery approach by orchestrating the adaptation of lightpath transmission along with the eNB functional split reconfiguration. This scheme allows recovery of the virtual DU (i.e., deployed in a virtual machine or container) and virtual CU connectivity, and at the same time it also fulfills the requirements of the fronthaul capacity [16.52].

The above-mentioned schemes, however, do not guarantee seamless communication between DU–CU during recovery. In fact, since the network takes time to detect the link outage and forward the connection over an alternative path, it is quite common to experience some packet loss during a recovery. Such a packet loss might put the DU–CU communication out of synchronization; as a consequence, the mobile network software implementations can cause disconnection. A disconnection forces the DU and CU to restart, which further prolongs recovery. Hence, recovery schemes taking into account the specific requirements in terms of recovery time and packet loss of the fronthaul for different functional splits need to be explored.

16.5.1 1+1 SDN and Packet-Based Protection Scheme

This section describes an SDN (i.e., OpenFlow) and packet-based 1+1 protection scheme to overcome a single fiber failure in a multilayer Ethernet/optical metro network fronthaul featuring an Intra-PHY split (i.e., split option 7-1). The main objective of this protection scheme is to minimize the loss of packets during recovery.

Figure 16.18 illustrates the operation of the 1+1 protection scheme. Two disjoint optical paths are computed and provisioned by the network orchestrator. The protection scheme is specifically implemented in kernel software because it is not readily available in Open-

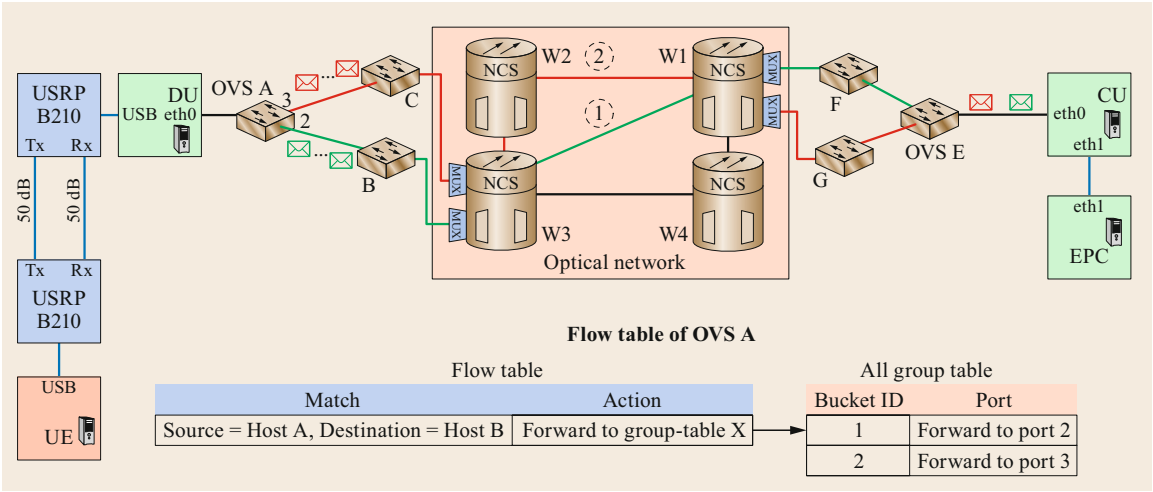


Fig. 16.18 1+1 protection mechanism as implemented in the PROnet testbed (after [16.50])

Flow/open V-switch (OVS) [16.53, 54]. The proposed scheme operates as follows: the ingress flow (coming from the transmitting DU) is forwarded to the two optical circuits (primary (1) and secondary (2)), which are interfaced by the ingress OVS by using the *all group table function* [16.54]. At the egress OVS, the two copies of the same flow arriving from the two optical circuits are processed in real time to remove duplicate packets before forwarding the traffic to the receiving CU. The same process takes place when the fronthaul traffic flows in the reverse direction. The egress OVS processes the packets from the two incoming flows using netfilter. Netfilter is a set of hooks in the Linux kernel network stack. The hooks can be used to define the custom functions to manipulate the network packets [16.55].

The following sequence of actions is performed each time a packet is captured by the hook: (i) a unique signature is created for the packet using some of the packet field(s); (ii) if the packet signature is found in the check list, the packet is discarded and the signature value is removed from the list; else (iii) the packet is passed onto the switch that carries out the necessary forwarding and the packet signature is stored in the check list.

Note that the packet signature must be computed to be unique for each received packet and its copy. For example, when transmitting packets over the fronthaul by using functional split option 7-1 (i.e., the intra-PHY split), the signature is computed using data in the radio link control (RLC) packet (i.e., frame, subframe, and symbol fields), as these three parameters combined form a unique triplet for each packet that is transmitted on the fronthaul.

16.5.2 Evaluation Scenario and Results

The scheme is experimentally evaluated in the programmable optical network (PROnet) [16.56] testbed, a two-layer – Ethernet-over-DWDM (dense wavelength division multiplexing) – network deployed at and around the University of Texas at Dallas. The PROnet configuration is shown in Fig. 16.18. Ettus B210 boards and Intel-i7 servers (acting as user equipment (UE), DU, CU, and the evolved packet core (EPC)) running the OpenAirInterface (OAI) [16.57] mobile network software are used for the experiment. The functional split IF4.5 (also referred to as option 7-1) is considered for the stated deployment. All used servers are equipped with 1GE NICs (Network Interface Cards) that are sufficient to sustain the single UE traffic demand with 5 MHz channel bandwidth. The DU and CU are connected to an OVS, which is controlled with OpenFlow 1.3 and equipped with multiple 1 GE interfaces. The Dell N2048 switch is used in a static configuration to connect the OVS to the optical muxponder. The DWDM layer consists of four Cisco NCS 2000 ROADM nodes, which are controlled and reconfigured dynamically through Netconf/YANG and TL1 interfaces.

The principles of SDN are extended in PROnet to control and coordinate the use of equipment at both the Ethernet and the DWDM layer. PROnet is controlled using the network (PROnet) orchestrator [16.58].

A more in-depth description of the PROnet testbed and its multilayer reliability mechanisms can be found in [16.56, 58]. In the 1 : 1 protection scheme, network resources are first provisioned through a network orchestrator and then the C-RAN modules (e.g., EPC, DU, CU, and UE) are started. Once the UE is connected, a fault is introduced by manually disconnecting one of the fiber spans used in the primary circuit.

As soon as the fault is introduced, the C-RAN stops working and the UE gets disconnected as shown in Fig. 16.19. This is because of the strict network recovery time/packet loss requirements that must be guaranteed between the CU and DU in the OAI C-RAN implementation. In fact, the fast failover table protec-

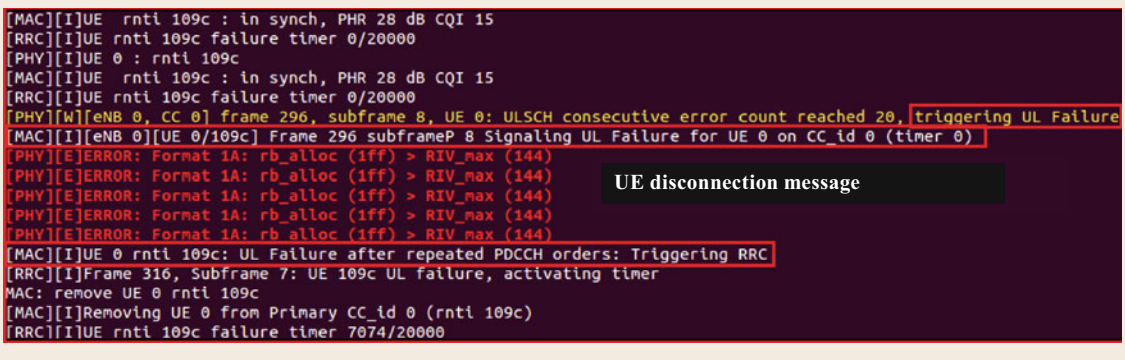


Fig. 16.19 UE disconnection message when 1 : 1 protection is implemented

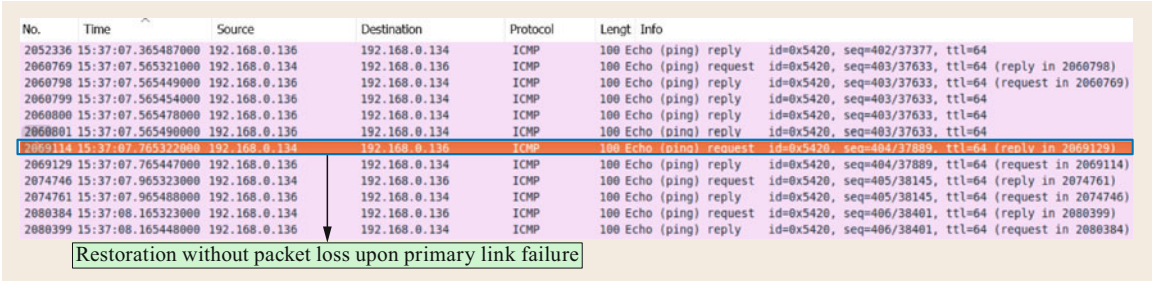


Fig. 16.20 Proposed 1+1 protection wireshark output at the server hosting the CU

tion mechanism requires a 100 ms recovery time, and it does not guarantee the delivery of all the transmitted packets over the fronthaul.

The experiment is also conducted with the 1+1 protection scheme. Figure 16.20 shows the wireshark output at the CU. The CU (IP address: 192.168.0.134) is set to transmit periodic echo request packets to the DU (IP address: 192.168.0.136), at intervals of about 200 ms. On start-up of the 1+1 protection mechanism, the kernel module at the egress node that drops duplicate packets is applicable only to C-RAN data packets and internet control message protocol (ICMP) packets (ping requests) therefore are not dropped. As a result, when ping operation is performed from the CU to the DU, both echo requests and echo reply packets are duplicated (e.g., 4 echo reply for 403 sequence number) by the 1+1 protection mechanism, as shown in Fig. 16.20.

Once the fault is introduced in one of the two optical circuits, both request and reply packets are received only once, as the echo packets cannot make it through the failed (bidirectional) optical circuit. By inspecting the wireshark output, the exact moment at which the optical circuit was disrupted can be identified. The continuous sequence numbers reported in the trace reveal that no packet is lost during the experiment, including the moment when the optical circuit is disrupted.

The next section will present converged wireless-access resource scheduling that implements coordinated multipoint (CoMP) while preserving user experience.

16.6 Application-Aware Converged Wireless-Access Resource Scheduling

Softwarization is a key technology for 5G mobile networks. Programmability and centralized control of network components will enable support of the high traffic demands required by 5G. The concept of SDN has been extended to mobile and optical domains. The software-defined features in mobile networks can increase core and network performances by effectively allocating the resources between core and access networks [16.59]. Moreover, the integration between access and aggregation networks can be achieved with the help of software-defined access (SDA) [16.60].

In next-generation mobile systems (e.g., 5G), high throughput can be achieved by the increasing deployment of small cells. However, the cell interference increases because of simultaneous transmission at the cell edge. Thus, coordinated scheduling (CS) is a coordinated multipoint (CoMP) technique introduced to improve the cell throughput by reducing the interference among the cells [16.61].

The timely exchange of messages among eNBs is of extreme importance to reach the expected performance improvement, because wrong decisions may derive from obsolete information. Thus, the time taken by updated information to reach all the cooperating evolved NodeBs (eNBs) (i.e., the CoMP cooperating set) is the key parameter that determines the performance gain of the adopted CS scheme [16.62].

Recently, ONUs of PONs have been proposed to connect eNBs. TDM-PON-based mobile fronthaul and backhaul are shown to be a feasible solution for next-generation radio access in [16.63]. For example, in [16.64], the PON-based mobile fronthaul solution has been proposed to reduce fronthaul latency up to 95% by calculating bandwidth allocation in the PON from the radio resource control (RRC) mobile scheduling. Moreover, a TWDM-PON is proposed for implementing the radio access network (RAN) in [16.65].

X2 and S1 are two application protocols utilized by the mobile network to implement eNBs' cooperation and to carry user plane traffic. While X2 is utilized for control information, S1 transports the effec-

tive user traffic; thus, its performance impacts the user experience. In this use case, we describe scheduling techniques for the traffic belonging to the X2 interface through which CoMP messages are exchanged with the help of a TDM-PON dynamic bandwidth allocation (DBA) scheme. In addition, guidelines are drawn for the dimensioning of PON in terms of number of eNBs (i.e., ONUs) taking into account target X2 exchange time and S1 user latency to implement application-aware CoMP-CS.

16.6.1 CoMP and Coordinated Scheduling Overview

CoMP techniques are based on joint processing and coordinated scheduling/beamforming (CS/CB). Moreover, CoMP can be applied to both homogeneous and heterogeneous networks [16.66].

Among the CoMP techniques, CS allows multiple cells to share up-to-date message information between the entities that participate in the scheduling process [16.67]. CS can be implemented in two different approaches: Centralized CS (CCS), where a central unit computes the scheduling, and distributed CS (DCS), where the eNB exchanges the CoMP hypothesis set to take the decision in a distributed fashion.

Different techniques have been proposed to address the requirement of optimal number of eNBs in the CoMP cooperation set to improve cell throughput in CS [16.61].The timeliness of the CS hypothesis set available at all eNBs in the cooperating set is the main factor determining CS performance. The performance of the CS is determined based on the time required (i.e., convergence delay) to exchange the CoMP hypothesis set among the eNB in the cooperating set. However, this time mainly depends on the latency of the underlying backhaul network [16.68, 69].

The CoMP hypothesis informs all collaborating eNBs about the resources that the sender eNB cannot utilize due to strong interference. The eNB sets the values in the CoMP hypotheses based on the measurement received by the user equipments (UEs). The UEs can send the channel state information channel quality indicator (CSI/CQI) through the physical uplink shared channel (PUSCH) and the physical uplink control channel (PUCCH) in both periodical and aperiodical ways. The information contained in CSI/CQI are applied by the eNB to formulate the CoMP hypothesis.

Load information messages are exchanged to implement CS that contains CoMP information and the CoMP hypothesis set as described in [16.70]. Here, the information element (IE) in CoMP information provides a CoMP hypothesis set and each CoMP hypothesis set represents a collection of CoMP hypothesis(es) belonging to one or multiple cells. The interested reader can refer to [16.70, 71] for a thorough explanation about CoMP information and the CoMP hypothesis set. The interference-protected resources that cannot be utilized are notified using the CoMP hypothesis to all collaborating eNBs. The eNB sets the values in the CoMP hypotheses based on the channel quality indicator (CQI) measurement received by the UEs. The UE uses PUSCH and PUCCH to send CSI/CQI.

The CQI periodicity is set by RRC during the handshake procedures (i.e., RRC connection setup and RRC connection reconfiguration), and expressed in terms of number of subframes as described in [16.72]. Moreover, the performance of CoMP CS is based on the periodicity of CQI transmission. The eNB produces a new CoMP hypothesis set after reception of updated CQI in order to share with the collaborating eNBs. Figure 16.21 shows the optimal scheduling that can be applied by the eNB after reception of CQI. Furthermore, the time required to apply the optimal scheduling mainly depends on the time that collaborating eNBs need to share their CoMP hypotheses (i.e., convergence delay) of the load information messages.

16.6.2 TDM-PON DBA for Supporting CS

Figure 16.22 shows the considered architecture, where the ONUs are attached to small-cell eNBs, and the OLT is connected to a macrocell. A macrocell is a cell that provides the largest area of coverage within a mobile network. As we can see, the small cells are deployed to increase the macrocell network capacity. The 10-Gb-capable PON (XG-PON) is considered to exploit the TDM-PON DBA for supporting CS. The XG-PON delivers traffic between the OLT and ONUs based on 10 GPON encapsulation method (XGEM) ports and transmission containers (T-CONTs). The OLT applies the DBA schemes to allocate the bandwidth to the ONUs.

Figure 16.23 shows the traditional DBA scheme, where the OLT grants an opportunity to transmit based on the report received from the ONUs. Thus, each grant size contains both X2 and S1 transmission. Figure 16.24 shows the proposed DBA scheme to reduce the waiting time of X2 to exchange CoMP hypothesis information among the collaborating eNBs in the network. Here, two different T-CONTs are adopted at the ONU for X2 and S1 interfaces. Thus, it helps to prioritize X2 traffic without involvement of the ONU upstream scheduler. When the OLT receives X2 T-CONTs transmission requests, it generates a bandwidth map to allocate all X2 traffic, and then all S1 traffic as shown in Fig. 16.24.

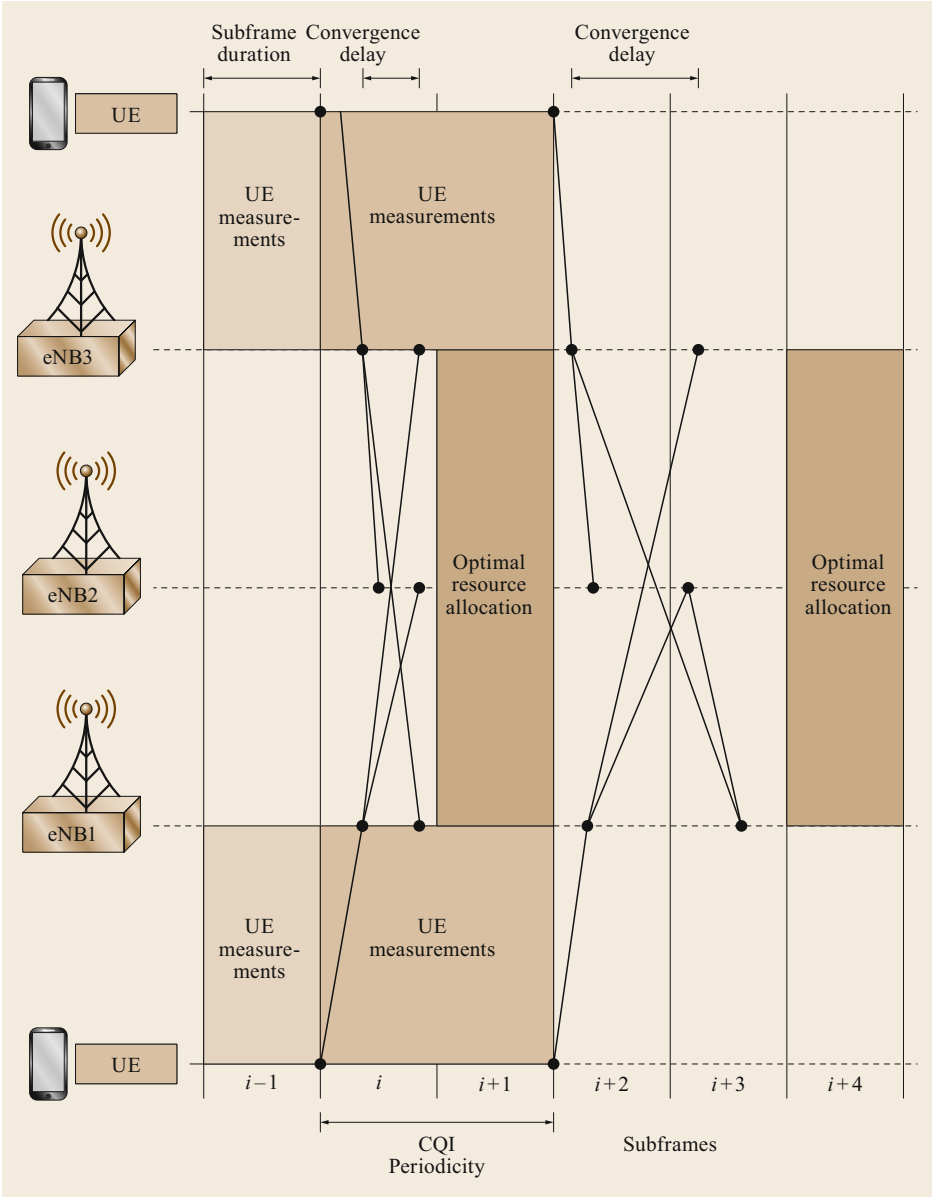


Fig. 16.21 UE measurement application

The X2 prioritization DBA scheme is applied in two different approaches: CCS and DCS. In the DCS approach, the ONUs in the cooperating set send X2 Application Protocol (X2AP) CoMP hypothesis messages to all the other ONUs. In the CCS approach, all the ONUs send their CoMP hypotheses to the OLT (i.e., macrocell eNB). Hence, the decision is taken in a centralized fashion and OLT informs its decision to all the eNBs.

The X2 prioritization could increase the delay of S1 traffic. One can model the X2 transmission as a *server vacation* for the S1 traffic and the *S1 waiting time*, introduced by vacation, can be approximated as half the average vacation time [16.73]. The X2 transmission time associated to the CoMP-CS process can be calculated as a function of the number of involved ONUs and exchanged messages; thus, it depends on the adopted CS scheme. In light of these considerations, the S1 waiting time in the presence of N ONUs transmitting X2 CoMP hypothesis messages of length L (in bytes) each, with an available upstream bandwidth BW_{UP}, can

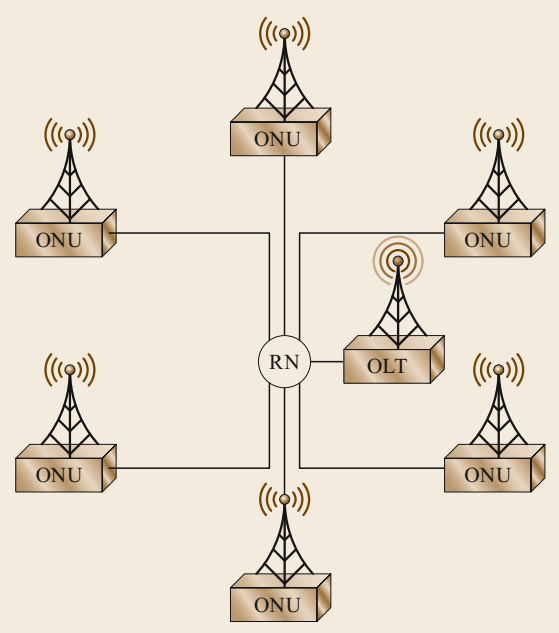


Fig. 16.22 Network configuration

be expressed in CCS as

$$W_{S1}^{CCS} = \frac{8NL}{2BW_{UP}} \tag{16.7}$$

and in DCS as

$$W_{S1}^{DCS} = \frac{[N(N-1)L8]}{2BW_{UP}}. \tag{16.8}$$

By comparing (16.7) with (16.8), it can be observed that, while W_{S1}^{CCS} grows linearly with the number of eNBs-ONUs, W_{S1}^{DCS} grows quadratically, because each eNB-ONU must transmit its CoMP hypothesis to the other $N-1$ eNBs, resulting in a total of $N(N-1)$ exchanged messages. Thus, in CCS, S1 traffic is less impacted by the prioritization of X2 traffic than in DCS.

16.6.3 Performance Evaluation and Results

The presented DBA scheme is implemented on the NS3 environment by modifying already available long term evolution (LTE) and XG-PON modules [16.74, 75]. The link (as shown in Fig. 16.22) distance between the ONUs and the OLT is set to 250 m. The considered link is based on a 10-Gb-capable PON (XG-PON), which is 10 Gb/s in downstream and 1.25 Gb/s in upstream direction.

The considered performance evaluation parameter is the *convergence delay*. The convergence delay in DCS is defined as the time required for the carried CoMP hypothesis to reach all eNBs in the cooperating set. The convergence delay in CCS is defined as the time that elapses between the transmission of scheduling information by all the eNBs in the cluster and the reception of the scheduling decision produced by the centralized coordinator.

Figure 16.25 shows the convergence delay obtained when DCS and CCS are supported by a DBA with and without X2 traffic prioritization. The presented DBA scheme which prioritizes X2 traffic over S1 traffic can potentially reduce convergence delay by 67% in both CCS and DCS.

Figure 16.26 shows convergence delay as a function of number of ONUs (ranging from 1 to 120). As we can see, the convergence delay is impacted as ONUs increase. When the considered approach is DCS, the convergence time of scheduling information increases with eNBs-ONUs due to PON upstream bandwidth saturation. Thus, DCS is not a suitable solution in the case

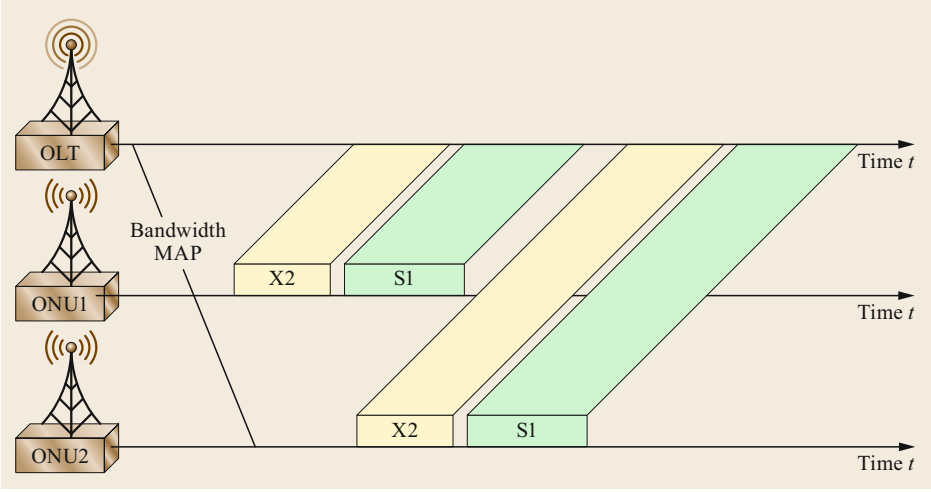


Fig. 16.23 Traditional DBA scheme

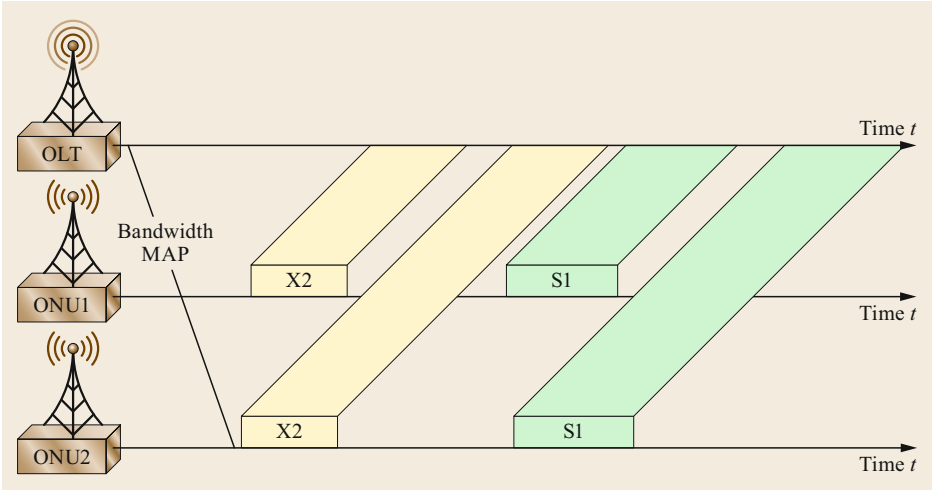


Fig. 16.24 X2 priority DBA scheme

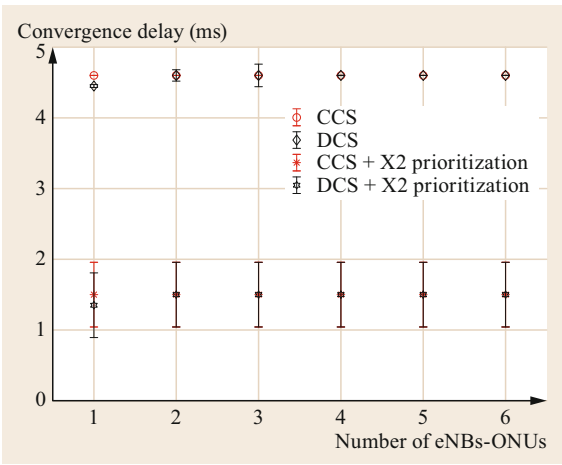


Fig. 16.25 Convergence delay for different CoMP methods

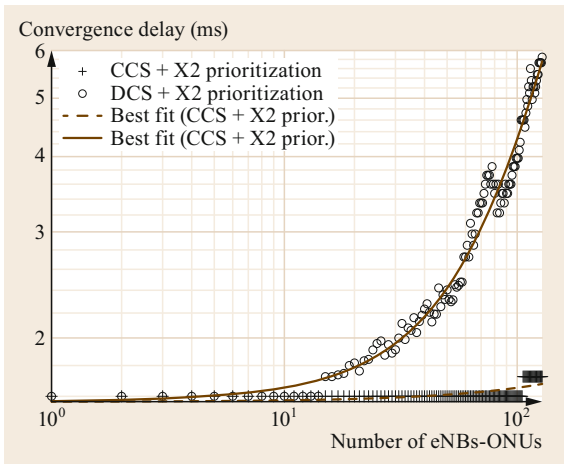


Fig. 16.26 Convergence delay with X2 prioritization (after [16.71])

of strict scheduling when the number of collaborating eNBs is greater than 20.

Figure 16.27 shows the S1 waiting time in the case of CCS and DCS as calculated in (16.7) and (16.8), respectively. Similar to the convergence delay, S1 waiting time is characterized by a linear growth with the number of eNBs-ONUs in the case of CCS while it grows quadratically in the case of DCS. In the case of DCS, the S1 waiting time is higher than 1 ms for a number of cooperating eNBs-ONUs greater than 60. The S1 waiting time represents the performance penalty introduced by X2 prioritization and contributes to the delay experienced by the mobile users. Since this penalty grows with the number of cooperating eNBs, the cluster size can be dimensioned according to the target S1 delay at the user plane, thus implementing an application-centric approach.

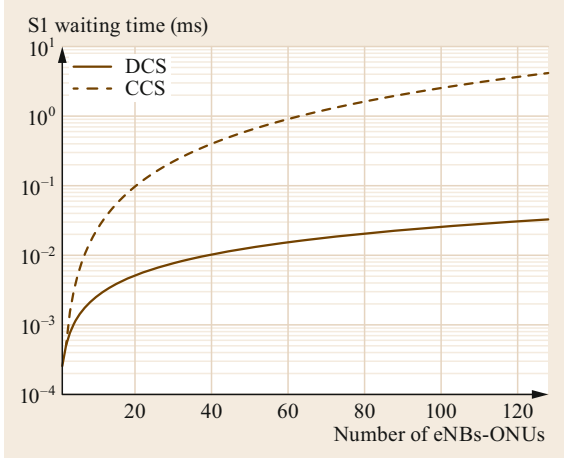


Fig. 16.27 S1 waiting time for CCS and DCS

16.7 Conclusion

This chapter discussed several techniques employed for cross-layer design in optical networks. The first part of the chapter focused on the cross-layer interactions between the physical layer and the network layer. Optically transmitted signals accumulate a variety of physical layer impairments as they traverse fiber spans and cross-connects, and may not be recoverable at the receiver unless appropriate measures are taken. The chapter introduced some of the main PLIs that impact optical signals and the network-level design and resource allocation algorithms to combat and mitigate the PLIs.

The second part of the chapter described application-aware networking paradigms. Firstly, an application-aware metro-access aggregation node architecture was introduced to compensate for the increased delay experienced in the access network (because of sleep mode) by reducing the delay with traffic shaping in the metro network segment. Experimental results showed that the end-to-end latency is preserved by coordinating the optical line terminal (OLT) sleep mode and traffic shaping at the metro aggregation nodes. Secondly, a software-defined networking (SDN) and packet-based 1+1 protection scheme is presented to recover the fronthaul connection carrying a lower layer functional split (i.e., Intra-PHY) without any packet loss in a multilayer Ethernet-over-DWDM metro network. Finally, wireless-access resource scheduling based on a coordinated multipoint (CoMP) technique is described, where dynamic bandwidth scheme (DBA) is implemented that prioritizes X2 traffic over S1 traffic. The results showed that the presented scheme reduces the convergence delay by up to 67%. Moreover, the impact of DBA schemes on centralized coordinated scheduling (CCS) and distributed coordinated scheduling (DCS) is also presented. The results showed that the CCS implementation is preferable with respect to DCS when the number of ONUs is large.

References

16.1 G.P. Agrawal: *Fiber Optic Communication Systems* (Wiley, New York 1992)

16.2 R. Ramaswami, K.N. Sivarajan: *Optical Networks: A Practical Perspective* (Morgan Kaufmann, New York 1998)

16.3 G.P. Agrawal: *Nonlinear Fiber Optics* (Academic Press, San Diego 2013)

16.4 T. Deng, S. Subramaniam, J. Xu: Crosstalk-aware wavelength assignment in dynamic wavelength-routed optical networks. In: *First Int. Conf. Broadband Netw., San Jose* (2004), https://doi.org/10.1109/BROADNETS.2004.26

16.5 M. Batayneh, D.A. Schupke, M. Hoffmann, A. Kirstaedter, B. Mukherjee: Optical network design for a multiline-rate carrier-grade Ethernet under transmission-range constraints, J. Lightwave Technol. **26**(1), 121–130 (2008)

16.6 Y. Pointurier, M. Brandt-Pearce, S. Subramaniam, B. Xu: Cross-layer adaptive routing and wavelength assignment in all-optical networks, IEEE J. Sel. Areas Commun. **26**(6), 32–44 (2008)

16.7 B. Xu, M. Brandt-Pearce: Comparison of FWM- and XPM-induced crosstalk using the Volterra series transfer function method, J. Lightwave Technol. **21**(1), 40–53 (2003)

16.8 Y. Pointurier, M. Brandt-Pearce: Analytical study of crosstalk propagation by fiber nonlinearity in all-optical networks using perturbation theory, J. Lightwave Technol. **23**(12), 4074–4083 (2005)

16.9 Y. Dong, S.-H. Zhao, Y.-H. Ni, H. Hong, X.-F. Tian: Effect of nonlinear phase noise on DQPSK modulation system. In: *Int. Conf. Electron. Commun. Control, Ningbo* (2011), https://doi.org/10.1109/ICECC.2011.6066318

16.10 A. Mafi, S. Raghavan: Nonlinear phase noise in optical communication systems using eigenfunction expansion method, Opt. Eng. **50**, 055003 (2011)

16.11 X. Zhu, S. Kumar: Nonlinear phase noise in coherent optical OFDM transmission systems, Opt. Express **18**(7), 7347–7360 (2010)

16.12 M. Malekiha, S. Kumar: Second-order theory for nonlinear phase noise in coherent fiber-optic system based on phase shift keying. In: *24th Can. Conf. Electr. Comput. Eng., Niagara Falls* (2011), https://doi.org/10.1109/CCECE.2011.6030494

16.13 M. Magarini, A. Spalvieri, F. Vacondio, M. Bertolini, M. Pepe, G. Gavioli: Empirical modeling and simulation of phase noise in long-haul coherent optical transmission systems, Opt. Express **19**(23), 22455–22461 (2011)

16.14 N. Sambo, M. Secondini, F. Cugini, G. Bottari, P. Iovanna, F. Cavaliere, P. Castoldi: Modeling and distributed provisioning in 10–40–100-Gb/s multirate wavelength switched optical networks, J. Lightwave Technol. **29**(9), 1248–1257 (2011)

16.15 B. Ramamurthy, D. Datta, H. Feng, J.P. Heritage, B. Mukherjee: Impact of transmission impairments on the teletraffic performance of wavelength-routed optical networks, J. Lightwave Technol. **17**(10), 1713–1723 (1999)

16.16 A. Mokhtar, M. Azizoglu: Adaptive wavelength routing in all-optical networks, IEEE ACM Trans. Netw. **6**(2), 197–206 (1998)

16.17 E. Karasan, E. Ayanoglu: Effects of wavelength routing and selection algorithms on wavelength con-

version gain in WDM optical networks, IEEE ACM Trans. Netw. **6**(2), 186–196 (1998)
16.18 A. Birman, A. Kershenbaum: Routing and wavelength assignment methods in single-hop all-optical networks with blocking. In: *Proc. IEEE Conf. Comput. Commun., Boston* (1995), https://doi.org/10.1109/INFCOM.1995.515906
16.19 R. Jain, D. Chiu, W. Hawe: A quantitative measure of fairness and discrimination for resource allocation in shared computer systems, DEC Research Report TR-301 (1984)
16.20 J. Sole-Pareta, S. Subramaniam, D. Careglio, S. Spadaro: Cross-layer approaches for planning and operating impairment-aware optical networks, Proc. IEEE **100**(5), 1118–1129 (2012)
16.21 A.G. Rahbar: Review of dynamic impairment-aware routing and wavelength assignment techniques in all-optical wavelength-routed networks, IEEE Commun. Surv. Tutor. **14**(4), 1065–1089 (2012)
16.22 A. Askarian, Y. Zhai, S. Subramaniam, Y. Pointurier, M. Brandt-Pearce: Protection and restoration from link failure in DWDM networks: A cross-layer study. In: *IEEE Int. Conf. Commun., Beijing* (2008), https://doi.org/10.1109/ICC.2008.1021
16.23 D. Zhou, S. Subramaniam: Survivability in optical networks, IEEE Network **14**(6), 16–23 (2000)
16.24 A. Askarian, Y. Zhai, S. Subramaniam, Y. Pointurier, M. Brandt-Pearce: Cross-layer approach to survivable DWDM network design, J. Opt. Commun. Netw. **2**(6), 319–331 (2010)
16.25 M. Medard, R.A. Barry, S.G. Finn, W. He, S. Lumetta: Generalized loop-back recovery in optical mesh networks, IEEE ACM Trans. Netw. **10**(1), 153–164 (2002)
16.26 B. Mukherjee, S. Ferdousi: The network user and its growing influence, Comput. Commun. **131**(3), 43–45 (2018)
16.27 B. Mukherjee, M.F. Habib, F. Dikbiyik: Network adaptability from disaster disruptions and cascading failures, IEEE Commun. Mag. **52**(5), 230–238 (2014)
16.28 F. T. Leighton: Connect to tomorrow, Akamai Edge Conference keynote speech (2017)
16.29 K.J. Kerpez, J.M. Cioffi, G. Ginis, M. Goldburg, S. Galli, P. Silverman: Software-defined access networks, IEEE Commun. Mag. **52**(9), 152–159 (2014)
16.30 T. Zinner, M. Jarschel, A. Blenk, F. Wamser, W. Kellerer: Dynamic application-aware resource management using software-defined networking: Implementation prospects and challenges. In: *IEEE Netw. Oper. Manag. Symp., Krakow* (2014), https://doi.org/10.1109/NOMS.2014.6838404
16.31 K. Kondepu, A. Sgambelluri, L. Valcarenghi, F. Cugini, P. Castoldi: Exploiting SDN for integrating green TWDM-PONS and metro networks preserving end-to-end delay, J. Opt. Commun. Netw. **9**(1), 67–74 (2017)
16.32 I.F. Akyildiz, A. Lee, P. Wang, M. Luo, W. Chou: A roadmap for traffic engineering in SDN-openflow networks, Comput. Netw. **71**, 1–30 (2014)
16.33 A. Lara, A. Kolasani, B. Ramamurthy: Network innovation using openflow: A survey, IEEE Commun. Surv. Tutor. **16**, 493–512 (2014)
16.34 D. Chitimalla, M. Tornatore, S. Lee, H. Lee, S. Park, H.S. Chung, B. Mukherjee: QoE enhancement schemes for video in converged OFDMA wireless networks and EPONs, J. Opt. Commun. Netw. **10**(3), 229–239 (2018)
16.35 N. Chander: *Transforming Mobile Backhaul Networks: Using SDN to Securely Scale Backhaul for LTE and the Next Generation of Mobile Broadband Services* (IDC, Framingham 2014)
16.36 C.J. Bernardos, A. De La Oliva, P. Serrano, A. Banchs, L.M. Contreras, H. Jin, J.C. Zuniga: An architecture for software defined wireless networking, IEEE Wirel. Commun. **21**(3), 52–61 (2014)
16.37 D. Bojic, E. Sasaki, N. Cvijetic, T. Wang, J. Kuno, J. Lessmann, S. Schmid, H. Ishii, S. Nakamura: Advanced wireless and optical technologies for small-cell mobile backhaul with dynamic software-defined management, IEEE Commun. Mag. **51**(9), 86–93 (2013)
16.38 J. Costa-Requena, J.L. Santos, V.F. Guasch: Mobile backhaul transport streamlined through SDN. In: *17th Int. Conf. Transpar. Opt. Netw., Budapest* (2015), https://doi.org/10.1109/ICTON.2015.7193588
16.39 R. Vilalta, R. Muñoz, R. Casellas, R. Martínez, F. Francois, S. Peng, R. Nejabati, D.E. Simeonidou, N. Yoshikane, T. Tsuritani, I. Morita, V. López, T. Szyrkowiec, A. Autenrieth: Network virtualization controller for abstraction and control of openflow-enabled multi-tenant multi-technology transport networks. In: *Opt. Fiber Commun. Conf., Los Angeles* (2015), https://doi.org/10.1364/OFC.2015.Th3J.6
16.40 W. Xia, Y. Wen, C.H. Foh, D. Niyato, H. Xie: A survey on software-defined networking, IEEE Commun. Surv. Tutor. **17**(1), 27–51 (2015)
16.41 P. Parol, M. Pawlowski: Towards networks of the future: SDN paradigm introduction to PON networking for business applications. In: *Fed. Conf. Comput. Sci. Inf. Syst., Krakow* (2013) pp. 829–836
16.42 ITU-T Recommendation Y.1541: Network performance objectives for IP-based services (2011)
16.43 L. Valcarenghi: Where do we stand and where are we heading in making NG-PONs more energy efficient? In: *Int. Conf. Opt. Netw. Des. Model., Pisa* (2015), https://doi.org/10.1109/ONDM.2015.7127300
16.44 S. Kaur, J. Singh, N.S. Ghumman: Network programmability using POX controller. In: *Int. Conf. Commun. Comput. Syst.* (2014) pp. 134–138
16.45 3GPP TR 38.801 V2.0.0: Study on new radio access technology: Radio access architecture and interfaces (2017)
16.46 C. Colman-Meixner, G.B. Figueiredo, M. Fiorani, M. Tornatore, B. Mukherjee: Resilient cloud network mapping with virtualized BBU placement for cloud-RAN. In: *IEEE Int. Conf. Adv. Netw. Telecommun. Syst., Bangalore* (2016), https://doi.org/10.1109/ANTS.2016.7947790
16.47 L. Valcarenghi, F. Cugini, F. Paolucci, P. Castoldi: Quality-of-service-aware fault tolerance for grid-

enabled applications, Opt. Switch. Netw. **5**(2–3), 150–158 (2008)
16.48 M.R. Rahman, R. Boutaba: SVNE: Survivable virtual network embedding algorithms for network virtualization, IEEE Trans. Netw. Serv. Manag. **10**(2), 105–118 (2013)
16.49 F. Gu, H. Alazemi, A. Rayes, N. Ghani: Survivable cloud networking services. In: *Int. Conf. Comput. Netw. Commun., San Diego* (2013), https://doi.org/10.1109/ICCNC.2013.6504230
16.50 S. Ramanathan, M. Tacca, M. Razo, B. Mirkhanzadeh, K. Kondepu, F. Giannone, L. Valcarenghi, A. Fumagalli: A programmable optical network testbed in support of C-RAN: A reliability study, Photonic Netw. Commun. (2019), https://doi.org/10.1007/s11107-018-00825-9
16.51 A. Fumagalli, L. Valcarenghi: IP restoration vs. WDM protection: Is there an optimal choice?, IEEE Network **14**(6), 34–41 (2000)
16.52 K. Kondepu, A. Sgambelluri, N. Sambo, F. Giannone, P. Castoldi, L. Valcarenghi: Orchestrating lightpath recovery and flexible functional split to preserve virtualized RAN connectivity, J. Opt. Commun. Netw. **10**(11), 843–851 (2018)
16.53 Linux Foundation: Open vswitch documentation, http://openvswitch.org/support/ (2016)
16.54 Open Network Foundation: OpenFlow switch specification, http://www.opennetworking.org/wp-content/uploads/2013/04/openflow-spec-v1.3.1.pdf (2013)
16.55 H. Welte, P.N. Ayuso: Netfilter.org project, http://www.netfilter.org (2014)
16.56 D. Hicks, C. Malina-Maxwell, M. Razo, M. Tacca, A. Fumagalli, D. Nguyen: PROnet: A programmable optical network prototype. In: *18th Int. Conf. Transpar. Opt. Netw., Trento* (2016), https://doi.org/10.1109/ICTON.2016.7550421
16.57 OpenAirInterface Wiki: https://gitlab.eurecom.fr/oai/openairinterface5g/wikis/home
16.58 B. Mirkhanzadeh, A. Shakeri, C. Shao, M. Razo, M. Tacca, G.M. Galimberti, G. Martinelli, M. Cardani, A. Fumagalli: An SDN-enabled multi-layer protection and restoration mechanism, Opt. Switch. Netw. **30**, 23–32 (2018)
16.59 T. Chen, M. Matinmikko, X. Chen, X. Zhou, P. Ahokangas: Software defined mobile networks: Concept, survey, and research directions, IEEE Commun. Mag. **53**(11), 126–133 (2015)
16.60 J. Elbers, K. Grobe, A. Magee: Software-defined access networks. In: *Eur. Conf. Opt. Commun., Cannes* (2014), https://doi.org/10.1109/ECOC.2014.6964232
16.61 NGMN Alliance: *RAN evolution project CoMP evaluation and enhancement* (NGMN, Frankfurt 2015)
16.62 A. Marotta, K. Kondepu, F. Giannone, S. Doddikrinda, D. Cassioli, C. Antonelli, L. Valcarenghi, P. Castoldi: Performance evaluation of CoMP coordinated scheduling over different backhaul infrastructures: A real use case scenario. In: *Int. Conf. Sci. Electr. Eng., Eilat* (2016), https://doi.org/10.1109/ICSEE.2016.7806105
16.63 S. Kuwano, J. Terada, N. Yoshimoto: Operator perspective on next-generation optical access for future radio access. In: *IEEE Int. Conf. Commun. Worksh., Sydney* (2014), https://doi.org/10.1109/ICCW.2014.6881226
16.64 T. Tashiro, S. Kuwano, J. Terada, T. Kawamura, N. Tanaka, S. Shigematsu, N. Yoshimoto: A novel DBA scheme for TDM-PON based mobile fronthaul. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.Tu3F.3
16.65 D. Iida, S. Kuwano, J. Kani, J. Terada: Dynamic TWDM-PON for mobile radio access networks, Opt. Express **21**(22), 26209–26218 (2013)
16.66 3GPP TR 36.819 V 11.2.0: Coordinated multi-point operation for LTE physical layer aspects (2013)
16.67 G.Y. Li, J. Niu, D. Lee, J. Fan, Y. Fu: Multi-cell coordinated scheduling and MIMO in LTE, IEEE Commun. Surv. Tutor. **16**(2), 761–775 (2014)
16.68 F. Musumeci, E. De Silva, M. Tornatore: Enhancing RAN throughput by optimized coMP controller placement in optical metro networks, IEEE J. Sel. Areas Commun. **36**(11), 2561–2569 (2018)
16.69 A. Marotta, K. Kondepu, F. Giannone, D. Cassioli, C. Antonelli, L. Valcarenghi, P. Castoldi: Impact of CoMP VNF placement on 5G coordinated scheduling performance. In: *Eur. Conf. Netw. Commun.* (2017), https://doi.org/10.1109/EuCNC.2017.7980776
16.70 3GPP TS 36.423 V 14.2.0: Technical specification group radio access network: Evolved universal terrestrial radio access network (E-UTRAN); X2 application protocol (X2AP)(release 14) (2017)
16.71 A. Marotta, F. Giannone, K. Kondepu, D. Cassioli, C. Antonelli, L. Valcarenghi, P. Castoldi: Reducing CoMP control message delay in PON backhauled 5G networks. In: *23th Eur. Wirel. Conf.* (2017) pp. 1–5
16.72 3GPP TS 36.213: Evolved universal terrestrial radio access (E-UTRA): Physical layer procedures (release 13) (2016)
16.73 L. Valcarenghi, M. Chincoli, P. Monti, L. Wosinska, P. Castoldi: Energy efficient PONs with service delay guarantees. In: *Sust. Internet ICT Sustain., Pisa* (2012)
16.74 G. Piro, N. Baldo, M. Miozzo: An LTE module for the NS-3 network simulator. In: *4th Int. ICST Conf. Simul. Tools Tech., Brussels* (2011) pp. 415–422
16.75 X. Wu, K. Brown, C. Sreenan, P. Alvarez, M. Ruffini, N. Marchetti, D.B. Payne, L. Doyle: An XG-PON module for the NS-3 network simulator. In: *Worksh. NS-3* (2013), https://doi.org/10.4108/simutools.2013.251605

Suresh Subramaniam

George Washington University
Washington, USA
suresh@gwu.edu

Suresh Subramaniam received the Ph.D. degree in electrical engineering from the University of Washington, Seattle, in 1997. He is a Professor of Electrical and Computer Engineering at the George Washington University, Washington DC. His research interests encompass several aspects of networking. He is a Fellow of the IEEE.

Koteswararao Kondepu

Scuola Superiore Sant'Anna
Pisa, Italy
koteswararao.kondepu@sssup.it

Koteswararao Kondepu is currently working as a Research Fellow at Scuola Superiore Sant'Anna, Pisa, Italy. He obtained his PhD in Computer Science and Engineering from the Institute for Advanced Studies Lucca (IMT), Italy, in July 2012. His research interests are 5G, cloud-native mobile systems, optical networks design, wired-wireless access convergence, communication networks reliability, and energy efficient schemes in access networks.

Andrea Marotta

DISIM
University of L'Aquila
L'Aquila, Italy
andrea.marotta@univaq.it

Andrea Marotta is currently working as a Research Fellow at the University of L'Aquila, L'Aquila, Italy. He obtained his PhD in Information and Communications Technology from the University of L'Aquila. He has been a visiting researcher at University of California, Davis. His research interests are 5G, network slicing, network reliability, software-defined access networks, and cooperative multipoint transmissions.

17. Optical Network Virtualization

Jie Zhang, Ricard Vilalta, Xiaosong Yu, Victor Lopez, Alejandro Aguado Martín

Optical network virtualization presents the necessary technologies to provide the specified set of network requirements, while providing the necessary isolation between network slices. In this chapter, technologies from both data and control plane perspectives are presented for virtual optical networks (VON). Firstly, the benefits provided by adopting software-defined networks (SDN) and network function virtualization (NFV) technologies are introduced. Secondly, the need for a network operating system is discussed, and a reference architecture for optical networks is presented. Thirdly, a methodology for describing virtual and physical network resources is provided. Fourthly, a review of different VON resource allocation algorithms is presented. Finally, an NFV architecture supporting VON is presented, among several use cases.

17.1 Optical Networks

As the cloud paradigm has profoundly impacted on the IT and software community, NFV and SDN approaches are expected to profoundly redefine the telecommunications industry. The networking industry needs to undergo profound and extreme changes in order to exploit virtualization technologies and software innovations in a truly effective manner. The introduction of these t echniques and software innovations is known as network softwarization or network programmability.

Network operators and service providers expect NFV and SDN to reduce both the capacity expenditure as well as the operations expenditure, taking into consideration the complexity and costs of deploying and managing heterogeneous networks and services. This promise is still unfulfilled, as currently network-wide abstractions to lay the foundation for true network programmability are still lacking.

The objective of NFV is to use IT virtualization techniques to provide entire classes of virtualized network node functionalities. A virtualized network function (VNF) is defined as a network function that was typically provided through custom hardware appliances, which runs as software on a single or several hosts, typically inside virtual machines (VMs) [17.1]. Examples of VNFs are authentication, authorization and accounting (AAA), load balancers, firewalls, or security network functions. Any data plane packet processing and control plane function is expected to be

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_17

adapted to NFV in fixed and mobile network infrastructures [17.1].

On the other hand, SDN improves network programmability and dynamic adjustment of network resources. The SDN controller is the center of SDN architecture and furnishes a logical centralized architecture that provides an abstraction of network resources through application programmable interfaces (APIs). This architecture ensures the ability to perform programmable control and management tasks of different network forwarding technologies, including packet/flow switches, circuit switching, and optical wavelength switched transport technologies. A single controller or a hierarchy/peer of SDN controllers can coordinate the multilayer architecture [17.2] with upper-layer applications.

This introduction will provide some basic concepts and definitions on NFV and SDN to complement the reader's knowledge on the topics in order to later address more advanced discussions fluently. First the concept of a network operating system will be introduced. Later, an introduction to optical network virtualization data plane techniques will be provided. Virtual optical network resource allocation algorithms will be discussed in the next section. Finally, optical network functions virtualization will be presented.

17.1.1 SDN

OpenFlow is the flagship protocol for SDN. OpenFlow is an open standard that enables researchers to run experimental protocols in ethernet networks. OpenFlow is added as a feature to commercial Ethernet switches, routers, and wireless access points, and provides a standardized hook to allow researchers to run experiments, without requiring vendors to expose the internal workings of their network devices [17.3]. OpenFlow enables the determination of the path of network L2 frames by running software on a separate server. In SDN, network intelligence and state are logically centralized, and the underlying network infrastructure is abstracted from the applications.

The OpenFlow protocol abstracts data plane switches as a flow table, in which a flow is defined as any combination of L2, L3, and L4 packet headers, providing the control plane with a common hardware abstraction through a standardized open interface (i.e., OpenFlow protocol) to program the flow table.

In general, OpenFlow assumes a logically centralized control plane, with an OpenFlow controller running a network-wide operating system (e.g., OpenDayLight and the Open Network Operating System (ONOS) are well-known SDN controllers). The network-wide operating system is usually composed of a single software controller, and a single network view database, that is kept in a database.

An SDN controller's network view can store the switch-level topology, the locations of users, hosts, middleboxes and other network elements, and the services (e.g., HTTP or NFS (network file system) being offered). The network view may also include all bindings between names and addresses.

Different applications can run on top of the SDN controller that uses this centralized network view to make control decisions, such as deciding how each flow is routed (i.e., routing), which flows are admitted (i.e., control access), where they are replicated (i.e., multicast), or configuring the switches in a vertical manner. Such a network-wide operating system allows these management applications to be written in as centralized programs. That is, they are written as if the entire network were present in a single machine. To this end, SDN controllers offer an API used by the management applications to create new functionalities and services to the network.

Since many vendors are yet to embrace OpenFlow, an in-house built OpenFlow agent placed on the node is typically used to provide OpenFlow abstractions. A modular OpenFlow agent might provide hardware abstractions to the controller via an OpenFlow interface. This agent utilizes the network equipment management interface (e.g., SNMP (simple network management protocol), TL-1) to communicate with the data plane. This can be understood as a transitional solution before OpenFlow protocol or other standard is implemented at network element level.

Optical OpenFlow

The OpenFlow architecture has also been proposed to control multilayer networks in a unified way [17.4]. An experimental OpenFlow unified control plane was proposed for packet and circuit switched networks. A simple proof-of-concept testbed was demonstrated, where a bidirectional wavelength circuit is dynamically created to transport a TCP flow. In [17.2], a unified control plane for packet, fixed, and flexible dense wavelength division multiplexing (DWDM) grid technologies proposed protocol extensions for different optical networks. In 2015, the ONF approved optical OpenFlow extensions standard [17.4].

Transport NBI

Network operators have been requesting multivendor interoperability in transport networks. Interoperability demonstrations help to align the proposed solutions of equipment manufacturers (vendors), while creating a competitive environment that enables innovation and provides new developments and product evolution.

Currently, network operators deploy optical networks on a regional basis. This means that each vendor uses their technology in a given area and operates it, which is very relevant for production environments. The use of the transport application programming interface (TAPI) as a northbound interface (NBI) of optical controllers allows the utilization of a common information model to support the optical services.

The definition of the TAPI within the Open Networking Foundation (ONF) is a pragmatic approach to obtain an end-to-end (E2E) software-defined network (SDN) infrastructure for transport networks. It supports network programmability in each optical domain without requiring full data plane interoperability. Multilayer/domain SDN will enable many innovative use cases, which will lead to new service product offerings and business opportunities like multitenancy, bandwidth on demand, and data center workload balancing. However, network operators also plan to automate today's internal operational processes like multilayer network optimization, multidomain link configuration, and restoration through SDN. This work has been made within the umbrella of the Optical Internetworking Forum (OIF) to coordinate the industry in this interworking activity.

17.1.2 ETSI NFV MANO Framework

The ETSI NFV ISG started developing a specification document that was released in 2014 with the aim of describing the main components and interfaces of an architectural framework for management an orchestration of NFV services. This specification document provides a common guideline for the correct operation of multivendor NFV solutions. Its ultimate goal is to provide a set of common interfaces and information models, and the mappings to nonstandard ones, in order to facilitate the automation of NFV service deployment in heterogeneous environments. This will reduce the complexity of integrating new components to the infrastructure by providing a consistent representation of the infrastructure and service management. Such integration can come either from southbound interface infrastructures (new computing and storage resources/architectures), from east–west systems such as operations support software (OSS) and business support software (BSS), or from northbound applications.

The NFV management and orchestration (MANO) architectural framework [17.5] is divided into three fundamental layers: the NFV orchestrator (NFVO), the VNF manager (VNFM), and the virtualized infrastructure manager (VIM). Later, we will describe each of these in detail.

NFV Orchestrator (NFVO)

The NFVO is responsible for the life cycle of NFV services, commonly distributed across the whole network infrastructure. Different VIMs, which can be physically distributed in different points-of-presence (PoPs), are orchestrated by the NFVO. It hosts a catalogue of the NFV images available for its use, organizing the deployment of each of them to finally create a service graph or service function chain (SFC). The NFVO also controls the access and authentication of users, creating a map of policies that are reinforced for scaling up and down the service graph. For this purpose, it works in collaboration with the VNFMs to monitor the load of each VNF and make sure that the current deployment can cope with the network service demands. Finally, it takes care of resource management, taking in account the infrastructure available and the VNFs already deployed.

VNF Manager (VNFM)

The VNFM is responsible for controlling the life cycle of one or multiple VNF instances. Most life cycle operations are assumed to be generic and shared among different VNFs, but the VNFM should also be able to control specific life cycle requirements associated to specific instances (e.g., how and why to scale up or down, other policies associated to an VNF or to an entire service). Among others, the main responsibilities of a VNFM include:

- Instantiation of the VNF, including initial configuration and/or template
- Active monitoring and performance measurements
- Scale operations up and down
- Any modification or update
- Coordination with the associated VIM
- Communication with the NFVO for policy checks and scaling operations
- Termination.

Infrastructure Manager (VIM)

The VIM is in charge of controlling the underlying network, and the computing and storage infrastructure on which the NFV services are built. A single VIM could potentially control all the resources on an operator's infrastructure, be distributed for each PoP, or even subdivide the resources within a PoP to be controlled by different VIMs. The VIM should be able to allocate the resources during the deployment of a VNF instance, while supporting the creation of the entire virtual services by creating the required virtual links, networks, or the security policies to isolate traffic and access for the different tenants.

17.2 Network Operating System

The concept of an operating system (OS), a system software that manages hardware and software resources and provides common services for computer applications, has inspired the concept of a network operating system (NetOS). A NetOS will manage network hardware and software resources, and allow network-enabled applications. It encourages the research and industrial communities to design, build, and deliver a running and stable set of novel network-enabled applications. As an initial effort toward a NetOS, the model that it is presented here is based on the work done in [17.6].

17.2.1 NetOS Concept

The NetOS concept is composed of three main components: i) drivers and devices, ii) the NetOS kernel, and iii) the user space. The proposed reference model is depicted in Fig. 17.1, which shows its main building blocks. For comparison, a general architecture of a UNIX operating system is presented in Fig. 17.2.

In the drivers and devices component, the network abstraction layer (NAL) is the element responsible for providing a unified southbound interface (SBI) towards the physical infrastructure, while abstracting its characteristics. That is, the NAL exposes the different network elements (NEs) to the NetOS kernel using drivers, thus enabling access to each specific element in a common way, hiding the complexity of the NEs to the upper layers.

The NetOS kernel deals with the essential coordination functions, including common models for network resource virtualization, while providing uniform mechanisms in key aspects, such as security, tenant separation, namespaces and location, and resource lifetime management. Moreover, the NetOS kernel maintains the state of the network by interfacing with the network elements through the NAL. Finally, the virtual network layer enables the deployment, control, and management of virtual networks over many possible underlying network technologies.

Finally, the user space simplifies the integration with operations support systems (OSS), business support systems (BSS), and controller frameworks. Moreover, through the definition of layered interfaces between the NetOS kernel and libraries, an application execution environment enables running network-wise apps.

The general NetOS reference model can be compared with a UNIX system reference model (Fig. 17.2). In this case, it is clear that in the NetOS case, the hardware is composed of groups of network devices. The controllers in the UNIX system correspond to the southbound interfaces or network drivers (e.g., OpenFlow), or even a control plane. The kernel interface to the hardware offers homogeneous access to network resources, which is basically the function of the network abstraction layer (NAL) in the NetOS case. Likewise, the NetOS kernel carries out the functions that the kernel

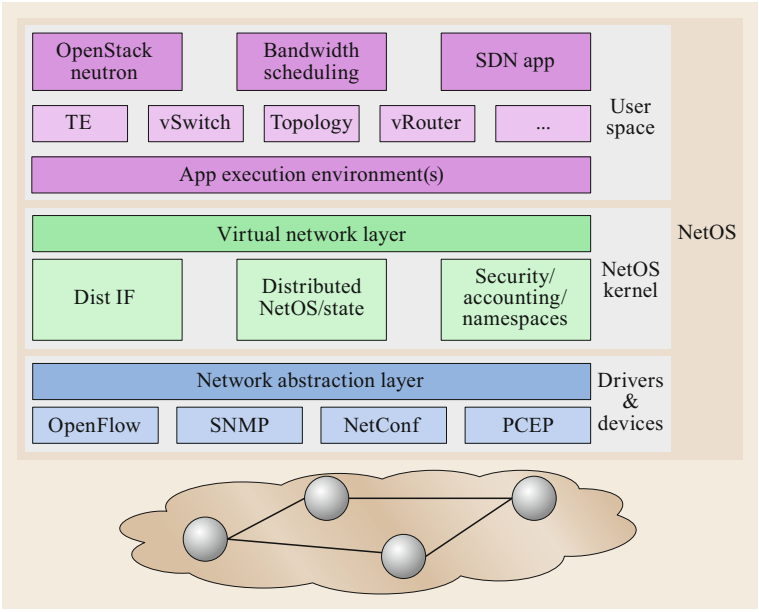


Fig. 17.1 Proposed layered structure of NetOS

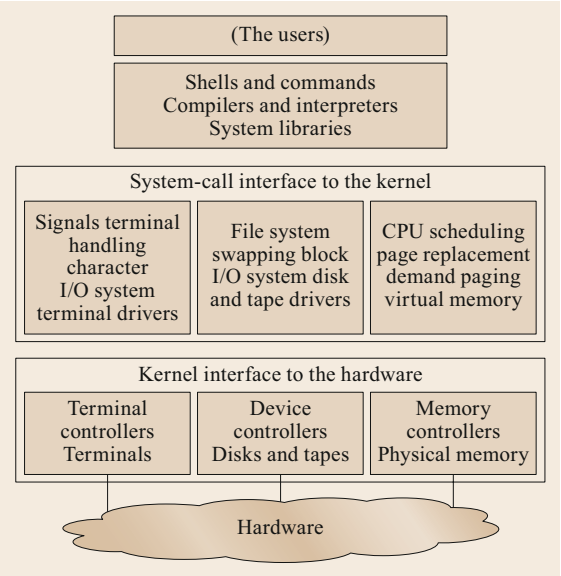


Fig. 17.2 UNIX System 1

does in a UNIX environment. For instance, it controls the daemons or processes that enable a broad spectrum of functions, such as configuration, session management, command queuing and scheduling, rollbacks, etc. These are core network applications that are embedded in the NetOS. Moreover, the virtual network layer offers a common method to access the kernel. Observe that the system libraries in the NetOS model are generic applications that can be used by multiple user applications. Some examples of these libraries are topology, vRouter, vSwitch, or traffic engineering (TE). Finally, the user applications that utilize the NetOS environment are specialized applications for business, like operation processes, operation styles and policies, right delegation, etc.

17.2.2 NetOS in the Scope of NFV

One of the main benefits of NFV is the increase of the flexibility and support for the launch and management of novel network services, while the operational cost for network operators is reduced. Network functions virtualization (NFV) is hosted at ETSI in the form of an ISG (Industry Specification Group), and it involves leading network operators around the world.

The relationship between NetOS and the ETSI NFV architecture is shown in Fig. 17.3. It can be observed that NetOS provides functionality as an NFV orchestrator (NFV-O), as well as an VNF manager (VNF-M) and virtualized infrastructure manager (VIM). Some of the proposed services that are being considered as possible NFV use cases to be run on NFV-enabled

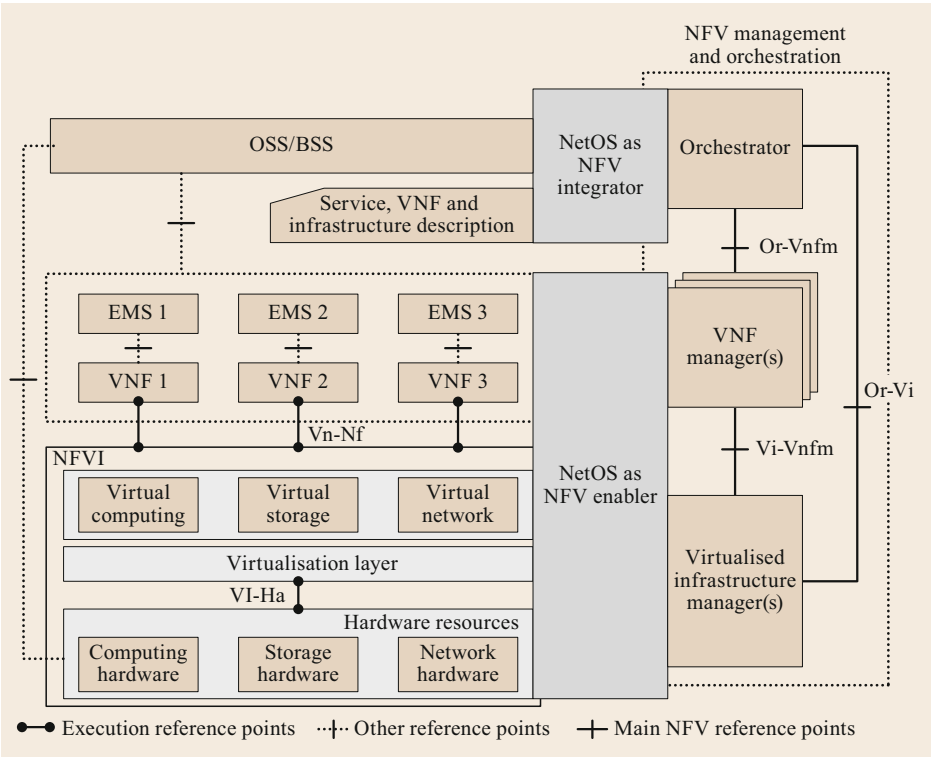


Fig. 17.3 NetOS in the scope of the ETSI NFV reference architectural model (after [17.5])

platforms cover practically all network segments and technologies: evolved packet core (EPC), broadband remote-access server (BRAS), content delivery network (CDN), or deep packet inspection (DPI).

It is important to note that virtual network functions have a complex dynamic life cycle and can scale in-/out on demand to rapidly adapt to changing conditions and load characteristics. NetOs abstractions simplify the management and orchestration of such services and their integration with nonvirtualized services.

17.2.3 Programmable Interfaces in Optical Networks

The industry and the research community have devoted significant efforts to designing tools that can autonomously configure and operate a variety of hardware appliances and network elements (NEs). With the objective of providing a unified view of the network elements, there have been multiple attempts to facilitate the configuration processes. These attempts have been located from element management systems (EMSs) that consider specific interfaces to the devices, up to the control plane, where several abstractions on services provided have been tackled.

Most current implementations of SDN for optical elements are either based on vendor-specific extensions or on an abstracted view of the optical layer. Typically, proprietary optical SDN controllers interact with vendor-specific network elements. These proprietary controllers have started to provide common northbound interfaces (NBI) to provide topological and provisioning services.

The IT and software communities have adopted the driver concept. However, the development of hybrid IT and networking fabrics is something that has not been done before. To this end, a NetOS requires defining the information models, protocols, and semantics in the southbound interfaces (SBI) for the plug-and-play of NEs.

In the following sections, the various necessary, commonly-accepted SBIs are introduced, for both topology recovery interfaces and for provisioning interfaces.

Topology Interfaces

BGP-LS. BGP-LS is an extension to Border Gateway Protocol (BGP) for distributing the link state (LS) of the network topology model to other network elements, such as the SDN controller. Typically, a node (e.g., a router) keeps at least one database for storing this information in a given network domain (based on the abstraction of nodes and links). If BGP is enabled, this node can collect this information and distribute it to its peers using the defined link state attributes.

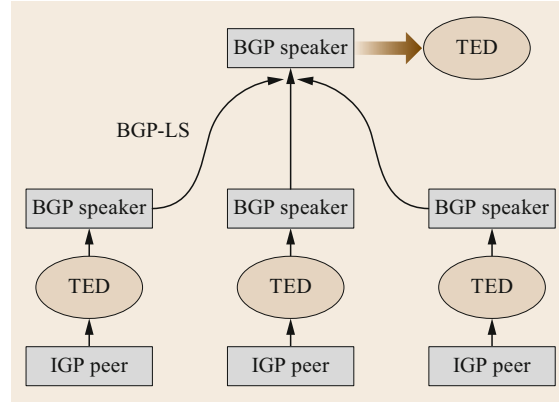


Fig. 17.4 TE distribution using BGP-LS

In Fig. 17.4, the BGP peers exporting the topology can represent child SDN controllers belonging to different domains, while the one importing the information can be the parent SDN controller learning about the network. In this model, a child SDN controller would provide enhanced computational power to enable optimal paths through its domain. This is possible due to BGP extending its applicability to carry not only reachability data but also LS information.

New BGP network layer reachability information (NLRI) and the corresponding link state attribute (BGP-LS attribute) are defined as extensions to support the optical domain description.

REST API. A representational state transfer (REST) API is an HTTP API that follows an architectural style defined by a set of constraints to be used for creating web services. There are several REST interfaces, one for each controller. Several controllers that can be considered are Floodlight, ONOS, and OpenDayLight (ODL). An SDN controller is typically able to run OpenFlow protocol and allows the development of applications built on top of it. The REST APIs made available by all running modules in the SDN controller can be accessed through the specified REST port (8080). Any client REST applications can now obtain information and invoke services by send-

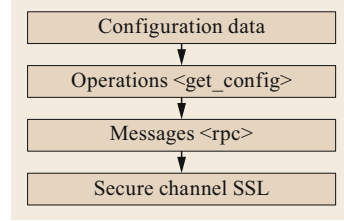


Fig. 17.5 NETCONF model

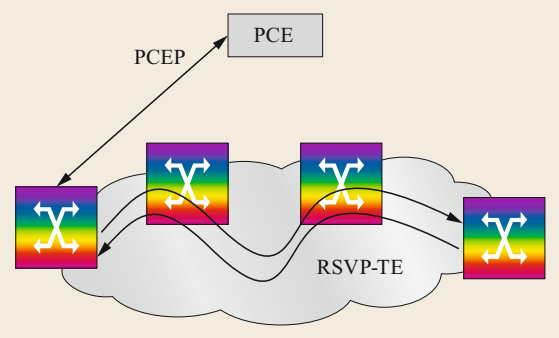


Fig. 17.6 PCEP to remotely initiate GMPLS LSP

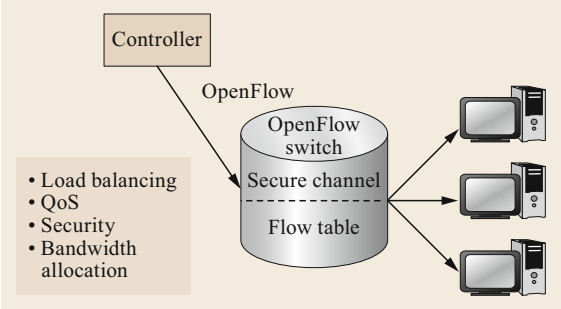


Fig. 17.7 OpenFlow scenario

ing HTTP commands (i.e., GET, PUT, POST, DELETE) to the controller REST port.

The topology service is the module in charge of computing topologies based on the link information it learns from the link discovery service, which is typically based on the Link Layer Discovery Protocol (LLDP). The topology service maintains a record of SDN control domains. A control domain is defined as a union of joined SDN network elements under the same instance of an SDN controller. Control domains can be interconnected using non-SDN switches in the same layer-2 domain.

Provisioning Interfaces

NETCONF. In Fig. 17.5, we can see the basic network configuration protocol (NETCONF) model that is used as a control interface (either for topology or provisioning). The operations layer defines the set of possible actions that NETCONF supports. These actions are requested as RPC (remote procedure call) methods with XML-encoded parameters. The framing and encoding the RPC objects is provided by the message layer. Finally, a dedicated channel is used to send the encoded message. NETCONF supports multiple transport protocol but is typically deployed with SSH. This protocol allows the client to dynamically discover the capabilities of the server and use them to adjust its behavior. Capabilities extend the basic operations of the device, describing additional operations and the content included in these operations.

PCEP. The extensions described in the Internet Engineering Task Force (IETF) draft [17.7] offer the possibility of using the Path Computation Element Protocol (PCEP) to set up generalized multi-protocol label switching (GMPLS) label-switched paths (LSP) remotely. As shown in Fig. 17.6, the idea behind it is to enable an active PCE to set up GMPLS LSPs.

However, these extensions can be used to request a path setup from any element in the network. The main advantage of PCEP is that the protocol has a rich syntax to express path details, and it is suited for any network technology. That is, the data model is standard and well defined, making PCEP protocol a good southbound candidate.

OpenFlow. As was previously introduced, OpenFlow provides a novel way of implementing and managing programmable networks [17.3]. OpenFlow presents open and well-defined data model that makes it possible for any developer to implement customized network functionalities.

The most common OpenFlow scenario consists of a logically centralized database containing up-to-date network information (Fig. 17.7). The controller is capable of interpreting this information and providing the switches with the optimum routing table for each data flow. This information can also be updated on demand if the network conditions fluctuate.

REST API. As was previously explained, a REST API allows the usage of semantical HTTP requests and responses in combination with JavaScript object notation (JSON) objects. JSON is a file format that uses text to transmit data objects consisting of attribute–value pairs and array data types. Using JSON over HTTP provides a highly efficient, low-cost mechanism to provision interfaces. The only requirement is to have a server listening in port 80 of the corresponding network element. By decoding the received JSON object fields, the

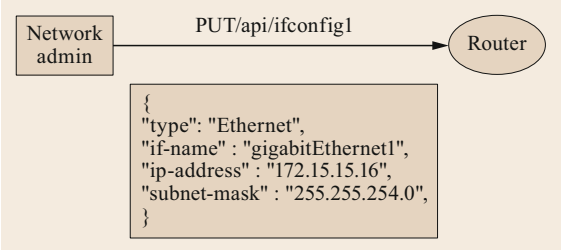


Fig. 17.8 JSON configuration example

necessary information to modify the configuration accordingly is obtained.

In Fig. 17.8, we can see a conceptual example of how we can use JSON with REST. The network administrator sends a post HTTP message containing a JSON field with the required configuration, and the destination router decodes the received message obtaining the new configuration. No session is needed, and every action can be performed asynchronously with the consequent reduction in complexity.

17.3 Network Resources Virtualization in Optical Networks

Optical network virtualization refers to the ability to create logical/virtual optical networks that are decoupled from underlying optical network hardware. To support logical optical networks, network hardware devices should be able to be virtualized for provisioning logical optical networks. There are several kinds of optical networks according to the multiplexing techniques used in fibers, i.e., wavelength-division multiplexing (WDM), elastic optical networks (EON), and space–domain multiplexing (SDM). Since network virtualization is not limited to any multiplexing techniques, we will use spectrum channels as an example, which can be implemented as wavelengths or subcarriers in fiber links. This section introduces optical network virtualization from viewpoint of the following three aspects: physical substrate virtualization, virtual resource management, and virtual networks provisioning.

17.3.1 Optical Network Substrate Virtualization

From a hardware perspective, optical networks are composed of optical switches and fiber links. To support network virtualization, optical network resources should be provisioned in a flexible manner. Abstraction is a key concept for optical network resource provisioning, and it can hide the underlying characteristics of network substrates from the major resources for virtual optical network provisioning. This section introduces two abstraction methods for optical-switch and link virtualization: participation and aggregation [17.8]. Using participation, one optical switch or link is sliced into multiple pieces for different virtual networks, and using aggregation, multiple resource pieces are integrated into one bigger virtual entity.

Optical-Switch Virtualization

Optical cross connect (OXC) and the reconfigurable optical add–drop multiplexer (ROADM) are two kinds of optical switches; they can switch optical signals from an input port to an output port. Inside an OXC or ROADM, spectrum channels are the major resources; they provide a natural dimension for partition and aggregation. Input and output ports other kinds of resources for optical switches, and the port is another dimension for switch virtualization. As is illustrated in Fig. 17.9, by partition, one entire optical switch can be sliced into multiple small virtual optical switches, and by aggregation, multiple optical switches can be represented as a super switch.

Partition. An optical switch can be sliced at either the port or the spectrum-channel dimension, or both dimensions, and a virtual switch can have parts of ports and spectrum channels from a physical switch. For a physical optical switch, which has K input/output ports and support spectrum channels $[1, N]$, it can support virtualized switches with k input/output ports and spectrum channels $[i, j]$, where $1 \leq k \leq K$ and $1 \leq i \leq j \leq N$. Figure 17.10 shows two virtualized optical switches sliced from one physical switch. Since spectrum channels are multiplexed in input/output ports, switch ports can be reused when a physical switch is sliced in the spectrum-channel dimension. However, one spectrum channel in a port cannot be reused by two or multiple virtual optical switches.

Aggregation. Similar to partition, aggregation can also be performed in ports and spectrum-channel dimensions. In the port dimension, multiple physical nonblocking switches, which can provide switching capa-

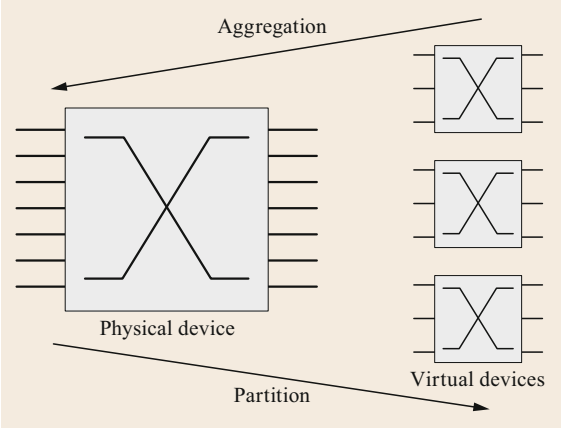


Fig. 17.9 Partition and aggregation

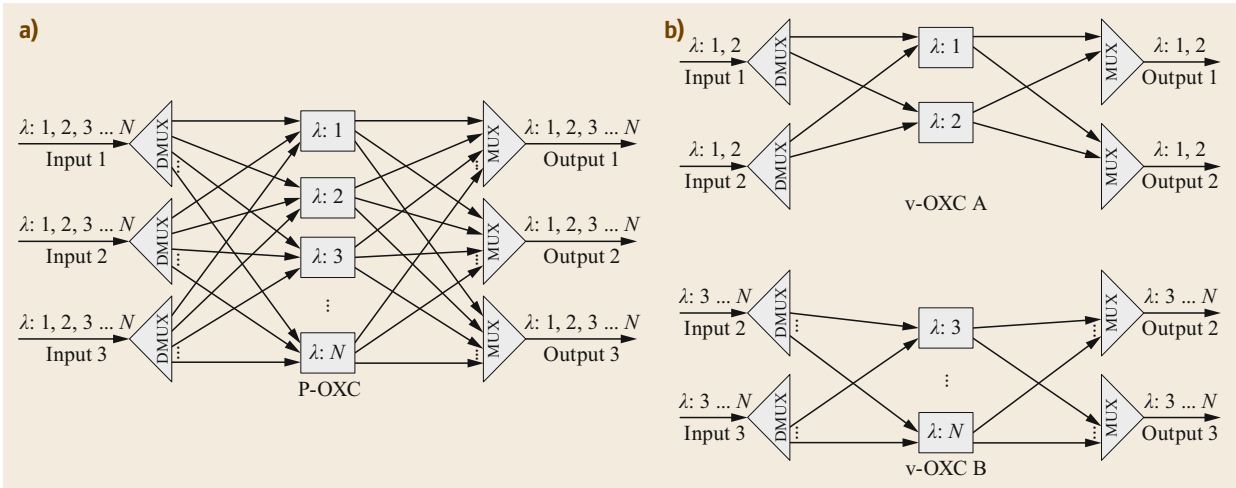


Fig. 17.10a,b Partition of optical switches **(a)** before partition; **(b)** after partition

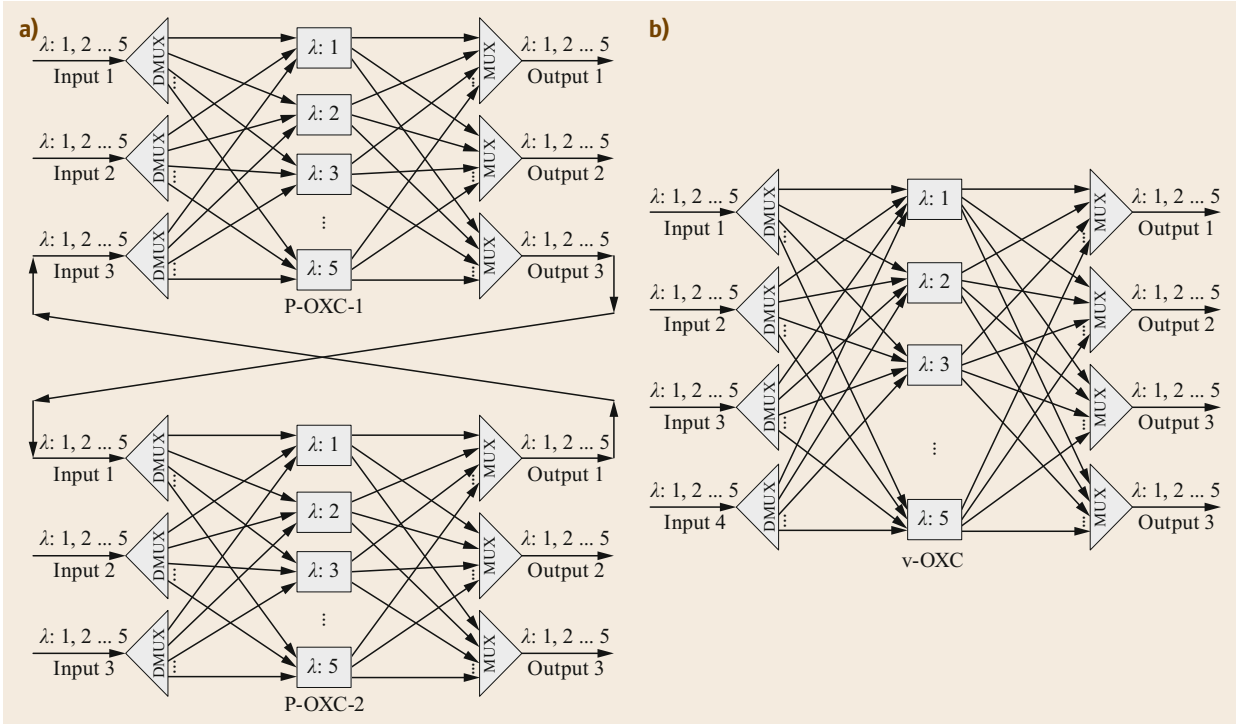


Fig. 17.11a,b Aggregation of optical switches in the port dimension **(a)** before aggregation; **(b)** after aggregation

bilities at the same spectrum channels, can be combined as a super nonblocking switch with the same spectrum channels available. Figure 17.11 shows an example of aggregating two three-input/output port switches, and the aggregated virtual switch has four input/output ports. To provide nonblocking switching capabilities, each physical switch needs to provide at least one extra input/output port to connect one peer. Given k physical switches $S_1, S_2 \ldots S_3$, whose number of input/output ports are $P_1, P_2 \ldots P_3$, the number of aggregated switches is calculated as $P_v = \sum_{\chi=1}^{k} P_\chi - 2(k-1)$. Note that connections for aggregation may go through other physical switches. In the spectrum-channel dimension, multiple physical optical switches, which can provide

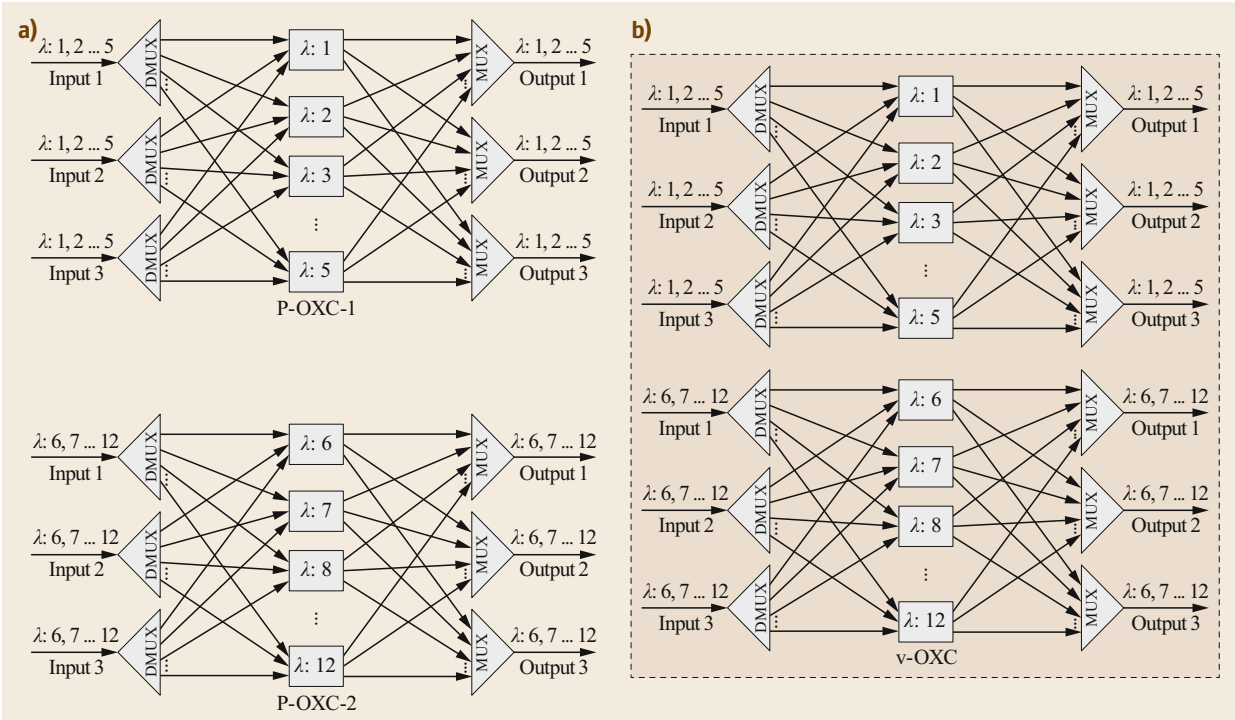


Fig. 17.12a,b Aggregation of optical switches in the spectrum-channel dimension **(a)** before aggregation; **(b)** after aggregation

switching capabilities for different spectrum channels, cannot be aggregated as a nonblocking super switch due to the physical isolation of ports on different switches. However, they can still be aggregated as a super switch with blocking restrictions. Figure 17.12 shows an example of aggregation in the spectrum-channel dimension. No physical connections are required for this kind of aggregation. However, signals from the input ports of one physical switch cannot be routed to the output ports of one another physical switch.

Optical Link Virtualization

Regarding optical link virtualization, the spectrum in optical fibers is the major resource that needs to be virtualized, and wavelength division provides a natural dimension for optical link virtualization.

Partition. In spectrum-channel dimension, one optical link can be sliced into multiple virtual links, and one virtual optical link can have parts of the spectrum channels from a physical link. Given a physical link with K spectrum channels, the number of spectrums of each virtual link k cannot be larger than K, i.e., $k < K$. Figure 17.13 shows an example of optical link partition. Note that one spectrum channel cannot be sliced into two or more virtual links.

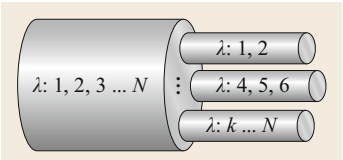


Fig. 17.13 Partition of optical links in the spectrum-channel dimension

Aggregation. Similar to the partition, aggregation can also be done in the spectrum-channel dimension. Spectrum channels from multiple physical links can be aggregated into one super virtual link. Figure 17.14 shows an example of optical link aggregation. Note that in EON, continuity of subcarriers should be considered for building super channels. Thus, there is a constraint for aggregated virtual links: super channels cannot cross physical links, even though the indexes of subcarriers from different physical links are contiguous. Further, inside a virtual link, its spectrum channels, which come from different physical links, can be same or not, as there are no inter physical link operations on spectrum channels.

Challenges of Physical Substrate Virtualization

In optical networks, physical layer constraints, such as wavelength continuity and physical layer impairments, must be considered for virtualization. Take spectrum

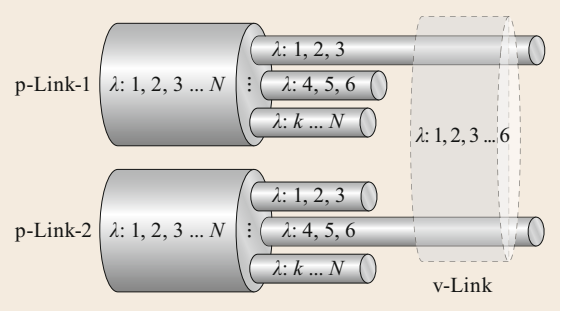


Fig. 17.14 Aggregation of optical links in the spectrum-channel dimension

continuity in EON as an example, guard band subcarriers should be reserved when slicing one contiguous spectrum segment into multiple virtual links. Reversely, spectrum segments from two or multiple physical EON links cannot be aggregated as a super contiguous spectrum segment in one virtual link, even though the indexes of the subcarriers are contiguous.

17.3.2 Features of Optical Network Virtualization

In order to provide logically-isolated optical networks in optical networks, physical resources are abstracted as virtual resources, and the virtual resources are interconnected to build different virtual networks. Virtual optical networks are isolated from each other and can program the virtual resources on the virtualization layer, which means that the virtual optical network can provide users with services similar to those provided by traditional networks deployed in physical optical networks. There are several key features to managing optical network virtualization to provide services as good as those on traditional platforms, which include:

- *Autonomy:* autonomy means that each virtual optical network represents a logically-independent system and has its own considerations about security, service requirements, etc.
- *Isolation:* isolation indicates that the operations on different virtual optical networks must not have any influence on each other, including security considerations, upper limits of bandwidth usage, and so on.
- *Network abstraction:* network abstraction is a feature that can hide the underlying characteristics of network resources and provide unified simplified interfaces to access the optical network resources.
- *Topology awareness:* topology awareness is a deeper imitation for topology convergence. It is indispensable for optimizing the usage of the virtual optical resources.
- *Performance:* a virtualization layer enables the creation of isolated resources on physical resources, while introducing additive overhead and degrades system performance. Then, the performance of virtual optical networks may be not good as the physical network. So, it is necessary to minimize the performance degradation.
- *Mobility:* apart from the above-mentioned characteristics, optical network virtualization should also support mobility that is a movement of virtual resources. Then, each virtual resource can be moved according to users' demands and in order to retain the performance of virtual optical networks.

17.3.3 Key Issues for Virtual Optical Networks

Coexistence of Virtual Networks

Network virtualization provides a more flexible way for optical network resource provisioning; the optical network operator allocates resources with corresponding control rights from the physical network substrate to multiple customers in terms of virtual optical networks. The coexistence of virtual networks on shared network substrates requires independent resource and control rights at the data and control planes, respectively. To support such a coexistence of multiple virtual networks, resource conflicts and interferences must be avoided when allocating resources to multiple virtual networks at the data plane, so that the configurations of one virtual network will not affect other virtual networks. At the control plane, control rights of the resources for each virtual network should be isolated and accurately mapped to corresponding customers to ensure that he virtual network operators' operations are independent.

Dynamic Reconfigurability

Flexibility refers to the capability of building a system and expanding it as needed in order to adapt to internal or external changes. Network virtualization provides quick reconfiguration of virtual networks to enhance flexibility to environmental changes [17.2]. In legacy networks, the scale of a network is restricted by the number of physical resources, so scaling out, i.e., adding additional physical resources to the network, is a simple methods to expand virtual networks. However, this approach cannot be taken in a flexible manner, because adding physical resources implies not only hardware cost, but also maintenance and operation costs for floor space, operations, rack occupancy, etc. Network virtualization allows the reuse of such re-

sources, thereby achieving scalability and efficiency in the use of network resources. Network virtualization allows adding or aggregating additional logical resources to a virtual resource in order to provide increased capability at a lower cost than by adding physical resources.

Network Level Operability

Through virtual optical networks, network operators provide not only bandwidth resources, but also the management interfaces of allocated resources to customers. On virtual optical networks, network customers can manage the allocated virtual resources as they were operating physical networks. For example, network customers can configure backup connections when setup a working lightpath. To support such network level manageability, network operators need to report constraints about spectrum-channel management to network customers. These constraints include blocking constraints in virtual optical nodes and constraints on spectrum-channel continuity in virtual elastic optical links. For example, spectrum channels inside the virtual link in Fig. 17.6 come from two physical links, and the spectrum discontinuity in such case should be reported to virtual network users. Being aware of such constraints, virtual network operators can avoid setting up superchannels with discontinuous spectrum channels.

17.4 Virtual Optical Network (VON)

17.4.1 Introduction of Virtual Optical Network (VON)

Based on the resource virtualization described in Sect. 17.3, a virtual optical network (VON) is proposed to facilitate the sharing of optical network infrastructures among different users and applications, as well as enable network operators to efficiently offer their network resources as a service [17.9–12].

General Concept Description

Multiple VONs can be constructed for multiple infrastructure renters or applications to share the optical network resources over the same optical network infrastructure, as shown in Fig. 17.15. A VON consists of several virtual nodes (VNs) interconnected by virtual optical links (VOLs) that need to be mapped to the optical network infrastructure. The optical network infrastructure is composed of several substrate nodes (SNs) interconnected by substrate fiber links (SFLs). By virtualizing SFL resources of an optical network infrastructure, VONs with widely varying characteristics can be created. Moreover, the VONs introduced by the resource virtualization mechanisms allow network operators to manage and modify optical networks in a highly flexible and dynamic way [17.13–15].

VON over Different Network Scenarios

VONs can be implemented in different types of networks, such as wavelength-division multiplexing (WDM) optical networks, elastic optical networks (EONs) and spatial division multiplexing (SDM) enabled EONs, which are introduced as follows.

VON over WDM Optical Networks. In WDM optical networks, the optical spectrum of an optical fiber is divided into separate parallel wavelength channels with a spectrum spacing of 50 or 100 GHz, as specified by ITU-T standards [17.16, 17]. The optical signal is transmitted over each wavelength channel by transponders that support fixed line rates (e.g., 10, 40, and 100 Gb/s). The wavelength channels in the SFL, i.e.,

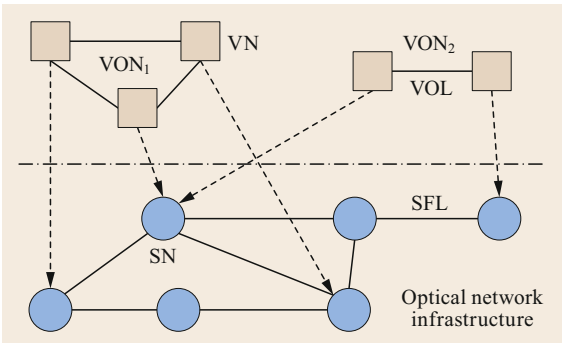


Fig. 17.15 The virtual optical network (VON) paradigm

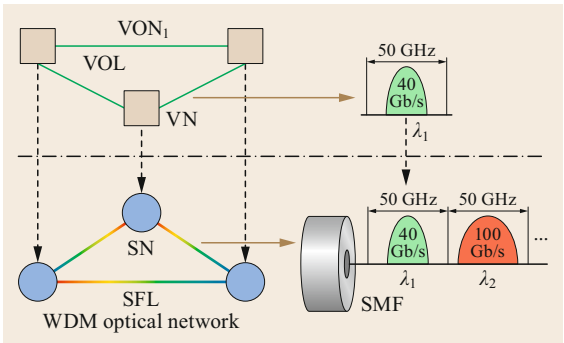


Fig. 17.16 VON over WDM optical networks

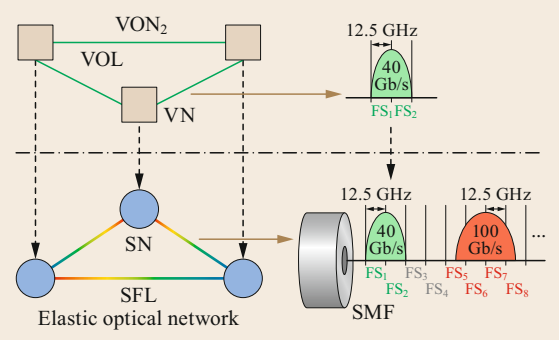


Fig. 17.17 VON over EONs

single-mode fiber (SMF) can be shared by multiple VONs, whereas one wavelength channel can only be utilized by a VOL in a VON at the same time. The VOL in VONs from a starting VN to a terminating VN is associated with its bandwidth demand. An example of VON over WDM networks is shown in Fig. 17.16. Different VONs can utilize different wavelengths and demand a different number of wavelengths, e.g., VON1 demands one wavelength and utilizes λ_1.

VON over Elastic Optical Networks (EONs). Elastic optical networks (EONs) based on sliceable bandwidth variable transponders (SBVTs) are proposed to efficiently utilize optical spectrum resources, which enables optical channels to span across multiple frequency slots (FSs) with a flexible grid (e.g., 12.5 GHz) [17.18]. The difference compared to WDM optical networks is that in a flexible-grid network the SN consists of a bandwidth-variable OXC that allows a flexible spectrum to be switched, which is also capable of traffic grooming. Mixed line rates (e.g., 40, 100, and 400 Gb/s) are allowed to coexist in EONs. An example of VON over EONs is shown in Fig. 17.17. Different VONs can utilize different FSs and demand a different number of FSs, e.g., VON2 demands two FSs and utilizes FS1 and FS2.

17.4.2 Virtual Optical Network (VON) Design

ILP Model for VON Embedding

In this section, an integer linear programming (ILP) model is introduced for virtual optical network embedding (VONE) over elastic optical networks (EONs). The ILP model is inspired by the all-or-nothing multi-commodity flow problem, and its details are as follows:

- Given:
  - $G^{\mathrm{S}}(V^{\mathrm{S}}, E^{\mathrm{S}})$: the substrate topology of the physical infrastructure
  - V^{S}: the set of substrate optical nodes
  - E^{S}: the set of substrate optical fiber links
  - v^{S}: a substrate node, $e^{\mathrm{s}} \in V^{\mathrm{S}}$
  - e^{S}: a substrate fiber link, $e^{\mathrm{s}} \in E^{\mathrm{S}}$
  - B^{S}: the total number of frequency slots on each fiber link E^{S}
  - $\omega^{\mathrm{S}}_{(\mu^{\mathrm{S}},v^{\mathrm{S}}),k}$: the starting frequency slot index of the k-th maximal contiguous slot blocks (MCSBs) on the substrate fiber link E^{S}=$(\mu^{\mathrm{S}}, v^{\mathrm{S}})$.
  - $Z^{\mathrm{S}}_{(\mu^{\mathrm{S}},v^{\mathrm{S}}),k}$: ending frequency slot index of the k-th MCSBs on the substrate fiber link $E^{\mathrm{S}} = (\mu^{\mathrm{S}}, v^{\mathrm{S}})$
  - $G^{\mathrm{r}}(V^{\mathrm{r}}, E^{\mathrm{r}})$: the VON request
  - V^{r}: the set of virtual optical nodes
  - E^{r}: the set of virtual optical links
  - v^{r}: a virtual optical node, $v^{\mathrm{r}} \in V^{\mathrm{r}}$
  - e^{r}: a virtual optical link, $e^{\mathrm{r}} \in V^{\mathrm{r}}$
  - $S^{\mathrm{r}}e^{\mathrm{r}}/d^{\mathrm{r}}e^{\mathrm{r}}$: end nodes of the virtual optical link e^{r}
  - n^{r}: Bandwidth requirement of each virtual optical link.
- Variables:
  - $\xi^{v^{\mathrm{r}},v^{\mathrm{s}}}$: a Boolean variable that equals 1, if the virtual node v^{r} is mapped onto the substrate node v^{s}, otherwise, $\xi^{v^{\mathrm{r}},v^{\mathrm{s}}} = 0$.
  - $f^{e^{\mathrm{r}}}_{u^{\mathrm{s}},v^{\mathrm{s}}}$: a Boolean all-or-nothing flow variable that equals 1 if the all-or-nothing flow is on the substrate fiber link $e^{\mathrm{S}} = (u^{\mathrm{s}}, v^{\mathrm{s}})$ for the virtual optical link $e^{\mathrm{r}} \in E^{\mathrm{r}}$, otherwise, it is 0.
  - $\delta^{k}_{u^{\mathrm{s}},v^{\mathrm{s}}}$: a Boolean variable that equals 1, if a contiguous spectrum block in the k-th MCSB on the substrate fiber link $e^{\mathrm{S}} = (u^{\mathrm{s}}, v^{\mathrm{s}})$ is assigned to the VON request $G^{\mathrm{r}}(V^{\mathrm{r}}, E^{\mathrm{r}})$, otherwise, it is 0.
  - ω: an integer variable that indicates the starting frequency slot index of the contiguous spectrum block assigned to the VON request $G^{\mathrm{r}}(V^{\mathrm{r}}, E^{\mathrm{r}})$, $\omega \in [1, B^{\mathrm{S}}]$.
  - Z: an integer variable that indicates the ending frequency slot index of the contiguous spectrum block assigned to the VON request $G^{\mathrm{r}}(V^{\mathrm{r}}, E^{\mathrm{r}})$, $Z \in [1, B^{\mathrm{S}}]$.
- Objective

$$\text{Minimize } n^{\mathrm{r}} \sum_{e^{\mathrm{r}} \in E^{\mathrm{r}}} \sum_{(u^{\mathrm{s}},v^{\mathrm{s}}) \in E^{\mathrm{s}}} f^{e^{\mathrm{r}}}_{u^{\mathrm{s}},v^{\mathrm{s}}} \,. \tag{17.1}$$

The objective minimizes the total number of frequency slots used for the VON request.

- Constraints:
  - Node mapping

$$\sum_{v^{\mathrm{s}} \in V^{\mathrm{s}}} \xi_{v^{\mathrm{r}},v^{\mathrm{s}}} = 1 \,, \quad \forall v^{\mathrm{r}} \in V^{\mathrm{r}} \,, \tag{17.2}$$

$$\sum_{v^{\mathrm{r}} \in V^{\mathrm{r}}} \xi_{v^{\mathrm{r}},v^{\mathrm{s}}} \le 1 \,, \quad \forall v^{\mathrm{s}} \in V^{\mathrm{s}} \,. \tag{17.3}$$

– Link mapping

$$\sum_{(u^{\mathrm{s}},v^{\mathrm{s}})\in E^{\mathrm{s}}} f^{e^{\mathrm{r}}}_{u^{\mathrm{s}},v^{\mathrm{s}}} - \sum_{(v^{\mathrm{s}},u^{\mathrm{s}})\in E^{\mathrm{s}}} f^{e^{\mathrm{r}}}_{v^{\mathrm{s}},u^{\mathrm{s}}} = \xi_{s^{\mathrm{r}}_{e^{\mathrm{r}}},u^{\mathrm{s}}} - \xi_{d^{\mathrm{r}}_{e^{\mathrm{r}}},u^{\mathrm{s}}}\,, \quad \forall e^{\mathrm{r}} \in E^{\mathrm{r}}\,, \forall u^{\mathrm{s}} \in V^{\mathrm{s}}\,, \tag{17.4}$$

$$\sum_{e^{\mathrm{r}}\in E^{\mathrm{r}}} (f^{e^{\mathrm{r}}}_{u^{\mathrm{s}},v^{\mathrm{s}}}) \le 1\,, \quad \forall (u^{\mathrm{s}},v^{\mathrm{s}}) \in E^{\mathrm{S}}\,, \tag{17.5}$$

$$z - \omega + 1 = n^{\mathrm{r}}\,, \tag{17.6}$$

$$\sum_{k} \delta^{k}_{u^{\mathrm{s}},v^{\mathrm{s}}} = \sum_{e^{\mathrm{r}}\in E^{\mathrm{r}}} (f^{e^{\mathrm{r}}}_{u^{\mathrm{s}},v^{\mathrm{s}}})\,, \quad \forall (u^{\mathrm{s}},v^{\mathrm{s}}) \in E^{\mathrm{S}}\,, \tag{17.7}$$

$$\omega \ge \omega^{\mathrm{s}}_{u^{\mathrm{s}},v^{\mathrm{s}},k}\delta^{k}_{u^{\mathrm{s}},v^{\mathrm{s}}}\,, \quad \forall (u^{\mathrm{s}},v^{\mathrm{s}}) \in E^{\mathrm{S}}\,, \forall k\,, \tag{17.8}$$

$$z \le (z^{\mathrm{s}}_{u^{\mathrm{s}},v^{\mathrm{s}},k} - B^{\mathrm{s}})\delta^{k}_{u^{\mathrm{s}},v^{\mathrm{s}}} + B^{\mathrm{s}}\,, \quad \forall (u^{\mathrm{s}},v^{\mathrm{s}}) \in E^{\mathrm{S}}\,, \forall k\,. \tag{17.9}$$

- Explanations for the constraints

Equations (17.2)–(17.3) ensure that each virtual node in the VON request is mapped onto a unique substrate node. Equation (17.4) is the flow conservation constraint. This constraint ensures that on all the substrate nodes, the total number of the inflows equals to that of the outflows, except for the embedded substrate nodes for the end nodes of the virtual optical link. Equation (17.5) ensures that for all the virtual optical links, their embedded lightpaths in the substrate topology are link-disjoint. Equation (17.6) ensures that the size of the assigned contiguous spectrum block for the VON request can satisfy its bandwidth requirement. Equations (17.7)–(17.9) make sure that the assigned contiguous spectrum block for $G^{\mathrm{r}}(V^{\mathrm{r}}, E^{\mathrm{r}})$ is located in a single MCSB on the substrate fiber links, and the size of the MCSB is bigger than or equal to that of the assigned contiguous spectrum block.

Heuristic Approach for VON Embedding Design

This section presents a heuristic approach to solve large problem instances of the virtual optical network design. For the large-scale networks, it is difficult to use traditional ILP methods to solve the VON embedding due to computational constraints. Heuristics become important when traditional linear programming (LP) methods due to computational constraints are time constrained. Different heuristic algorithms achieve different optimization targets, including energy efficiency and minimized cost. Aiming at the cost-efficient design problem of a survivable virtual infrastructure [17.19], survivable virtual infrastructure mapping approaches are developed to minimize the network resource costs while guaranteeing virtual infrastructure survivability. Energy and spectrum-aware VON mapping approaches are proposed by using the minimum submatrix scheme, which achieves the minimizing energy consumption and reduces spectrum usage [17.20]. In [17.21], the role of high-performance dynamic optical networks in cloud computing environments is addressed. The results provide guidelines for the infrastructure providers to effectively and optimally provision virtual infrastructure services. Meanwhile, VON embedding problems are investigated by considering both transparent and opaque VONs over elastic optical networks [17.4].

A new infrastructure as a service architecture is proposed by utilizing optical network virtualization and addressing the physical impairment problem on a physical optical network [17.22]. In [17.23], mechanisms for embedding virtual cloud networks to minimize the overall power consumption, i.e., the aggregate of the power consumption for communication and computing in the data centers (DCs), were developed. An embedding approach based on candidate mapping patterns was proposed in [17.24] to provide the requested resources according to the shortest path routing.

The design of VON mainly focuses on the problem of virtual optical networks embedding (VONE), which deals with the allocation of virtual resources both in optical nodes and links. Therefore, the VONE problem can be formulated as two subproblems: virtual node mapping (VNM) and virtual link mapping (VLM). In the VNM subproblem, virtual optical nodes must be allocated in physical nodes. In the VLM subproblem, virtual links connecting these virtual nodes must be mapped to lightpaths in the underlying optical network. Thus, we can solve the VONE problem by addressing VNM and VLM subproblems. The VNM and VLM subproblems can be solved in two different ways: uncoordinated mapping and coordinated mapping. In uncoordinated mapping, each subproblem is solved in an isolated and independent way. In this case, VNM must be solved first, since the result of VNM provides the input for VLM. The drawback is that the optimization in VNM may jeopardize the optimization in VLM. If VNM is performed without considering its relation to link mapping, the solution space is restricted, and the overall performance of the embedding decreases. On the contrary, some VONE approaches have improved the performance of the solution by providing coordination between the two phases. In this section, we focus on the coordinated mapping approaches.

In this part, we choose one example to show the procedure of the heuristic algorithm. In this approach [17.25], based on a breadth-first search (BFS), VONE is solved in one stage considering the impact of node mapping on the link mapping stage. The first

step is to measure the topology incidence of each node and conduct the node ranking. If node mapping incorporates topology attributes, the link mapping efficiency and the acceptance ratio are improved. If two nodes have the same resources, the node with the more capable neighborhood is chosen, which leads to a higher success probability for the embedding. Then a BFS tree is built, where the root node is the virtual node with greatest rank, and the children are placed from left to right based on their rank. VONE is performed by going through the BFS tree. Each virtual node is mapped in the first feasible substrate node of its list. At the same time, each virtual link is mapped to the shortest path of the substrate that can satisfy the bandwidth demand. Based on the results of VNM, VLM can be solved in two different ways: single-path mapping and multiple-path mapping. In the first case, each virtual link must be mapped to a single path in the underlying optical networks. In the second case, each virtual link demand can be carried by several paths in the substrate networks.

Input: a physical network, a set of VON requests;
Output: results of VON mapping on the optical network;

for each VON request **do**
  Calculate node ranks for substrate and virtual nodes;
  Builds a breadth-first search (BFS) tree, where the root node is the virtual node with greatest rank, and the children are placed from left to right based on their rank;
  for all virtual nodes and virtual links in BFS tree **do**
    Map each virtual node in the first feasible substrate node;
    Map the virtual links incident to that virtual node onto the substrate shortest paths;
    if the bandwidth is enough to meet the capacity demand **then**
      Perform the VON embedding;
    else
      Block the VON request;
    end if
  end for
end for

17.4.3 Survivability of Virtual Optical Network

Optical network survivability is an important issue in the design of VON. In particular, protection against failures is important failures in order to realize reliable VON provisioning. Link or node failures in the physical networks will lead to the failure of one or multiple VONs. Therefore, we need to consider the possible network failures when mapping VON to the underlying optical networks. In general, dedicated or shared protection can be employed for each virtual node or link. Note that we focus on the failure of single network elements in this section. According to the location of network failures, survivable VON mapping can be classified into two types.

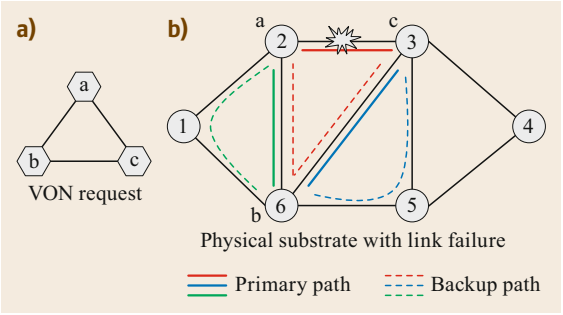


Fig. 17.18a,b Survivable VON mapping for link failure. **(a)** VON request; **(b)** survivable VON mapping for link failure

Link Failure

Considering link failure in a substrate network, each virtual link is mapped to two link-disjoint paths, the primary path and the backup path. Figure 17.18 shows the link failure in the survivable VON mapping. We can see that the VON request is a three-node ring network. Three virtual nodes are mapped to optical node 2, 6, 3, respectively. Virtual link (a, c) is mapped to two physical routes (2, 3) and (2, 6, 3). The former is the primary path, and the latter is the backup path. Since primary and backup paths have no common physical links, link failure will only affect one path at most. Therefore, the survivability of VON can be guaranteed.

Input: a physical network, a set of VON requests;
Output: survivable VON mapping for link failure;

for each VON request **do**
  for all all descending orders of the virtual node set based on the resource demand **do**
    Select one virtual optical node with the largest resource demand;
    for all all descending orders of the physical node set based on the capacity of available resources **do**
      Select one optical node with the largest resource provisioning;
      if the resource is enough to meet the capacity requirement **then**

```
        Map the virtual node to physical node;
      else
        Block the VON request;
      end if
    end for
  end for
  for all the virtual links in VON request do
    Search two link-disjoint paths with enough spectrum resource;
    if two link-disjoint paths can be found from the k-shortest paths then
      Map the virtual link to two physical paths;
    else
      Block the VON request;
    end if
  end for
end for
```

Node Failure

Figure 17.19 shows the node failure in survivable VON mapping. We assume that optical node failure will lead to the failure of adjacent optical links. Meanwhile, node failure will cause the failure of multiple VONs. For protection against potential node failures in VON mapping, each virtual node is mapped to two physical nodes, a primary node and a backup node. Each virtual link should be mapped to two node-disjoint physical routes.

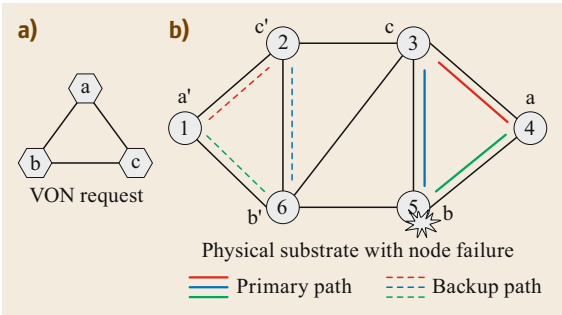


Fig. 17.19a,b Survivable VON mapping for node failure

For instance, virtual node *a* can be mapped to two physical nodes #4 and #1. Virtual link (a, c) can be mapped to physical routes (4, 3) and (1, 2). Since these two routes do not have any common physical nodes, node or link failure will not affect the survivability of this virtual link.

```
Input: a physical network, a set of VON requests;
Output: survivable VON mapping for node failure;

for each VON request do
  for all all descending orders of virtual node set based on the resource demand do
    Select one virtual optical node with the largest resource demand;
    for all all descending orders of physical node set based on the capacity of available resources do
      Select two optical nodes with the largest resource provisioning;
      if the resource on two optical nodes is enough to meet the capacity requirement then
        Map the virtual node to two physical nodes;
      else
        Block the VON request;
      end if
    end for
  end for
  for all the virtual links in VON request do
    Search two node-disjoint paths with enough spectrum resource;
    if two node-disjoint paths can be found in the substrate network then
      Map the virtual link to two node-disjoint physical paths;
    else
      Block the VON request;
    end if
  end for
end for
```

17.5 Network Function Virtualization in Optical Networks

In order to take advantage of SDN/NFV paradigms in optical networks, it is of key importance to justify an architecture for SDN/NFV transport networks [17.26]. This section presents the first architecture for SDN/NFV transport networks and later introduces two examples of VNFs: the control plane and the path computation element (PCE). Such examples are experimentally validated to show that these VNFs can be deployed in reality.

17.5.1 Architecture for SDN/NFV Transport Networks

Figure 17.21 depicts the architecture for providing such NFV services. The main components of this architecture are: NFV orchestrator, VNF manager, and the integrated cloud and network orchestrator.

As previously presented in this chapter, the ETSI [17.5] defines the virtual infrastructure manager

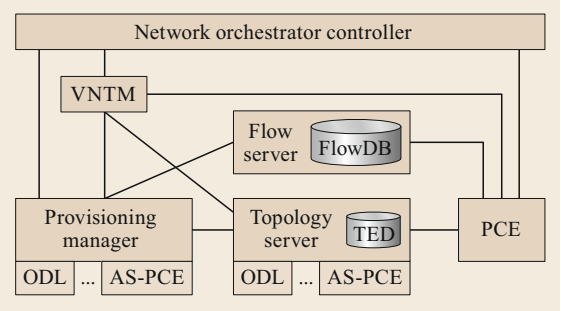


Fig. 17.20 Multidomain SDN orchestrator architecture

(VIM). The VIM has to manage and control the interaction of a VNF with the network and IT resources, as well as the virtualization of these resources. Here, the integrated cloud and network orchestrator [17.27] takes the role of a VIM. The integrated cloud and network orchestrator is not only able to interact with a cloud controller (e.g., OpenStack) but also with a multidomain SDN orchestrator (MSO) [17.28]. The integrated cloud and network orchestrator interacts with the MSO by declaring the necessary services that need to be established for the E2E path provisioning of several virtual machines (VM).

The following sections describe the different components of the proposed architecture following a bottom-up approach.

Multidomain SDN Orchestrator

When deploying services across different network domains through per-domain SDN/OF or GMPLS/PCE controllers, an entity to support such end-to-end connectivity by orchestrating is needed. This component is called the multidomain SDN orchestrator (MSO), and it is depicted in Fig. 17.20. The MSO considers the heterogeneous underlying network resources (e.g., multidomain, multilayer, and multicontrol network resources), and it assumes that the underlying SDN controllers can to provide network topology information and flow programming functions. The MSO enables multilayer and multidomain network orchestration, as demonstrated in [17.28]. The MSO has to implement different plugins for each SDN controllers, which are typically technology and vendor dependent.

The MSO architecture is composed of the following components: network orchestration controller, virtual network topology manager (VNTM), topology server, provisioning manager, and path computation element (PCE). This architecture is based on the application-based network operations (ABNO) framework [17.29, 30], which has been standardized by IETF. Subsequently, we briefly describe the main role of the components:

- The network orchestration controller is responsible to handle different requests in order to provision end-to-end connectivity services. For this purpose, an NBI is exposed to other applications, so they can use the exposed API.
- The topology server is the component that is able to recover the network topology from each control domain and to later provide a unified network topology, which is stored in the traffic engineering database (TED). The TED is the database that includes all the necessary information regarding network links and nodes. The PCE recovers the topology from the TED and computes routes across the network. Moreover, it requires some inventory information for some operations.
- VNTM is the component responsible for managing the necessary multilayer connectivity. The VNTM establishes the necessary optical lightpaths, which are then offered as logical links to the upper layers to satisfy any incoming connectivity service requests.
- Through different plugins the provisioning manager consumes the necessary provisioning interfaces to command the establishment of connectivity services to underlying domain controllers. The Flow server stores the connections established in the network into a Flow DataBase (FlowDB).

Integrated Cloud and Network Orchestrator

The integrated cloud and network orchestrator must deal with different network and IT resources. The integrated cloud and network orchestrator provides the following services: the VM create, read, update, delete (CRUD) mechanism; VON CRUD; the network CRUD mechanism; and VM migration.

The VM CRUD Mechanism. The VM CRUD mechanism enables to create, read, update or delete a VM using a REST API. A VM can be requested based on its availability zone, the disk image to be loaded, or its hardware resources (i.e., flavor). The VM must be allocated inside a network. A second management network is configured by default to the VM in order to provide management access to the VM by the different applications (e.g., VNF managers).

The VON CRUD Mechanism. Similarly, the VON CRUD mechanism allows us to create, read, update or delete a VON.

The Network CRUD Mechanism. The network CRUD mechanism allows to create, read, update, or delete a network. To create a network a valid IP range is necessary, which is used to select an IP address to

be assigned to a VM virtual network interface card (NIC). A network consists of end-to-end paths that are created between each VM attached to the network (resulting in $(n(n-1)/2)-n$ connections). The setup of an end-to-end path between two VMs is described in Fig. 17.24. The integrated cloud and network orchestrator requests connectivity service between VM1 and VM2 to MSO. The MSO computes the route, and it interconnects the computing resources through the packet network (intra-DC) or through an inter-DC connection.

In the intra-DC scenario, the MSO uses the necessary underlying SDN controller NBI to establish the OpenFlow commands (i.e., flowmod) to the virtual host switches (e.g., OpenVSwitch—OVS) and into the intra-DC switches. In the inter-DC scenario, the MSO first establishes an optical lightpath between the DCs using the AS-PCE PCInitiate command. When the optical lightpath has been established, the SDN Controller discovers the new layer-2 link established between the DCs. A new path is computed, and the necessary commands are sent to the necessary SDN controller, following the same procedure as detailed in the intra-DC scenario.

VM Migration. The most common types of VM migration are live and block migrations. Live migrations allow moving a VM without stopping its inside running processes. Block migration requires stopping the VM, which involves a certain service downtime. In the proposed architecture, VMs do not run isolatedly. It is typical that a VM is interconnected with other VMs in the same network in order to offer a joint service (i.e., network service in the NFV architecture). So, it is of the essence that if one of the VMs is migrated, it is necessary to maintain its connection state. This is known as VM seamless migration, which is performed by the integrated cloud and network orchestrator [17.27].

VNF Manager

A VNF manager is responsible for the life-cycle management of a VNF, which means the creation, configuration, and removal of a VNF. Moreover, the VNF scale out/in is also controlled by the VNF manager. These functions could be carried out by a single VNF manager or by multiple ones.

The proposed VNF manager has the following basic components: request for virtual IT resources, VNF life-cycle management, and overall control of the VNF. The interface of the VNF manager to the NFV orchestrator enables the control of the VNF. The integrated cloud and network orchestrator must obtain the necessary IT and network resources to deploy a VNF (and to scale them in the case of need). The virtual IT resources component interacts with the integrated cloud and network orchestrator to obtain such computing and network resources. The VNF manager can access the computing resources obtained to control the VNF, so it has access to the allocated VM. Lastly, the VNF life-cycle component handles the creation/configuration/removal (i.e., life cycle) of the VNF. Moreover, it also monitors its behaviors and notifies the VNF controller of incidences.

NFV Orchestrator

As was previously presented in this chapter, ETSI [17.5] defines the NFV manager and orchestrator (MANO) architecture. Let us highlight their responsibilities, which are twofold: first, to support the infrastructure virtualization by handling the life-cycle management of the physical and software resources and, second, to create/destroy the different VNFs. The NFV orchestrator (within NFV MANO) is in charge of handling the various VNF managers and of offering the aforementioned services. The northbound interface (NBI) of the NFV orchestrator must be standard, because users or applications will use it to request the NFV services.

17.5.2 Control Plane Virtualization for GMPLS and OpenFlow Environments

Use Case Definition

SDN and NFV are key concepts for understanding the current network evolution. In recent year, the main objective of telecom operators has been to adopt the innovations of the IT industry in a gradual manner. SDN orchestration [17.31] proposes as a realistic approach to dealing with multiple network technologies, vendors, and control domains, even with interconnected data centers that are able to provide virtualized network functions. SDN orchestration covers the necessary requirements for end-to-end (E2E) connections, which focus on high-bandwidth provisioning and low-latency connections. SDN orchestration focuses on network control and abstraction through several control domains whilst using standard protocols and modules. A network control domain is understood as a set of network elements (NE) under a logically centralized SDN controller. Multidomain SDN orchestration can be studied in several contexts, such as pure OpenFlow (OF)-enabled networks, or in more complex scenarios with heterogeneously-controlled networks (GMPLS/PCE and OF). Several initiatives in standardization organizations such as ONF transport API or IETF transport NBI advocate for the necessity of SDN orchestration.

OIF-ONF showed the results of their global transport SDN prototype demonstration in [17.32], which included several SDN controllers and hierarchical control levels. The IETF abstraction and control of transport networks (ACTN) framework [17.33] provides a number of

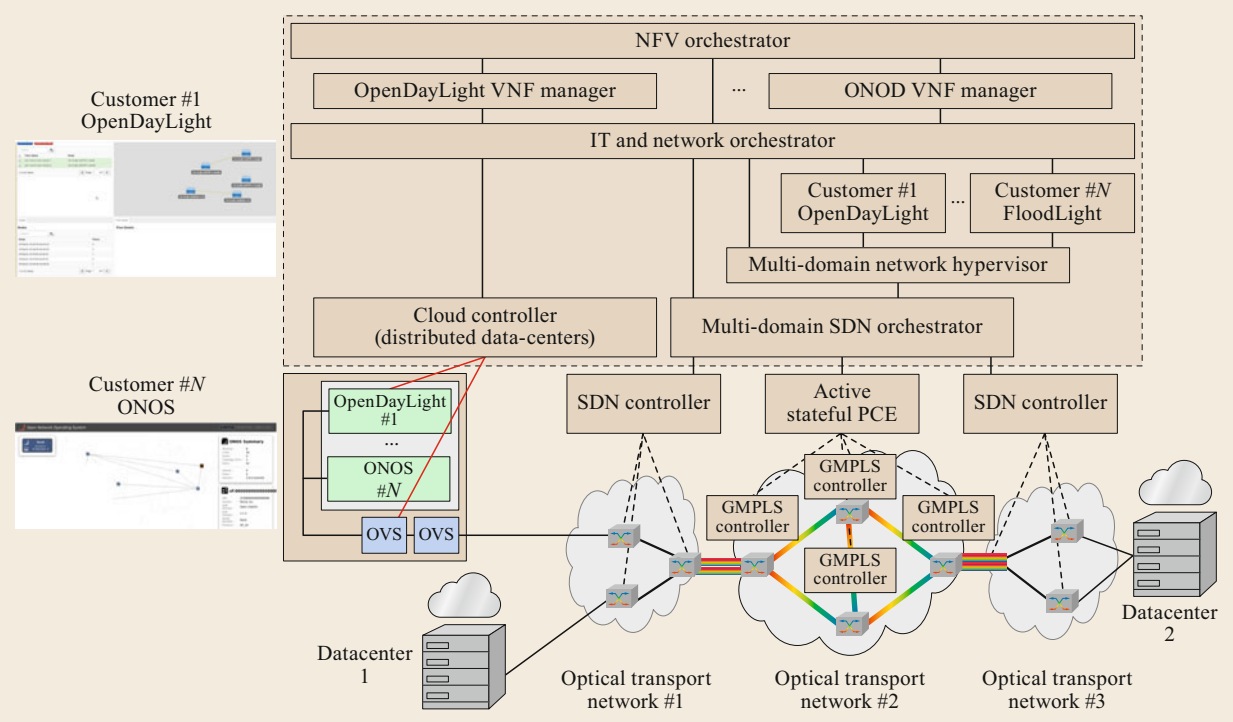


Fig. 17.21 Transport network function virtualization architecture

capabilities and features such as: i) seamless hierarchical service coordination across multitenants, ii) control of virtual and physical network domains, and iii) horizontal E2E service coordination across multidomain networks.

The SDN approach can solve two points from the virtualization perspective. There must be a mechanism to split the physical resources into multiple virtual optical networks (VON). This can be done with SDN. Secondly, the control plane software to carry out such operations can be deployed as a VNF in a data center. As the control plane software is decoupled from the data plane, the control plane can use several technologies to operate the network, like OpenFlow or GMPLS. Moreover, the tenant can control the VON deployed through a tenant SDN controller run on the cloud. The traditional architecture to provide VON consists of dedicated hosts to run the SDN controllers of each VON. Several software implementations, such as OpenDayLight or ONOS, can be deployed to configure the VON.

The idea of NFV is that the creation of VNFs is dynamic. Therefore, the deployment of independent SDN controller instances must be automated, so every time a new VON is deployed, its SDN controller is also deployed. The dynamic deployment not only has the dynamicity to provide the service, but also offers additional advantages such as the lack of hardware maintenance downtime, along with faster recovery times in the case of a network disaster. A virtual SDN controller can be quickly and easily moved between physical hosts within a data center when hardware maintenance is required.

The next sections present how the use case for control plane virtualization can be deployed on the architecture for SDN/NFV transport networks presented above. This includes the dynamic deployment of VONs and their respective customer SDN controller (CSC) as VNFs in data centers.

Multidomain SDN Orchestrator Role

The physical substrate in optical networks consist of multiple domains and heterogeneous transport technologies (e.g., Flexigrid EON, OPS (optical packet switching) and MPLS-TP (MPLS transport profile)). The VON resource-allocation algorithm considers the information of physical networks, that is the network topology, including the connectivity of the interdomain links and the availability of physical resources (e.g., switch ports or spectrum slots), as well as specific attributes and constraints of the multitechnology domains (e.g., spectrum continuity and impairments). The MSO is in charge of such E2E provisioning. Let us assume a scenario as depicted in Fig. 17.21, where domains are controlled by GMPLS and SDN controllers.

Figure 17.22 shows the proposed message exchange between the MSO and the underlying SDN and GM-

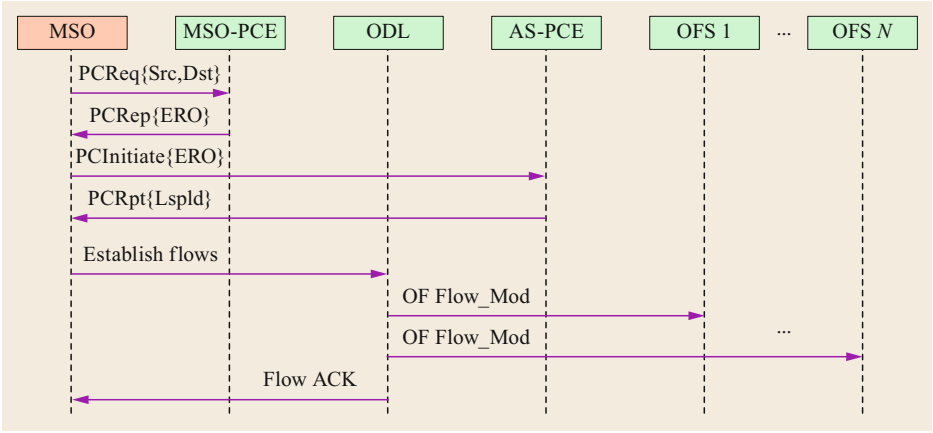


Fig. 17.22 Multidomain SDN orchestrator workflow

PLS/PCE controllers. Let us assume that there are two different control plane technologies in the network. There is one domain with a PCE/GMPLS control plane and a second domain with an OpenFlow SDN controller (OpenDayLight). The steps of the workflow are as follows:

- The initial trigger is an E2E connection to the MSO.
- The MSO asks the MSO PCE to compute the necessary path, taking into account the overall topology of the network domains. In this example, PCEP is used for this interface.
- Once the multidomain path is known, the MSO first requests an optical lightpath to the AS-PCE, which acts as an optical SDN controller. That is, it configures the underlying optical devices, which are based in GMPLS. The AS-PCE returns a PCRpt once the configuration has been done in the domain.
- After the first domain has been configured, the MSO configures the second domain. It requests the SDN controller to configure the remaining part of the E2E connection.
- The ODL controller configures each device one by one, as OpenFlow is a per-device protocol.
- Finally, the ODL controller returns a confirmation to the MSO to acknowledge that the configuration has been carried out.

In the previous example, it is clear that the MSO must configure each domain using a different interface. The Control Orchestration Protocol (COP) [17.34] is proposed as an abstraction protocol, which provides a common set of control plane functions used by a number of SDN controllers. This approach allows the interworking of heterogeneous control plane paradigms (i.e., OpenFlow, GMPLS/PCE), but using the same interface. The COP YANG data model (yet another next generation data model) can be transported using the IETF RESTconf protocol, which is being adopted by the industry. The COP data models cover a network topology service, a call service for establishing E2E connections, and a path computation service. It has been demonstrated that the use of COP or similar data models eases the deployment of SDN orchestration. As has been mentioned, ONF and IETF have proposed similar protocols to carry out similar operations.

Multidomain Network Hypervisor

The multidomain network hypervisor (MNH) is the entity responsible for network virtualization [17.35]. It interacts with an MSO to provision virtual links, which compose the VON, not only in a each domain, but also in E2E fashion. Such integration means that it can request a VON to move towards an MNH and use the customer (i.e., tenant) SDN controller. The MNH is in charge of providing the virtualization and abstraction of the underlying network resources (Fig. 17.23). This provides a network overlay, which allows dynamically deploying multitenant virtual networks on top of networks orchestrated by the MSO.

The MNH architecture is detailed in [17.35], and a short reference is provided here. The MNH interface to request both virtual switches and virtual links to actually deploy a VON is provided by the component VON

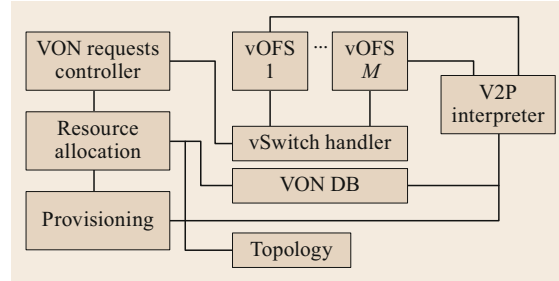


Fig. 17.23 Multidomain network hypervisor architecture

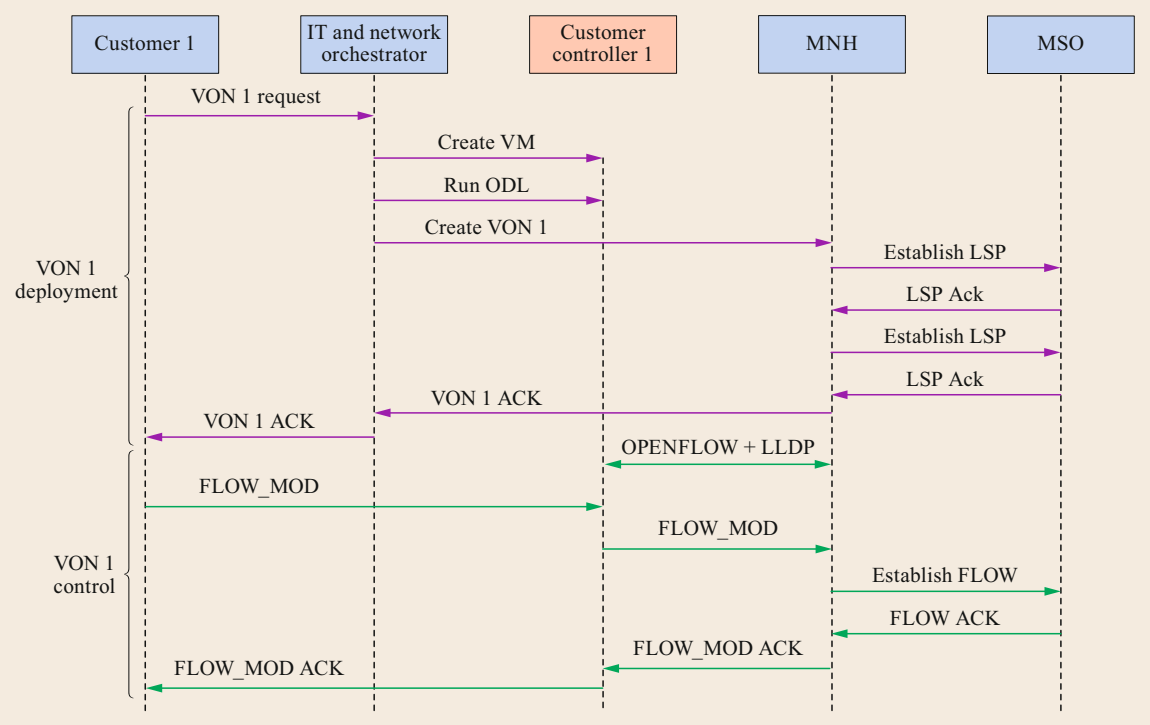


Fig. 17.24 Multidomain network hypervisor workflow

request controller. The CSC is identified by its own IP address. This CSC IP address is used by the virtual switch handler to provide an abstract network view of the VON allocated to the CSC.

A virtual switch request consists of a number of virtual switches, which include virtual ports and virtual links. The virtual ports relate to the underlying NE (which might have been abstracted as nodes by the MSO). A virtual link request includes source and destination virtual switches. The allocation of the physical ports of the physical domains to the virtual switches is performed by the resource allocation (RA) component. It is also responsible for requesting from the MSO (through the provisioning component) the necessary multidomain connections to interconnect the requested virtual switches, which are related to physical domains. Once the connections have been established, the RA allocates the virtual port identifiers to which the connections are related.

For each VON, the virtual switch handler provides the necessary OpenFlow datapaths towards the IP address of the virtual tenant SDN Controller (CSC). Each OF datapath is provided by an emulated OpenFlow virtual switch. To provide the virtual VON topology to the CSC, the different emulated OF virtual switches are interconnected with virtual links. The CSC triggers the automatic topology discovery using the link-layer discovery protocol (LLDP) by using OpenFlow packet Out commands to the emulated virtual switches. The destination switches of the packet Out command respond to the CSC with the OpenFlow packet In commands.

The virtual-to-physical (V2P) interpreter is responsible for translating the received OF command (e.g., FLOW_MOD (flow table modification message)) from the CSC using the abstract VON topological view, to the allocated physical resources. The emulated virtual OF switches are connected to it. To this end, the V2P interpreter consults the VON database for the allocated physical ports and the established connections. Later, the processed requests are sent to the provisioning module, which is responsible for requesting the provisioning of the physical resources to the MSO. This workflow is detailed in [17.35].

Figure 17.24 shows a workflow for the dynamic deployment of an SDN-controlled VON and its control. The steps of the creation are explained as follows:

- The VON deployment starts with the creation of the VM that will act as the CSC.
- Once the CSC is running, the integrated cloud and network orchestrator requests a VON to the MNH.
- The MNH interacts with the MSO in order to provision the necessary virtual links. In this example, LSP terminology is used, but it can be OpenFlow

commands or any other technology that the MNH can manage.
- Once the VON is finished, its creation is acknowledged by the MSO.
- CSC is connected to the MNH and is able to handle the allocated VON resources.
- The OF commands to the VON are translated by the MNH.
- Finally, the OF commands are also acknowledged.

An interesting approach for disaster resiliency of the control plane is provided in [17.36], where it is proposed to design it as a virtual network, which can be solved using virtual network mapping techniques. The appropriate mapping of the controllers over the physical network is selected such that the connectivity among the controllers (controller-to-controller) and between the switches to the controllers (switch-to-controllers) is not compromised by physical infrastructure failures caused by disasters.

17.5.3 The Path Computation Element as a Network Service

Use Case Definition

An example of a transport network function is the transport path computation element (PCE). The PCE is a functional component of the control plane able to perform constrained path computation on a graph representing a network. The graph and its topological information are known as a traffic engineering database (TED) [17.37]. The IETF has standardized the architecture of the PCE and its communication protocol (the path computation element protocol—PCEP).

The PCE is an application that can be run on top of commercial off-the-shelf (COTS) equipment [17.28]. The deployment of PCEs was easily justified by the increasing complexity of path computation, so a dedicated element was proposed to do the computation. Later, an active stateful path computation element (AS-PCE) [17.38] was proposed to maintain not only the traffic engineering information (link and node states), but also the state of the active connections in the network. Moreover, an active AS-PCE can manage the active paths controlled by the nodes, allowing the PCE to modify or tear down the connections established in the data plane. This AS-PCE is a transport SDN controller, by acting as an SDN-enabler for a GMPLS-controlled optical transport network.

The operations done in the PCE may be computationally intensive when running the path computation for transport connection provisioning or reoptimization on large production networks. Several solutions have been proposed to avoid scalability limitations, such as hierarchical PCE and front-end/back-end PCE [17.37]. In this use case, it is proposed to extend the concept of NFV to transport networks by removing a dedicated PCE server and virtualizing the PCE functionality.

To demonstrate the feasibility of deploying network services on top of the proposed integrated cloud and network orchestrator, we propose the adoption of the NFV architecture to deploy a PCE dedicated to path computation of a transport network as a network service, consisting of a single VNF. Although the NFV architecture has successfully been demonstrated for mobile networks, there have been only few attempts to introduce this architecture to transport networks.

An NFV orchestrator is introduced, so that the proposed transport PCE network service is able to handle intense peak loads of path computation requests. The NFV orchestrator dynamically deploys virtual PCEs (vPCEs) on demand to keep the quality of the network service (e.g., in terms of latency, request processing time, dedicated algorithms, etc.). A vPCE is defined as a PCE instance that runs as a software application on an NFV infrastructure point of presence (e.g., a virtual machine). A PCE domain name server (DNS) [17.39] is introduced to offer the vPCEs deployed as a single function perceived by the different path computation clients (PCC).

The next section shows how this use case can be deployed on the architecture for an SDN/NFV transport network presented on this chapter. Moreover, it includes an experimental validation of the architecture and provides the results of deploying a transport PCE as a VNF.

Transport PCE Virtual Network Function

The proposed transport NFV architecture is shown in Fig. 17.25. It consists of several components: a) an NFV orchestrator, b) a VNF manager, and c) a virtual infrastructure manager and a NFV infrastructure point of presence.

An NFV orchestrator is the entity responsible for the deployment of the PCE as a network service, consisting of a single VNF. The NFV orchestrator interacts with the VNF manager. The PCE VNF manager consists of three separated modules: a) a PCE VNF controller, b) virtual IT resources, and c) PCE computation load monitoring:

- The necessary logic for deploying the necessary vPCE in order to guarantee the quality of the service provided VNF is provided by the PCE VNF controller. The necessary data to deploy a new instance of a vPCE or to delete one is obtained by the interaction with the PCE computation load monitoring module.
- The virtual IT resources module is responsible for requesting the necessary resources from the integrated cloud and network orchestrator (acting as the NFV VIM). The VIM allows dynamically deploying and releasing virtual machines with custom images

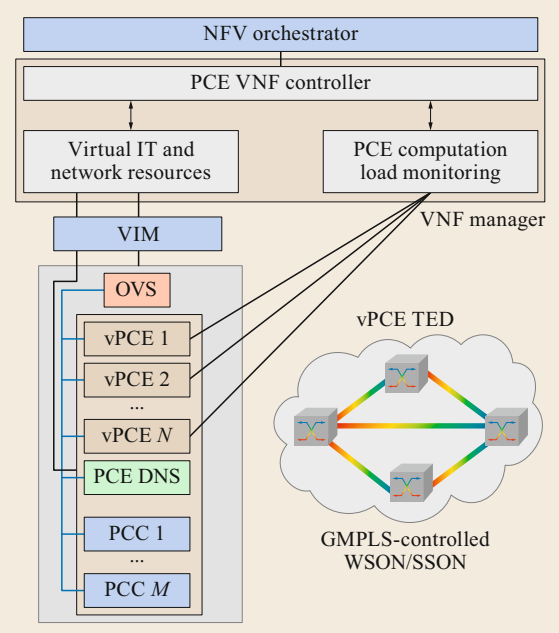


Fig. 17.25 Transport NFV architecture

running vPCE as an application and interconnecting them, although they might be geographically distributed. To each vPCE instance VIM must assign a new IP address from a set of available ones. This IP address is parsed, and the PCE DNS is notified.

- The PCE computation load monitoring module monitors the quality of the VNF. The parameters monitored are a set of the PCE monitoring parameters, such as concurrent path computation requests. These monitoring parameters are defined in [17.40] and are exposed by an HTTP server running on each vPCE. A significant parameter is the mean path processing time. If the mean path processing time exceeds a certain threshold, the PCE VNF can deploy a new vPCE to reduce the peak request load in the PCE VNF.

Two scenarios might require intensive usage of a path computation element: a) intense dynamic usage of the network, or b) the need to perform in-operation network replanning or network recovery. Under these conditions, there appears the need for the deployment of vPCEs to perform the necessary path computations. The PCE computation load monitoring module will detect the end of the situation that has generated the need for path computation and turn down the unnecessary vPCEs.

There exist two different approaches for offering the running vPCE to be perceived as a single PCE by the different path computation clients. The first approach is described in [17.39] and consists in the usage of a PCE DNS to handle all the incoming requests and redirecting them to the allocated vPCE. The second approach is described in [17.41] and promotes the usage of the front-end/back-end PCE architecture in order to use a front-end PCE as a proxy for all the vPCEs deployed, which act as back-end PCEs. A more complete description of the different workflows can be found in [17.42].

17.6 Conclusion

This chapter has presented the optical network virtualization techniques, resource allocation algorithms, and NFV use cases that can be introduced in optical networks. Some of the benefits have been discussed, such as improvement of network usage, availability, and resilience. Importantly, some of the most significant difficulties in deploying virtual optical networks have also been discussed. These are the need for interoperability between network domains, the combination of SDN and NFV technologies (e.g., the lack of flexible support for end-to-end multisite installations), and consolidation of the initiatives to avoid the additional development needed to integrate the application/service on the platform.

References

17.1 M. Chiosi, D. Clarke, P. Willis, A. Reid, J. Feger, M. Bugenhagen, W. Khan, M. Fargano, C. Cui, H. Deng, J. Benitez, U. Michel, H. Damker, K. Ogaki, T. Matsuzaki, M. Fukui, K. Shimano, D. Delisle, Q. Loudier, C. Kolias, I. Guardini, E. Demaria, R. Minerva, A. Manzalini, D. Lopez, F. Ramon, F. Ruhl, P. Sen: Network functions virtualisation, an introduction, benefits, challenges and call for action. SDN and OpenFlow World Congress (2012). https://portal.etsi.org/NFV/NFV_White_Paper.pdf

17.2 M. Channegowda, R. Nejabati, M. Rashidifard, S. Peng, N. Amaya, G. Zervas, D. Simeonidou, R. Vilalta, R. Casellas, R. Martínez, R. Muñoz, L. Liu, T. Tsuritani, I. Morita, A. Autenrieth, J.P. Elbers, P. Kostecki, P. Kaczmarek: First demonstration of an OpenFlow based software-defined optical network employing packet, fixed and flexible DWDM grid technologies on an international multi-domain testbed. In: *Eur. Conf. Opt. Commun., Amsterdam* (2012), Th.3.D.2

17.3 Open Networking Foundation: OpenFlow Switch Specification, Version 1.5.1, ONF TS-025 (2015)
17.4 Open Networking Foundation: Optical Transport Protocol Extensions, ONF TS-022
17.5 ETSI: Network Function Virtualization (NFV): Architectural Framework, ETSI GS NFV 002, v.1.1.1 (2013)
17.6 V. López, O. González de Dios, B. Fuentes, M. Yannuzzi, J.P. Fernández-Palacios, D. López: Towards a network operating system. In: *Proc. Opt. Fiber Conf. (OFC)* (2014), Th3I.6
17.7 Z. Ali, S. Sivabalan, C. Filsfils, R. Varga, V. Lopez, O. Gonzalez de Dios, X. Zhang: Path Computation Element Communication Protocol (PCEP) Extensions for Remote-initiated GMPLS LSP Setup, IETF Draft, draft-ietf-pce-remote-initiated-gmpls-lsp-03 (2016)
17.8 R. Nejabati, E. Escalona, S. Peng, D. Simeonidou: Optical network virtualization. 15th Int. Conf. Opt. Netw. Des. Model. (ONDM) (2011). https://ieeexplore.ieee.org/abstract/document/5753389/
17.9 A. Fischer, J.F. Botero, M.T. Beck, H. De Meer, X. Hesselbach: Virtual network embedding: a survey, IEEE Commun. Surv. Tutor. **15**(4), 1888–1906 (2013)
17.10 L. Gong, Z. Zhu: Virtual optical network embedding (VONE) over elastic optical networks, J. Lightwave Technol. **32**(3), 450–460 (2013)
17.11 A. Pagès, J. Perelló, S. Spadaro: Virtual network embedding in optical infrastructures. In: *Int. Conf. Transpar. Opt. Netw* (2012), https://doi.org/10.1109/ICTON.2012.6253704
17.12 B. Mukherjee, D. Banerjee, S. Ramamurthy, A. Mukherjee: Some principles for designing a wide-area WDM optical network, IEEE ACM Trans. Netw. **4**(5), 684–696 (1996)
17.13 J. Zhang, J. Yuefeng, M. Song, H. Li, R. Gu, Y. Zhao, J. Zhang: Dynamic virtual network embedding over multilayer optical networks, IEEE OSA J. Opt. Commun. Netw. **7**(9), 918–927 (2015)
17.14 S. Zhang, L. Shi, C.S.K. Vadrevu, B. Mukherjee: Network virtualization over WDM and flexible-grid optical networks, Opt. Switch. Netw. **10**(4), 291–300 (2013)
17.15 A. Pages, J. Perello, S. Spadaro, G. Junyent: Strategies for virtual optical network allocation, IEEE Commun. Lett. **16**(2), 268–271 (2012)
17.16 J.M. Simmons: *Optical Network Design and Planning*, 2nd edn. (Springer, New York 2014)
17.17 B. Mukherjee, S. Ramamurthy, D. Banerjee, A. Mukherjee: Some principles for designing a wide-area optical network, IEEE ACM Trans. Netw. **4**(5), 684–696 (1996)
17.18 O. Gerstel, M. Jinno, A. Lord, S.J.B. Yoo: Elastic optical networking: a new dawn for the optical layer?, Commun. Mag. IEEE **50**(2), s12–s20 (2012)
17.19 H. Yu, V. Anand, C. Qiao, G. Sun: Cost efficient design of survivable virtual infrastructure to recover from facility node failures. In: *IEEE Int. Conf. Commun.* (2011), https://doi.org/10.1109/icc.2011.5962604
17.20 B. Chen, J.P. Jue, J. Zhang, S. Huang, W. Gu, W. Xie, Y. Zhao: Energy and spectrum efficiency with multi-flow transponders and elastic regenerators in survivable flexible bandwidth virtual optical networks. In: *Opt. Fiber Commun. Conf. Exhib.* (2014), https://doi.org/10.1364/OFC.2014.W2A.27
17.21 D. Simeonidou, R. Nejabati, S. Peng: Role of optical network virtualization in cloud computing, IEEE OSA J. Opt. Commun. Netw. **5**(10), A162–A170 (2013)
17.22 S. Peng, R. Nejabati, D. Simeonidou: Impairment-aware optical network virtualization in single-line-rate and mixed-line-rate WDM networks, IEEE OSA J. Opt. Commun. Netw. **5**(4), 283–293 (2013)
17.23 J.M.H. Elmirghani, L. Nonde, T.E.H. Elgorashi: Energy efficient virtual network embedding for cloud networks, J. Lightwave Technol. **33**(9), 1828–1849 (2015)
17.24 X. Wang, Q. Zhang, I. Kim, P. Palacharla, M. Sekiya: Virtual network provisioning over distance-adaptive flexible-grid optical networks, IEEE OSA J. Opt. Commun. Netw. **7**(2), A318–A325 (2014)
17.25 X. Cheng, S. Su, Z. Zhang, H. Wang, F. Yang, Y. Luo, J. Wang: Virtual network embedding through topology-aware node ranking, ACM Sigcomm Comput. Commun. Rev. **41**(2), 38–47 (2011)
17.26 R. Vilalta, R. Muñoz, A. Mayoral, R. Casellas, R. Martínez, D. López, V. López: Transport network function virtualization, J. Lightwave Technol. **33**(5), 1–8 (2015)
17.27 A. Mayoral, R. Vilalta, R. Muñoz, R. Casellas, R. Martínez: Experimental seamless virtual machine migration using an integrated SDN IT and network orchestrator. Proc. Opt. Fiber Commun. Conf. (OFC) (2015). https://ieeexplore.ieee.org/abstract/document/7121650/
17.28 Y. Yoshida, A. Maruta, K. Kitayama, M. Nishihara, T. Tanaka, T. Takahara, J.C. Rasmussen, N. Yoshikane, T. Tsuritani, I. Morita, S. Yan, Y. Shu, Y. Yan, R. Nejabati, G. Zervas, D. Simeonidou, R. Vilalta, R. Muñoz, R. Casellas, R. Martínez, A. Aguado, V. López, J. Marhuenda: SDN-based network orchestration of variable-capacity optical packet switching network over programmable flexi-grid elastic optical path network, IEEE OSA J. Lightwave Technol. **33**(3), 609–617 (2015)
17.29 D. King, A. Farrel: A PCE-Based Architecture for Application-Based Network Operations, IETF Internet Eng. Task Force (2015)
17.30 A. Aguado, V. Lopez, J. Marhuenda, O. Gonzalez de Dios, J.P. Fernandez-Palacios: ABNO: a feasible SDN approach for multivendor IP and optical networks, J. Opt. Commun. Netw. **7**(2), A356–A362 (2015)
17.31 Open Networking Foundation: SDN Architecture, Issue 1 (2014)
17.32 Optical Internetworking Forum (OIF), Open Networking Foundation (ONF): Global Transport SDN Prototype Demonstration, White Paper (2014)
17.33 Y. Lee, D. Dhody, S. Belotti, K. Pithewan, D. Ceccarelli: Requirements for Abstraction and Control of TE Networks, Network Working Group, draft_actn_requirement_02.txt (2015)
17.34 R. Vilalta, V. López, A. Mayoral, N. Yoshikane, M. Ruffini, D. Siracusa, R. Martínez, T. Szyrkowiec, A. Autenrieth, S. Peng, R. Casellas, R. Nejabati, D. Simeonidou, X. Cao, T. Tsuritani, I. Morita,

J.P. Fernández-Palacios, R. Muñoz: The need for a control orchestration protocol in research projects on optical networking. In: *Proc. Eur. Conf. Netw. Commun.* (2015), https://doi.org/10.1109/EuCNC.2015.7194095

17.35 R. Vilalta, R. Muñoz, R. Casellas, R. Martínez, F. Francois, S. Peng, R. Nejabati, D. Simeonidou, N. Yoshikane, T. Tsuritani, I. Morita, V. López, T. Szyrkowiec, A. Autenrieth: Network virtualization controller for abstraction and control of OpenFlow-enabled multi-tenant multi-technology transport networks. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2015), paper Th3J.6

17.36 S.S. Savas, M. Tornatore, M.F. Habib, P. Chowdhury, B. Mukherjee: Disaster-resilient control plane design and mapping in software-defined networks. In: *2015 IEEE 16th Int. Conf. High Perform. Switch. Rout. (HPSR)* (2015), https://doi.org/10.1109/HPSR.2015.7483086

17.37 R. Casellas, R. Muñoz, R. Martínez, R. Vilalta: Applications and Status of PCE, J. Opt. Commun. Netw. **5**(10), A57–A65 (2013)

17.38 E. Crabbe, I. Minei, J. Medved, R. Varga: PCEP Extensions for Stateful PCE, draft-ietf-pce-stateful-pce-09, IETF (2014)

17.39 Q. Wu, D. Dhody, D. King, D. López, J. Tantsura: PCE Discovery Using DNS, draft-wu-pce-dns-pce-discovery-06, IETF (2014)

17.40 J.P. Vasseur, J.L. Le Roux, Y. Ikejiri: A Set of Monitoring Tools for PCE-Based Architecture, RFC 5886, IETF (2010)

17.41 L. Velasco, A. Castro, D. King, O. Gerstel, R. Casellas, V. Lopez: In-operation network planning, IEEE Commun. Mag. **52**(1), 52–60 (2014)

17.42 R. Vilalta, A. Mayoral, R. Muñoz, R. Casellas, R. Martínez: Multi-tenant transport networks with SDN/NFV, J. Lightwave Technol. **34**(6), 1509 (2016)

Jie Zhang

Institute of Information Photonics and Optical Communications (IPOC)
Beijing University of Posts and Telecommunications (BUPT)
Beijing, China
lgr24@bupt.edu.cn

Professor Jie Zhang is the Dean of Institute of Information Photonics and Optical Communications at Beijing University of Posts and Telecommunications (BUPT), China. He received his PhD in Electromagnetic Field and Microwave Technology from BUPT in 1998. His research interests include all-optical networks and optical transport networks, network control and management, planning, optimization, and survivability of optical networks.

Ricard Vilalta

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)
Castelldefels, Spain
ricard.vilalta@cttc.es

Ricard Vilalta is a senior researcher at CTTC. He has led several international, EU, national, and industrial research projects. He is an active contributor in several standardization bodies such as ONF (OTCC), ETSI (NFV, ZSM), IETF, and ETSI OpenSource MANO (OSM). He has co-authored several book chapters, more than 40 journal articles, 160 conference papers, and 20 invited talks.

Xiaosong Yu

Institute of Information Photonics and Optical Communications (IPOC)
Beijing University of Posts and Telecommunications (BUPT)
Beijing, China
xiaosongyu@bupt.edu.cn

Xiaosong Yu received his PhD from Beijing University of Posts and Telecommunications (BUPT) in 2015. He is currently Assistant Professor at the Institute of Information Photonics and Optical Communications (IPOC) at BUPT. His research interests focus on elastic optical networks (EONs), spatial division multiplexing (SDM) optical networks, software-defined optical networking (SDON), data center networks, optical network security, and more.

Víctor López

Telefonica
Madrid, Spain
victor.lopezalvarez@telefonica.com

Víctor López is a Technology Expert at the Technology and Planning for Transport, IP and Interconnection Networks Department at Telefonica Global CTO. He obtained his PhD and worked as Assistant Professor at UAM (2009). He joined Telefonica I+D in 2011 as Technology Specialist, working on funded research projects from the Telefonica group and the European Commission.

Alejandro Aguado Martín

Center for Computational Simulation
Universidad Politécnica de Madrid
Madrid, Spain
a.aguadom@fi.upm.es

Alejandro Aguado graduated in mathematics and computer science from the Universidad Autónoma de Madrid in 2014. He is currently a researcher at the Center for Computational Simulation at Universidad Politécnica de Madrid working on quantum key distribution networking and for Telefonica GCTIO on its Software-Defined Transport Network (SDTN) strategy. He also collaborates at standardization institutes, such as ETSI and IETF.

18. Metropolitan Networks

Ashwin Gumaste

Metropolitan area networks, or MANs are at the confluence of business and home users—connecting enterprises to core networks and residential users to the rest of the Internet. This important segment of the network spans cities, regions, districts and municipalities and is a prime driver segment of broadband networking as well as being pivotal in providing connectivity to enterprises. In this chapter we begin by describing the premise of technology in metro networks (shortened form of MANs). After a detailed overview of how metro networks plug into the larger service provider scheme of things, we delve into individual technologies that are intrinsic to metro network architecture. These technologies include SONET/SDH, OTN, optical networks, WDM, IP/MPLS and carrier Ethernet. Each technology is described from the metro standpoint and how it is used as a service offering medium. We then focus on futuristic technologies such as SDN and NFV. A detailed guide towards best practices for provider networks summarizes the chapter.

Since the development of humanity into industrial zones, there has been a migration of humans to population centers or cities. Cities represent metropolitan regions where large populations reside. Metropolitan regions include residential areas, businesses, government, public services, international landing sites, public utilities. These entities interact with each other, requiring a robust networking backbone. Given the number of people and business in such cities, the metropolitan area network represents perhaps the most important aspect of a provider's network. In fact, in [18.1] it was stated that the revenue for providers from metro networks is almost 50% of their overall revenues, and constitutes the bulk of enterprise traffic, mobile backhaul traffic as well as inter-data-center traffic. A key aspect of the metro network architecture is that is requires ultimate carrier-class service features such as 99.999% availability [18.2], and excellent reliability and manageability.

The metro network began as a layer 0-3 network with an optical backbone, over which layer 1 transport and layer 2 framing resided. An IP/MPLS overlay made up for the routing protocol. Select points of presence (POPs) along the metro network lead to distribution networks from where various types of access networking technologies are manifested. These access networking technologies could be coaxial cable, hybrid copper-fiber cable, various types of passive optical networks, various types of wireless technologies and light-over-air communication systems [18.3]. One technology that is no longer growing, but where things historically began is the plain old telephone system (POTS), which connects RJ11 based phone-lines to an exchange and multiples of such exchanges are mapped on to a metro POP. Enterprises are connected to the metro either via this type of distribution network or directly to a metro POP.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_18

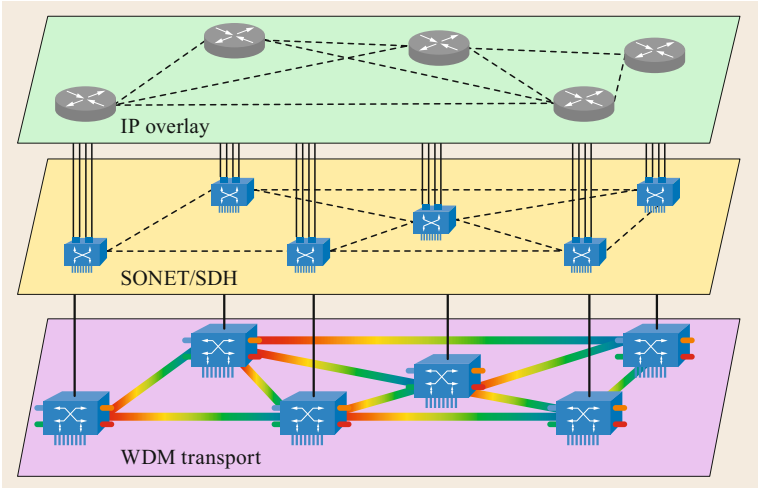


Fig. 18.1 Multiple layers of the metro hierarchy

The range of communication technologies and the spectrum of customers connected to the metro network implies that this domain manifests multiple types of services ranging from legacy voice, video, and data to enterprise services, data-centers, clouds etc. The service domains in the metro network have grown significantly over the past decade or so. The advent of the cloud has meant that cloud networking directly fits into a metro network and with virtualization technologies metropolitan area clouds have become a mainstay of today's online businesses or business processes. It is in fact estimated that more than 80% of all data is in the cloud, and most of these clouds are regionally active, even though they may be globally present [18.4]. Architecting a cloud on the top of a metro network is seen as both a challenge and an opportunity for service providers, equipment vendors, and network system integrators. Figure 18.1 shows a typical multilayer provider hierarchy.

18.1 Metro Transport Network Architecture

Shown in Fig. 18.2 is an architectural depiction of a classical service provider network spanning across a country or a large region. The network is roughly divided into four areas of operation—an access network, a metropolitan network, a core network and a data-center network. This network layout is in line with leading service providers' visions such as Domain 2.0 from AT&T [domain2.0] and also includes new technology concepts like SDN [18.5]. The access network consists of a plethora of technologies such as wireless, cable, fiber and powerline [18.3].

Much of the access is dominated by the last mile network, where wireless plays the role of providing ubiquitous services and fiber provides raw bandwidth pipes and provisioned services to homes and enterprises. The access network terminates in a central office CO, and multiple COs are backhauled into a metropolitan point of presence (POP). Multiple metro POPs are connected to each other via a ring network that manifests as the principle metropolitan area network (MAN). The CO-to-POP interconnection is across a layer 3 router which could be a provider edge (PE) router or an edge router (ER). Both PE and ER are similar routers, with different nomenclature used in different networks. For example, in a provider domain, PE is more prevalent, while in an enterprise the ER name dominates. The access network may have its own edge router to facilitate layer 3 services and terminations from end users and enterprises. The metropolitan network connects multiple POPs and also serves as a collector of traffic that is routed to a core network. The core network may be a nationwide backbone or an international landing point. The metro network has multiple technologies at each layer. At the fiber layer there is a need to transport big chunks of data, and hence the use of SONET/SDH, which is now being replaced with packet-optical integration with technologies such as carrier Ethernet and OTN. These technologies are further mapped to optical wavelengths that are transported by optical add-drop multiplexers at nodes. By making these add-drop multiplexers reconfigurable we are able to achieve end-to-end optical communication within the metro domain. To this end, the ROADM (reconfigurable optical add drop multiplexer) technology has been in use. The inter-

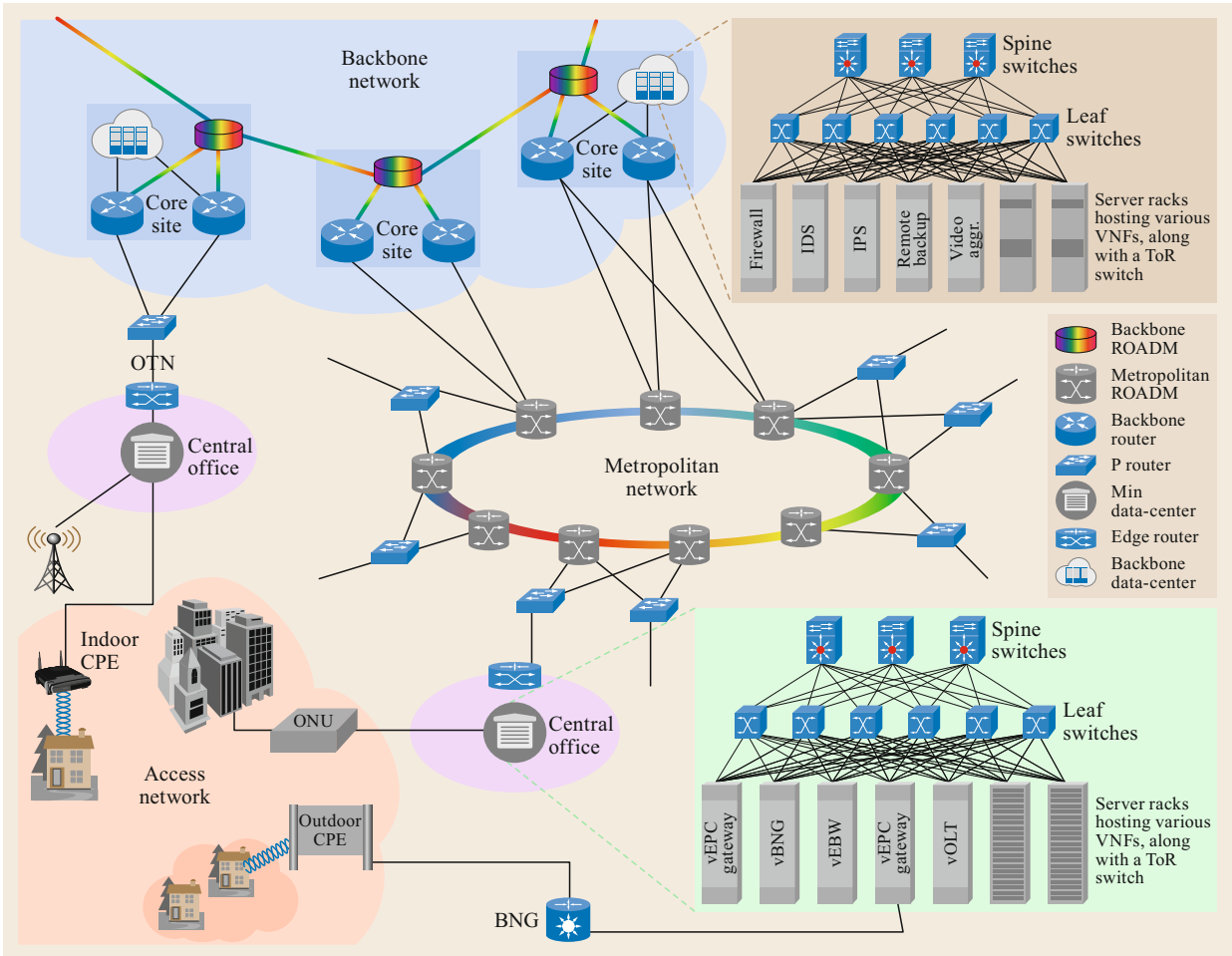


Fig. 18.2 A classical 3-tier network

connection point between the metro and core is usually through a P-router or a provider backbone (core) router, which is also sometimes called a core router or a border router (BR). This router runs IP and MPLS protocols and can support several 100 Gbps or a few Tbps of switching/routing capability.

In the above-mentioned architecture, we also have data-centers at core and edge locations, together forming internet clouds and application clouds. This type of a multicloud environment facilitates virtualization and automation of technologies, protocols and data. We now discuss the different technologies that make up a metro network. Figure 18.2 depicts a classical three tier provider architecture with data-centers as well as movement of such an architecture towards a new technology such as SDN/NFV.

18.2 SONET/SDH

We now describe the metro network architecture that has evolved over the past decade or so. As the Internet exploded as a connectivity tool in the mid-1990s, POPs in the metro environment had to be connected with transport technologies. Wireless did not have sufficient bandwidth, while copper and coaxial cable had significant issues of reach. Fiber seemed the direct and natural choice as a medium for transport of data between POPs in the metro. The first fiber technology to appear in the metro was based on SONET/SDH. SONET (synchronous optical network) and its Eurasian cousin SDH (synchronous digital hier-

Table 18.1 SONET/SDH designations and bandwidth [18.6]

SONET optical carrier level	SONET frame format	SDH level and frame format	Payload bandwidth	Line rate (kbit/s)
OC-1	STS-1	STM-0	50 112	51 840
OC-3	STS-3	STM-1	150 336	155 520
OC-12	STS-12	STM-4	601 344	622 080
OC-48	STS-48	STM-16	2 405 376	2 488 320
OC-192	STS-192	STM-64	9 621 504	9 953 280
OC-768	STS-768	STM-256	38 486 016	39 813 120

archy) were both excellent technologies for information transport.

SONET/SDH evolved as a need for maintaining the TDM hierarchy for carrying voice communication. It is not incorrect to say that SONET/SDH was more of a direct need to transport voice lines as they grew in number and in distance. SONET/SDH facilitated as a direct infrastructure that could be readily used for transport of non-voice services, such as those that were prevalent in the metro environment. The entire SONET/SDH framework was developed on the TDM concept, whereby several lower rate lines were time-division multiplexed to form a single higher rate line, which became the client signal for further multiplexing at the next level. This sort of multiplexing led to a hierarchy of line-rates that facilitated efficient transport of data from one place to another in a network. SONET/SDH has the peculiarity that each multiplexing hierarchy exceeds the previous one by a factor of four, which means that if the base line-rate of a channel is A bits per second, then four such lines are multiplexed together to create the next hierarchical channel, which is approximately at $4A$ bits per-second, though in reality, the line rate is slightly higher than $4A$ bits per second due to extra protocol bits used for monitoring, resiliency, restoration etc. Shown in Table 18.1 is a set of line rates supported by SONET/SDH. The base SONET line-rate is 155 Mbps and this line is called an OC-3 (optical circuit 3). The corresponding equivalent line-rate in the SDN domain is called the synchronous transport medium (STM). SONET/SDH grows in multiples of four, with the next line rate approximately 622 Mbps or OC-12, and then 2.488 Gbps or OC-48, followed by 10 Gbps or OC-192 and subsequently 43 Gbps or OC-768. Multiplexing in SONET/SDH follows an interesting pattern, whereby frames are multiplexed row-wise from left-to-right and top-to-bottom. The advantage of such a technique is the spatial equivalence produced due to the uniform structure of the frame which helps in easily identifying the constituent signals.

Initial deployments of SONET/SDH systems were linear or point-to-point networks with four fiber-based systems that used two fibers for working traffic and two fibers for protection traffic (which was used in the event of a failure). Each pair of fibers was used for duplex communication. This sort of protection was termed unidirectional path switched ring (UPSR), and bidirectional line switched ring (BLSR). The concept of path and line for UPSR and BLSR shall be explained subsequently. The problem with such an approach was that the work and protection routes had to be geographically disjoint, or else a failure such as a cut in location would result in all the traffic going down.

SONET/SDH communication networks are characterized by a three-layer model. The three layers are path, line and section. The layers are intertwined. Multiple sections make up a line, and multiple concatenated lines make up a path. The section is the smallest indistinguishable part of a SONET/SDH network. Regenerators that rejuvenate an optical signal by converting it into an electronic signal and re-transmitting the signal back into the optical fiber are used to connect two adjacent sections. The line-rate on either side of a regenerator is usually the same, implying that the multiplexing hierarchy is well maintained. Multiple concatenated sections are part of a line. A SONET/SDH line consists of add-drop multiplexers at either end. A SONET/SDH ADM consists of optics and electronics that can create, process, transmit and receive a SONET/SDH signal.

A path consists of several lines and sections concatenated together. A path also consists of the multiplexing hierarchy. At the ends of a path, we have SONET/SDH ADMs that can add or drop slower tributaries of traffic by taking advantage of time-division multiplexing principles.

SONET/SDH networks due to the TDM hierarchy are ideal for voice centric communication. Over the years, the SONET/SDH community has directed efforts towards making SONET/SDH more amenable to data-communication. These efforts are called VCAT and LCAS (virtual concatenation and link channel adjustment scheme). However, with the growth of data-traffic SONET/SDH is being actively replaced by packet-networks. Most modern day metro networks have retired SONET/SDH, and it is a poor design choice to inculcate SONET/SDH in contemporary/future networks, given the massive growth of data and video traffic. An anatomy of an SONET/SDH circuit is as shown in Fig. 18.3.

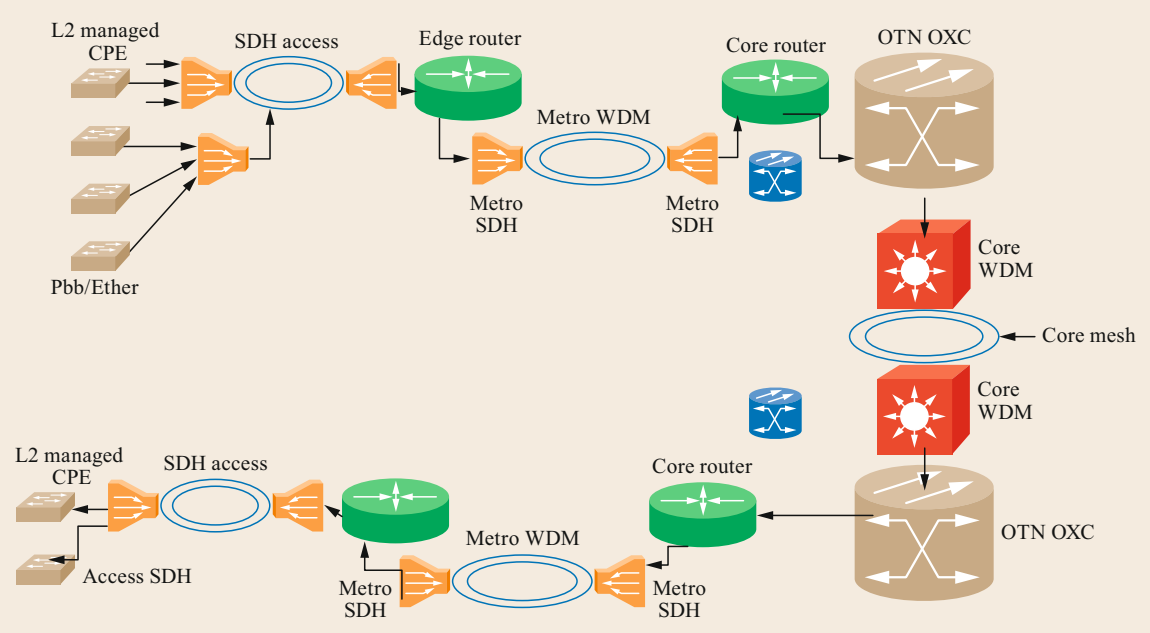


Fig. 18.3 End-to-end anatomy of a circuit

18.3 Optical Transport Network (OTN)

The ITU G.709 or optical transport network (OTN) is perhaps one of the most important standards that governs the transport of data through the Internet, especially in metro and core networks. The OTN has evolved as a telecommunication class transport mechanism that does away with some of the non-data-centric issues posed by SONET/SDH and paves the way to meet the future of transport technology such as for high line-rates, flexible granularity support and doing away with basic TDM hierarchy while providing for excellent OAMP capabilities. The OTN philosophy is well adapted to meet the changing data paradigm with requirements of IP over WDM, enabling IP routers to have OTN compatible interfaces and facilitating IP data to be mapped to wavelengths in a WDM network. The OTN is a true telecom-class transport protocol that was designed as a mechanism to transport contemporary services as well as serve as an enabler transport platform for emerging services.

The OTN has evolved as an impressive transport technology—ideal for replacing SONET/SDH—taking into consideration the requirements of *both* the overlay and the underlay network. The overlay of IP, fiber-channel, and Ethernet services is characterized by somewhat bursty, packet-oriented or user-defined and sub-wavelength bandwidth granular requirements. Mapping the plethora of overlay protocols on SONET/SDH has been shown to be inefficient as well as cost-wise unaffordable. SONET/SDH was designed as a hierarchy to transport TDM signals whose base line-rate was a DS0 (a digital signal for base communication at 64 kbps—implying that a DS-x would be x times 64 kbps). With the fast proliferation of IP and revenue bearing services now pushing the limits of applications such as storage networks and data centers, and the importance of sustaining a hierarchy based on the basic DS0 signal is rapidly vanishing. Mobile backhaul, especially in the 3G/4G/5G variants also has traffic that is bursty, yet with newer technology, such as required with 5G, there is a need for strong latency guarantees, implying a corresponding need for a stable transport layer [18.7].

The underlying network is primarily based on WDM technologies. Mapping user defined signals to WDM is possible using a multitude of protocols, such as 10GigE/100GigE, SONET/SDH and even OTN. The OTN however, is especially well-suited for the WDM layer—it was designed keeping in consideration the fact that WDM would be its primary physical layer exponent. The OTN provides a mechanism to transport signals directly over lambdas—a unique feature considering the fact that variable rate client signals can

now be transported and effectively recovered through a WDM network. The OTN provides for digital monitoring techniques and forwarding equivalence class (FEC)—both of which act as force multipliers leading to higher line-rate per wavelength. The future of the data layer is actively moving towards Ethernet—and as this happens, OTN has a strong role to play—specific OTN mechanisms are available to transport 10 Gbps, 40 Gbps Ethernet, and 100 Gbps Ethernet through the MAN and WAN. The architecture of OTN also works well with ROADMs. ROADMs are the mainstay of optical networking in the metro and core. The OTN-ROADM combination is especially useful when we consider multi-degree ROADM hubs that allow seamless movement of wavelengths across any two ports. The per-channel monitoring capabilities induced by the OTN layer enable us to provision end-to-end all-optical circuits over a ROADMized core relegating the per-circuit management to the OTN layer.

The aforementioned discussion makes a compelling case for OTN as the transport technology of the future, perfectly mapping to both WDM below it and the vast number of overlaid protocols that bring in traffic.

18.3.1 OTN Workings

The emergence of WDM technology has allowed good utilization of the installed fiber-plant enabling a provider to add new services on existing fibers efficiently. Critical components like ROADMs based on wavelength selectable switching (WSS) technology furthered the use of WDM networks with the flexibility that the service providers needed in designing the network to accommodate emerging demand-based services. Service providers are now subject to the need for a new transport layer infrastructure that:

1. Supports the transport of a diverse range of client signals (arising from the need to support a diverse set of services such as mobile-backhaul, video, data-center, and storage networking)
2. Is flexible to accommodate the varying bandwidth granularities
3. Is agnostic to client signal types—thus enabling the use of a single transport layer for all service types
4. Enables inter-networking of equipment from different vendors and interworking of networks from different carriers.

The contemporary solution of using SONET/SDH as the transport layer has limitations in addressing many of these requirements. Besides SONET/SDH was not designed for the kind of emerging services that are seen on the horizon today. Primary limitations include its rigidity to support transparent transport of local area network (LAN) type signals and limitations in supporting network-wide provider operations.

The optical transport network architecture defined by ITU.T. G.872 consists of the following three layers: an optical transport layer (OTS), an optical multiplex layer (OMS) and an optical channel layer (OCh). The optical signals at each of these layers are terminated at different points in the network. For example, the OTS is terminated at every optical-to-electrical-to-optical (OEO) regenerator while the OMS is terminated at the elements of an optical network which in most cases is the ROADM. The OCh layer maps onto a wavelength, and its termination is at the point where the OMS is terminated for adding or dropping the client signals. Overheads are added at each of these layers and are essential for provisioning, recovery and management. While client signals of any type—SONET/SDH, IP, Ethernet etc.—could be directly transported over a wavelength (OCh), service providers needed support for individual wavelength monitoring capability and full OAMP. The native optical layer proves to be expensive to meet the needs of full OAMP support, especially for fractional-wavelength granular flows.

The ITU-T G.709 recommendation defines, interfaces to the OTN [18.8] that expands the OCh to add per-wavelength OAMP support. It makes use of the OEO conversion that is currently required at the 3R (reshaping, retiming and reamplification of the signal) regenerator points in the network to provide OAMP capabilities in the electrical domain. The G.709 recommendations in addition to adding OAMP related overheads to the client signal at the wavelength level, adds FEC that helps in reducing the span between 3R regeneration points resulting in OPEX savings. It also provides a layered structure comprising of optical channel transport unit (OTU), optical channel data unit (ODU), and optical channel payload unit (OPU) layers for mapping client signal payload. Note the following factors:

1. Overheads at each of these layers
2. Support for mapping a wide variety of client signal types—legacy TDM and emerging packet services
3. The ability to support any client signal rate—rigid SONET rates and packet data at non-SONET rates
4. The per-wavelength OAM&P capability that is agnostic to client signal type—that make OTN a perfect transport layer for next generation networks that require converged transport platform that exploits the bandwidth potential offered by DWDM technology.

G.709 OTN: The OTU layer defined in OTN is the electrical equivalent of the OCh. Supported OTU*k* line rates

are 2.7, 10.7, 40 and 111.8 Gbps. These rates are specifically selected to accommodate SONET/SDH client payloads at 2.5, 10, 40, and 100 Gbps rates. The OTU layer signal corresponding to these rates were OTU1, OTU2, OTU3, and OTU4 respectively and the common representation is OTUk where $k = 1, 2, 3,$ or 4.

The OTUk layer is the section layer equivalent of SONET/SDH and is terminated along with the OCh. The OTUk is hence the innermost part of a connection. The client payload is mapped into the OPUk payload and the information about payload type and rate justification is added in the OPUk overhead.

The ODUk layer is equivalent to the path layer in SONET/SDH and it encapsulates the client data in the OPUk layer with overheads that provide necessary information for end-to-end supervision.

The ODUk path layer data rates are defined in the ITU.T. G.709 recommendation and includes three SONET/SDH rates of 2.5, 10, and 40 Gbps. The second revision of the recommendation extended these to include an even higher 100 Gbps rate (ODU4) client payload, as well as a lower 1.25 Gbps (ODU0) rate payload, and finally a flexible container (ODUFlex) for constant bit rate (CBR) payload at any rate above 2.5 Gbps. The ODU0 rate is added to support bandwidth efficient transport of a native 1 Gigabit Ethernet signal and any CBR signal of rates less than 1.25 Gbps (the ODU0 rate). The ODU2e rate has been added to transparently transport the 10G BASE R LAN signal. ODUFlex was added to handle the data traffic requirements arising from services with payloads whose rates do not match the SONET/SDH or Ethernet rates. The addition of these rates, make OTN a converged transport layer. *Note that ODU0, ODU2e, and ODUFlex signals do not have a corresponding OTU layer signal.* These are mapped into the payloads of higher order ODUk where $k = 1, 2, 3,$ or 4 and transported over the corresponding OTUk layer signal.

The G.709 standard also supports TDM of lower rate ODUj signals, into a higher rate ODUk signal where $k > j$. This provides the sub-wavelength networking capability as will be shown in one of our use cases. For this purpose, the first revision of the standard divided the higher rate ODUk signal into multiple ODU1s—for example, ODU2 is divided into four ODU1s and ODU3 into eight ODU1s. The second revision of the standard supports dividing all higher rate ODUk signals into 1.25 Gbps time slots, corresponding to ODU0 rate.

OTN cross connect: OTN is being used as a solution to create OTN cross connects, whereby ODU-level switching is facilitated. Large OTN based cross connects are being developed and deployed. OTN based systems score over IP/MPLS or carrier-Ethernet (CE) systems for wide area networking on two counts: (1) The reach enhancement and monitoring of channels due to OTN based FEC and tandem channel monitoring like features. (2) Creating mesh based restoration within 50 ms. In many high-speed networks across the providers there are significant deployments of OTN cross connects that aid in the transport network. OTN as a technology has significantly matured to meet the requirements of telecom class networking.

One of the key reasons to deploy OTN in networks is to map multiple protocols seamlessly in a carrier domain. This facilitates networks to cater for disparate user-needs in terms of service provisioning while adhering to a carrier-class 50 ms restorative backbone network. 50 ms is critical in the telecom domain as that is the maximum time that a voice circuit can survive a failure without the human ear detecting the failure. What this means is that the perceptive capability of the human ear is 100 ms, half of which is the time we allocate for rerouting a signal post-a-failure and hence the time-duration of 50 ms is sacrosanct in the telecoms world.

18.4 Metro IP/MPLS

The key layer 3 protocol has always been IP. IP began as a best-effort datagram paradigm to connect nodes and facilitate distributed routing. IP was enhanced by routing protocols such as IGP and OSPF [18.9] that have been able to provide intra-domain routing capabilities using distance-vector and link-state techniques respectively. Routers periodically exchange messages by pinging each-others' specific TCP ports. The responses facilitate routers to know if the link is active and its parameters. Based on this information the routers further exchange the longest prefix information thereby building forwarding tables for IP addresses. This sort of routing and forwarding is based on the longest-prefix match techniques. IP routers were always subtended with a strong layer 2 network and corresponding interfaces. Contemporary routers have SONET/SDH, Ethernet and OTN interfaces. Some of these interfaces may be colored—exemplifying the use of WDM technology.

IP routers began as simple central processing unit (CPU) based manifestations in the early 90s with a few thousand entries in its forwarding tables. However, with

the prolific growth of the Internet in the mid 90s, there was a need for larger routers. Just around that time, VLSI technology had matured and there was an abundance of application-specific chips in the market. Taking advantage of this development, the industry imbibed VLSI technology to build IP routers. As interface rates grew along with forwarding table sizes, the size of the IP router also grew. What began as a 10 Mbps capable IP router became a gigabit router towards the end of the last century.

The advent of VLSI technology and its use in designing custom ASICs was a major boost to router design and development. ASIC or merchant silicon based designed began it's manifestation in various forms and towards various functions. The switch ASIC was one of them. Using switch ASICs [18.10], one was able to design large capacity routers that facilitated huge movement of data between ports. The reliance on CPUs had ceased in router design.

A typical router consisted of multiple IO cards that were connected to a routing or switching fabric. IO cards could now connect to a variety of protocols such as SONET/SDH interfaces, 1, 10, and 100 Gigabit interfaces and OTN interfaces. Furthermore, all these interfaces could have gray optics for short ranges (10 km) or colored optics for longer ranges (40 km and above).

The switching card of a router had two premier innovations: the ternary content addressable memory (TCAM) [18.11] and the switch fabric itself. The TCAM was a major development that facilitated large port count and high throughput routers. TCAM chips could map forwarding entries to ports and to service classifications. The challenge prior to TCAMs was to facilitate fast selection of entries for mapping incoming packet header specifics (such as IP addresses etc.) to corresponding forwarding information (output ports etc.). Traditional memories such as dynamic random-access memories (RAMs) were plagued by latency issues. Static RAMs (SRAMs) on the other hand were fast memories but were often plagued by size issues. TCAMs in contrast could be much larger in depth and routinely manage several thousand forwarding entries.

Router IO cards dealt with packet parsing which also included header stripping, processing and forwarding. Packet parsing includes checking packets that arrive from IO ports and then stripping the header after the correcting of a packet is achieved. The header is then mapped against tables and metadata is created that is forwarded to a routing/switching fabric.

Multiple IO cards are connected to a redundant switching card in a Clos fabric fashion as shown in Fig. 18.4. Usually a router would be designed such that it is completely non-blocking. By non-blocking it is implied that if an output port is free and an input port desires to transfer packets to this free output port, then this connection would be supported. The current state of the art is such that terabit capacity routers are not uncommon. In such routers, usually per-slot a 400 Gbps fabric is made available. For such an IO slot, one may have a combination of 10, 40, and 100 Gbps IOs. A typical IO slot may have 4×10^2 Gbps IOs, or 48×10^1 Gbps ports etc. It is expected that in the near future a terabit per second per slot of IP processing would be easily achievable.

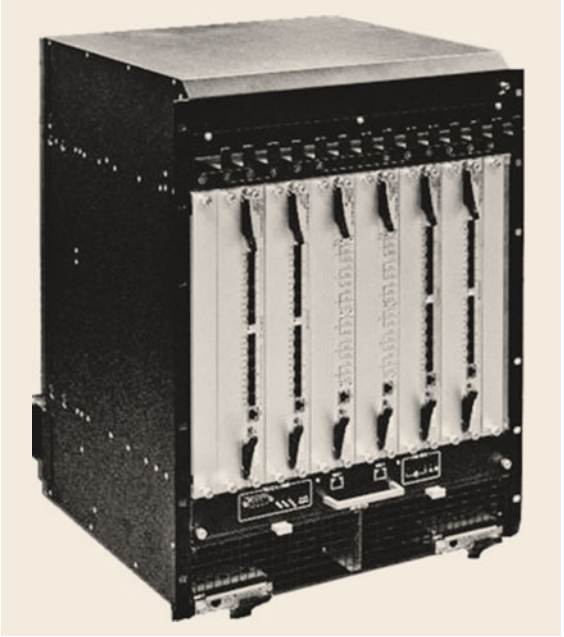


Fig. 18.4 Sample IP-SDN router. (Image courtesy of Nokia)

Routers are measured in terms of their throughput in addition to the latency that they offer on a port-to-port basis. Another figure of merit is the QoS that can be accorded to flows through a router. Currently four or eight QoS levels are common in provider networks. Another important characteristic of IP routers is to provide rate-limiting and access control list (ACL) support. Rate-limiting allows users to ensure that flows entering a router are limited based on earlier arrangements between a customer and a provider. ACLs facilitate IP address or some other parameter based forwarding, dropping or some other QoS manifestation as may be the case. ACLs require significant time for processing and are known as a bottleneck in the forwarding speed of a router. Conversely, the number of ACLs supported by a router is a good parameter to indicate router stability and performance.

18.4.1 Multiprotocol Label Switching

MPLS was defined with the aim of simplifying packet forwarding in IP networks and to address core IP router performance issues [18.12]. MPLS uses label switch paths (LSPs) to define the path followed by packets in the network. The use of MPLS labels enables routers to avoid complex IP-like processing overhead. Major advantages of MPLS technology over IP are speed, scalability, QoS support using resource reservation, and traffic engineering capability. MPLS fast re-route (FRR) [18.13] supports 50 ms restoration of services after failure. However, originally IP/MPLS technology is not purely carrier-class (on account of deficient OAM characteristics) and IP/MPLS routers turn-out to be expensive. As the technology has evolved, IP/MPLS has acquired some of the carrier class features that make it a good fit in large provider networks, including the metropolitan area.

MPLS manifests itself as label switched paths that are created using labels assigned to flows/packets and are forwarded/switched using labels. Labels are inserted at an ingress node, which is also called a label edge router (LER) (for that particular flow). Once a packet flow is classified by an LER and a label is inserted on to the packet, then the flow is forwarded thereafter based only on the inserted label. Labels are based on a concept of forwarding equivalence class or FEC. The idea is to build a label that corresponds to a longest prefix match IP address and ties this longest prefix match address to an output port at a router. Naturally, FECs are generated in the reverse direction to the flow and are thereby updated at each node as they traverse to a particular source node. FEC to port mapping is locally valid but the FEC mapping itself is valid for all the nodes along the path. FECs could be announced by downstream nodes to upstream nodes declaring a label for a group of addresses, or FECs could be requested by upstream nodes. A node maintains the next hop label forwarding entry (NHLFE) that is essential to compute the appropriate output port on which to forward a packet. For assigning labels a label distribution protocol (LDP) is used. In the past there were two variants of the LDP—the constrained routed LDP (CR-LDP) and the resource reservation protocol (RSVP). Currently it appears that most LDP variants now are based on RSVP, and in particular RSVP has been refined to include traffic engineering (TE) as an extension, thereby calling it RSVP-TE [18.14].

The MPLS label sits just after the L2 header, which is critical to demarcate a packet from the earlier packet. In that sense, due to its situation just after the L2 header, the MPLS protocol suite is generally considered as an L2.5 protocol [18.15]. The MPLS label is a 32 bit entity, of which is a 20-bit shim (label identifier) that uniquely distinguishes a label from another label. Shim values need not be globally unique but must be router-wise unique. In addition to the shim is a three bit QoS value that accords eight priority levels to the packet. A single bit EXP (experimental) value is used that can further enhance QoS to 16 priority levels. An eight bit time-to-live (TTL) is also part of a label and when a packet enters an MPLS domain, the TTL value from the IP header is copied into the TTL placeholder in the MPLS label. When a packet passes through an LSR, the TTL value is decremented. When a packet leaves an MPLS domain or if an MPLS label is removed, then the TTL value is copied back into the IP header. In MPLS, there are several key functions that are used to act upon a packet. Among these the three most important ones are pop, push, and swap a label. When a packet enters an MPLS domain an incoming label map (ILM) checks if the packets' IP address or port or VLAN tag can be mapped to an existing label entry. If an entry exists in the forwarding database of the LER, then the corresponding label is inserted onto the packet. This sort of an operation of inserting a label on an unlabeled packet is called push. The packet now travels along a label switched path across multiple core routers called label switched routers (LSRs). At each LSR, the labeled packet is fed to a label forwarding table, which stores a forwarding information base (FIB). The FIB is a standard table that contains prefix and label mappings to ports. At the penultimate node along a label switched path, the packet is stripped of the MPLS labels and then forwarded as the original non-encapsulated packet as was received in the MPLS domain. This is done so that the last node need not always support MPLS—thus facilitating efficient interoperability between an MPLS and IP network. This type of penultimate hop popping of labels is a key characteristic of MPLS. At times an intermediate LSR might feel that it has a better route to the destination than is earmarked by the packets' existing label hierarchy. In such a case the intermediate LSR replaces the outermost label with a more appropriate label that could imply a shorter path for example. This sort of replacement of an existing label with another label is called label swapping. The act of removing a label from a packet is called label popping. These three operations together, i.e., push, pop, and swap define MPLS label operations. Another key operation that defines MPLS is path-merging. Two paths going towards a common destination can be merged at an interconnection point using label swapping.

MPLS networks have dominated much of provider layer 2.5/3 networks worldwide. One of the key aspects of MPLS that makes the technology important to providers is traffic engineering. Traffic engineering allows a provider to plan network traffic such that pain-

points are avoided and traffic guarantees are met in an efficient manner. Traffic engineering allows users to choose routes thereby optimally using network resources and avoiding congestion. Traffic engineering is especially important to ensure that service level agreements are met in a cost-effective manner. Traffic engineering facilitates many important characteristics of modern provider networks. Among these congestion control, fast re-route after failure, customer priority identification are some of the features. Load-balancing can easily be achieved by inculcating traffic engineering into an MPLS domain. A load balanced network, can be set up using traffic engineering by routing multiple streams (LSPs) between the same source-destination pair across different routes thereby completely balancing the network.

An important aspect of MPLS networks is protection. There are many techniques of providing protection to MPLS networks. The most commonly used is called FRR or fast re-route. Traditional protection techniques require that the source and destination signal to each other about a prospective failure, subsequent to which traffic is moved from a work path to a protection path. This process or the derivatives of this process take significant amounts of time to converge especially for large networks, and hence it is difficult to achieve 50 ms carrier-class protection. Fast reroute alleviates this signaling problem by precomputing restoration paths for node or link failures and routing locally, wherever a failure has occurred. The dual process of local bypass of the failed apparatus along with the relaxed need for signaling leads to faster restoration times.

Despite its advantages, there are some issues with traffic engineering, such as the requirement for manual set up of LSPs, performance degradation post-FRR and protocol dependency. Some of these can possibly be reduced using an automatic control plane like the generalized MPLS (GMPLS) control plane but studies have shown that scalability and cost issues exist with adoption of such an approach.

Carrier-class MPLS: It is imperative to discuss whether MPLS is carrier-class from a provider perspective. MPLS per-se is a best-effort service, however, with the add-ons of traffic engineering, there is a significant use of MPLS in provider networks. As far as carrier-class services are considered we define these with three parameters:

1. Deterministic latency across a domain irrespective of traffic load.
2. Measurable parameters of the provisioned service such as ability to bill, manage, provision, and operate the service and network.
3. 50 ms protection.

The latter two features are well integrated in the MPLS framework, however, due to its forwarding plane behavior, an MPLS LSR unless well provisioned through manual traffic engineering does exhibit a best-effort characteristic, especially for the scenario of large networks with heavy traffic. To make this kind of data-plane carrier-class the IETF and ITU are standardized MPLS-TP (transport profile) [18.16]. This is described later.

18.5 Metro Optical Network

One of the most important parts of the metropolitan network is the optical network as it carries all the traffic between enterprises and various points of presence. The optical network has traditionally been implemented using optical equipment, and since its inception the goal has been to attempt to switch data at nodes in an all-optical manner. All-optical switching of data has its advantages in the sense that the bandwidth made available by the fiber is of the order of 1–2 Tbps leading to fast, low cost communication. It is well-known that keeping data in the lower strata of the network, i.e., the optical layer leads to lower cost-per-bit, lower vulnerability and lower energy consumption. Hence it makes significant sense to keep the data in the optical layer. However, using all optical processing has some inherent limitations. To begin with, the rate at which a laser can be modulated using external modulation formats is significantly lower than the bandwidth that a fiber offers. Even with techniques such as coherent communication and use of orthogonal frequency division multiplexing (OFDM) and Nyquist modulation techniques there is a significant gap between the optical offered bandwidth and the electrical techniques available. This difference between optical and electrical bandwidth is known as an opto-electric bottleneck and one way to resolve this is to divide the optical spectrum (in the 1.5 μm range) into channels and individually sending data on these channels. This kind of multiplexing is termed wavelength division multiplexing (in the C and L band from 1520–1590 nm or so). If the channels are far-apart say 20 nm apart then the technique is further classified as coarse wavelength division multiplexing (CWDM) and if the channels are spectrally packed close to each other with 0.4 or 0.8 nm

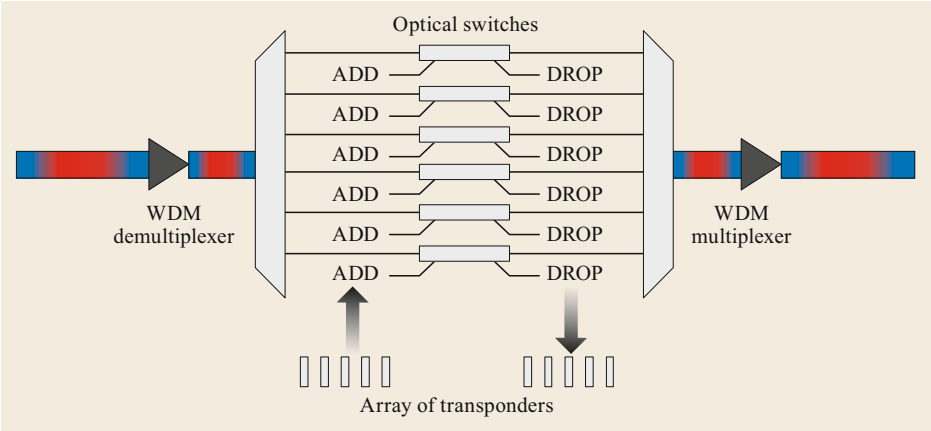


Fig. 18.5 A second generation ROADMs with AWGs

spacing, then the technique is called dense wavelength division multiplexing (DWDM).

The next issue is transporting data. There are two ways of transporting data through an optical network—as packets and as circuits. Naturally, due to the immense use of IP datagrams at the network layer, there is a desire to use packets in the optical layer. Optical packet switching however, is a technology in its infancy on account of the inability to decode packet headers in the optical domain as well as provide extremely fast switching that would make packet-switching viable. As a result, almost all commercial deployments today focus on optical circuit switching technologies. Optical circuits or lightpaths are set up between a source and destination node and it is desired that the circuit be all-optical from the source to the destination so as to avoid expensive opto-electro-opto (OEO) conversion and regeneration of the signal. Power, chromatic and polarization-mode dispersion and noise resulting in lower optical signal to noise ratio (OSNR) are the three principle ailments that impact an optical signal. An important innovation that can boost up a signal power is the erbium-doped fiber amplifier (EDFA) that can transfer the power from a pump signal to the signal and noise spectrum. The EDFA does amplify the signal but also amplifies the noise associated with the signal and thus a good figure of merit to determine reachability of a signal in conformance to a particular bit error rate is the OSNR. Even when we assume that a signal can reach from a source to a destination node, it may have to go through several intermediate nodes that need to switch this signal from the appropriate ingress fiber to the egress fiber and in some cases from an ingress wavelength to a corresponding egress wavelength. To do so, there have been various devices proposed for providing interconnection switching. Most of these devices provide wavelength-transparent switching. The first generation of such devices included a patch panel that required operators to switch manually between ingress and egress ports. This process was both time-consuming (inefficient) and not sustainable against dynamic traffic variations. The direct impact of a patch panel was on provider personnel requirement and the inability to smooth out management processes to dispatch operational personnel on time who could make required changes. The next major innovation in the metro optical space was hence the reconfigurable optical add drop multiplexer (ROADM).

The ROADM could add or drop wavelengths (channels) from a fiber in different configurations. A subsystem of merit in the ROADM system was the transponder that could take in a client signal and convert it to a DWDM compliant wavelength (in the ITU band). Transponders began as simple translators of data from a client wavelength to an ITU-grid enabled wavelength, but today do much more. With the advent of coherent optics, transponders multiplex multiple smaller rate channels into a coherently modulated signal that is further sent into the fiber. Transponders are the starting and stopping point of every signal in a metro optical network. In conjunction with a ROADM, the transponders make up the full optical transmission system that is necessary for the transport of high-speed data-channels.

ROADMs are classified based on the degree of their flexibility. The first generation of ROADMs could add and drop channels at specific ports and only a few channels could be added or dropped at a node. This system meant that the overall inventory of wavelength specific transponders had to be carefully planned and led to a rigid environment. In the second generation, the ROADM was built using an arrayed waveguide, as shown in Fig. 18.5. In this embodiment, a composite WDM signal enters a node that has an arrayed waveguide (AWG) pair arranged in a mirror-conjugate model—the first AWG demultiplexes the signal into constituent wavelengths, while the second AWG multi-

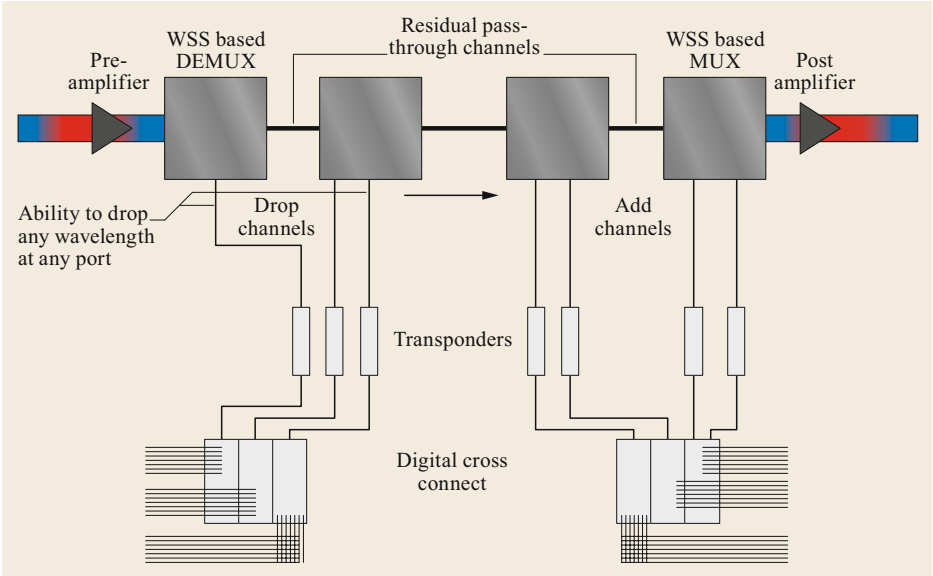


Fig. 18.6 Modern ROADMs with WSS technology

plexes individual signals to re-create a composite WDM signal. In between the two AWGs are a series of 2×2 optical switches, one each for each wavelength. If the switch is in the bar state then the wavelength passing through the switch is a pass-through for that node, while if the switch is in the cross state then the wavelength can be dropped at the node and the same wavelength (with locally injected data) can be sent into the network (with that node acting as an ingress node). This kind of a ROADM system, in which all the channels in the WDM band can be added or dropped at any node, but with the caveat that there is a strict port-to-wavelength mapping, was popular until the advent of the wavelength selective switch (WSS) technology.

The WSS technology revolutionized ROADMs and make their appearance completely dynamic to traffic needs and facilitated mesh networking, which was difficult in the earlier generation of ROADMs.

With the advent of WSS technology, the ROADM was made more agile to meet dynamic traffic needs. A WSS device of $1\times N$ port count can route any combination of wavelengths between the input and any of the output ports. In another embodiment, a $M\times N$ WSS can route wavelengths from any of the M input ports to any of the N output ports to achieve the behavior of an optical circuit router. By making use of multiplex WSS one can create a mesh networking element. Figure 18.6 is an illustration of a ROADM with WSS.

A point of merit of a ROADM developed from WSS is the ability to support a non-blocking wavelength routing fabric. In this regard multiple ROADM implementations are proposed based on three parameters: colorless, directionless and contentionless designs.

Colorless designs allow any wavelength to be dropped or added from any port. Directionless designs allow that any wavelength be sent to any fiber (especially valid in an $N\times N$ fiber system at a node). In a directionless design any wavelength can also be dropped from any fiber at the node. Finally in a contentionless design we can drop multiple identical wavelengths (from multiple fibers) at any port at a node. This means that if we have a 4×4 ROADM supporting a 4-degree mesh node, then we can drop λ 1 from fiber 1, λ 1 from fiber 2, λ 1 from fiber 3, and λ 1 from fiber 4 all at the local node across adjacent or any add/drop ports that we desire.

Achieving colorless, directionless and contentionless (CDC) in ROADMs is expensive and many architectures have been proposed to this end. As a compromise a full CD (colorless/directionless) ROADM is easy to build with some partial degree of contention. The reason why full contention is difficult is because multiple similar wavelengths from multiple fibers when they traverse to the same WSS, get lost as noise—the WSS cannot differentiate between two similar wavelengths—it can only differentiate across wavelengths. So other node elements such as port power splitters, blockers, power switches, multicast switches (MCS) are used to build a ROADM—which all lead to cost increase and power penalties. It is safe to say that the ultimate CDC ROADM that can work at low-prices is yet to be built and perhaps requires some out-of-the-box approach. As a compromise, partial CDC ROADMs (see Fig. 18.7 for an architectural illustration of a partial CDC ROADM) that breaks add drop ports at a node into banks are developed,

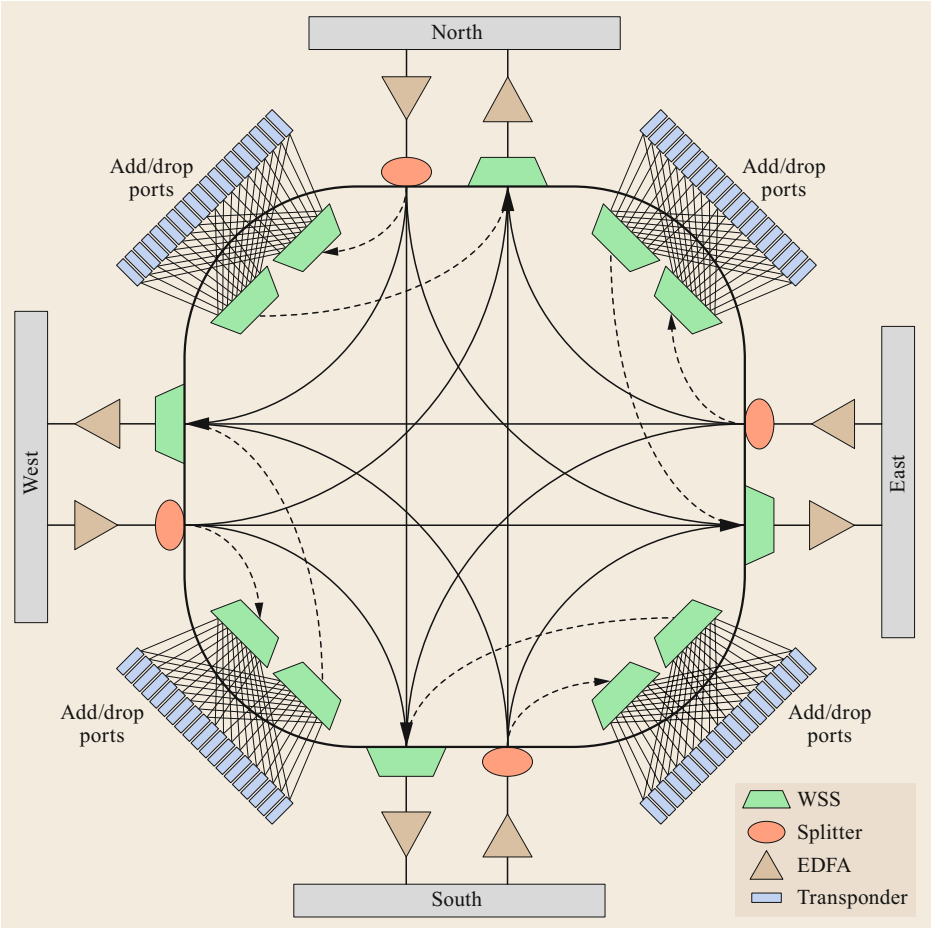


Fig. 18.7 Typical partial CDC ROADM

whereby there is no contention across banks, but there is contention within a bank—i.e., only one instance of a particular wavelength can be added/dropped at a bank.

A final feature of the ROADM technology is the gridless feature, whereby the traditional ITU-mandated spacing between channels is relaxed [18.17]. In the gridless design, channels of variable spectral width can be multiplexed together in the frequency domain thus facilitating better reach, multiplexing advantages, and circumvention to impairments. Gridless designs also facilitate coherent communication with the advent of the superchannel concept that facilitates upto a Tbps of data on a carrier-group system.

18.6 Metro Ethernet

Ethernet is the most successful and dominant LAN technology world-wide for over three decades with close to half a trillion Ethernet interfaces across the globe and counting. Ethernet interfaces are almost always available across all sorts of devices irrespective of the medium. Ethernet interfaces span across fiber, copper and wireless domains implying absolute ubiquity. One of the reasons why Ethernet has been so successful is due to its simplicity, cost-efficiency, and ease of deployment. Ethernet was initially designed for data communication across machines over a campus LAN environment. Ethernet began as 1 Mbps Ethernet in the Xerox corporation. Since then, Ethernet has come a long way—to 100 Gbps and beyond. Ethernet standards at 200 and 400 Gbps are ready, and it is just a question of time before terabit Ethernet arrives. Figure 18.8 shows the basic transition of the protocol stack with the advent of carrier Ethernet.

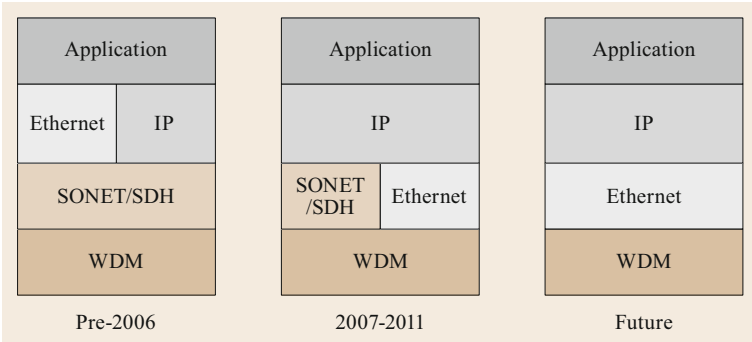


Fig. 18.8 Ethernet transition

Ethernet began as a LAN technology in the late 1970s, and was standardized in February 1980 (hence 802) under the premise of the IEEE as a LAN technology. The earliest standardized avatar of Ethernet was 1 and 10 Mbps Ethernet and could potentially connect hosts (computer systems). The technology deployed was using carrier-sensing. Hosts would sense if there was transmission on a carrier and if that was the case would not transmit till the existing transmission had subsided. This led to issues such as fairness, inordinate delays and limited efficiency (two hosts who transmitted simultaneously or in overlapping intervals due to distance separating them would result in loss of data for both hosts). This carrier sense multiple access (CSMA) standard was reasonably acceptable for small sized LAN environments with occasional data spurts. An improvement over this standard was the collision detection (CD) mechanism where by a CSMA-CD system would continue to listen to a communication channel and upon detecting a collision, would stop sending data immediately as well as sending a jamming signal, thereby resulting in saving precious time that would otherwise have been lost in transmission of collided frames by the two or more hosts involved (in the collision). CSMA-CD technology improves the Ethernet transmission system. It however, continues to be probabilistic in nature (probability of a collision is the probability of another station transmitting during the ranging time of a frame already being transmitted). To improve CSMA-CD, switched Ethernet was proposed. One of the premises of switched Ethernet was the use of bridging technology. Ethernet bridges, standardized under 802.1D was the first approach towards making Ethernet venture across LAN segments covering entire enterprises. An Ethernet bridge, at an abstract level was analogous to a 1×1 switch connecting two LAN segments on either side of the bridge. Each LAN segment is characterized by a spanning tree. Spanning tree protocol (STP) runs in a LAN segment, such that bridge protocol data units (BPDUs) are exchanged between hosts, which over a period of time results in ports being declared as forwarding, filtering, or blocking, such that every host now has a reasonably short path to the bridge (here called the root bridge). The prime goal of STP is to create a loop-free topology such that hosts can facilitate communication with each other using MAC-bridges. While STP is a reasonably good way of achieving loop-free communication, the time required for STP convergence post a failure or change of state of hosts (host dropping, adding, or new ports being made active) is probabilistic in nature leading to significantly non-deterministic provisioning times. Learning of MAC addresses and bridge port IDs and then pruning the spanning tree got worse as the number of prospective nodes in the tree increased.

An improvement over MAC bridges was the introduction of the 802.1Q standard also called the virtual LAN (VLAN) standard [18.18]. The idea was simple—insert a four byte tag called a VLAN tag, between the MAC address and the IP address in an Ethernet frame such that the inserted tag would contain VLAN ID and priority information. VLANs allowed partitioning of a LAN environment into several VLANs that could each have their own broadcast domain (since spanning tree works on the broadcast principle). The advantage of such an approach is to limit the size of the broadcast domain to just a VLAN (as opposed to the entire LAN). A host can be part of multiple VLANs and that way can still possibly communicate with any node in the network. To begin with, the 802.1Q protocol classified three types of VLAN frames—a tagged frame, a port based VLAN (or port VLAN) and a QoS VLAN. VLANs had two bytes reserved for definition purposes—to denote that the frame conforms to 802.1Q. Other VLANs have also been defined such as a customer tag (CTAG) and a service tag (STAG), each of which has its unique 2-byte Ethertype. In addition to the Ethertype, we have 12 bit VLAN ID, called VID that leads to 4096 instances of VLANs that can be supported. The first (all zeros) and the last (all Fs in hexadecimal or ones in binary), are reserved, leaving 4094 instances. For purposes of QoS, 3 bits are used leading

to eight QoS levels. The last remaining bit is reserved as a drop eligibility indicator (DEI) bit for traffic shaping or as a canonical format indicator (CFI) bit.

With VLAN technology, spanning tree domains were restricted in size and new spanning tree protocols such as rapid spanning tree protocol (RSTP) and multiple spanning tree protocol (MSTP) [18.19] were proposed. These helped consolidate VLAN based systems within enterprises. One could now differentiate frames by their VIDs or by their QoS values (a feature that traditional Ethernet has lacked). A frame with a VLAN tag is called a tagged VLAN frame, and a port that supports VLANs is called port-VLAN. While this system has worked well, it is still dependent on spanning tree protocol, implying a probabilistic nature to forwarding that has resulted in non-deterministic latency, and lack of service-class behavior especially for sensitive services like voice, and video.

Despite the non-carrier class nature of 802.1Q, the advantages of VLANs had propelled providers to use bridging and VLANs. The problem however, was that the number of VLANs supported was then restricted to 4094. This led to three issues:

1. The provider could have at the most 4094 unique 802.1Q customers—a bad way to scale business.
2. The provider now had to run its own STP and learn all the MAC addresses of all its customers—which meant that it had to deal with issues of mobility, capacity and have a robust algorithm for learning.
3. Finally learning MAC addresses of customers could potentially compromise the security of providers (due to MAC spoofing and similar such approaches).

As a way out of this situation, the first approach proposed was to 802.1ad or provider bridges. Provider bridges facilitates the separation of provider and customer equipment (and hence MAC addresses) into two domains. To do so, a new tag with a unique ethertype was proposed—called the STAG. Each customer now could run its own spanning tree and support 4094 VLAN instances within its (customers') domain, while the provider could now have its own STP and VLAN. Though this did not solve the scalability issues, the approach of two tags (also called Q-in-Q) [18.20] did solve an important approach of demarcating the customer domain from the provider domain. In cases of 802.1ad supporting frames, there are up to two tags presents—an outer provider tag called the STAG and an inner customer induced tag called the CTAG. Once a frame enters the provider domain it is handled exclusively using the STAG in conjunction with the destination MAC address. The advantage of the Q-in-Q approach is that it gives a semblance of an Ethernet service outside of the LAN environment, i.e., in the provider MAN or WAN domain. The type of service could be point-to-point or multipoint-to-multipoint service.

The problem with the Q-in-Q approach was the persistent fact that the provider had to learn customer MAC addresses—which led to compromises in security and the hassle of learning customer MAC addresses. Provider backbone bridges, under the IEEE premise of 802.1ah were hence proposed as a method to avoid learning customer MAC addresses and provide a complete demarcation between the operator and all its customers. This notion of provider backbone bridges includes the famous MAC-in-MAC approach, whereby a customer frame was completely engulfed in a provider frame at the edge of a metro Ethernet network. Providers were no longer required to run a spanning tree implementation all the way into a customer network which meant that provider security was not compromised. Provider backbone bridges (PBB) is a service-oriented technology, though not completely carrier-class (as it does not support deterministic latency all the time). Beyond PBB, the Ethernet standards define three types of Ethernet services [18.21]:

1. ELINE: A point-to-point bidirectional Ethernet virtual circuit (EVC) that facilitates packet flow between two end-points in a network.
2. ELAN: A multi-point to multi-point (MP2MP) connection between multiple end-points.
3. ETREE: A one-to-many (root to leaves) simplex connection, and a leaf to root simplex connection in the reverse direction.

These three service types are defined by the metro Ethernet forum. For each service type it is mandated that some sort of operations and maintenance standard be used, and by far the leader of the OAM pack is the IEEE802.1ag connectivity fault management (CFM) standard. In the 802.1ag standard, we set up maintenance domains which are zones of activity of traffic such that connections set up in a domain are the responsibility of the domain peers for maintenance.

CFM works as follows: for each ELINE, ELAN or ETREE, we set up maintenance end points (MEPs) at each end node of a service. The MEPs exchange heartbeat messages called connectivity check messages or CCMs. Loss of three CCM messages indicates a fault and switch over from the work path to a protection path. The idea is to provision the work and protection paths such that the two paths are node-and-edge disjoint—a non-trivial task for multicast ELANs. In addition to CCMs, there are loopback messages (LBMs)

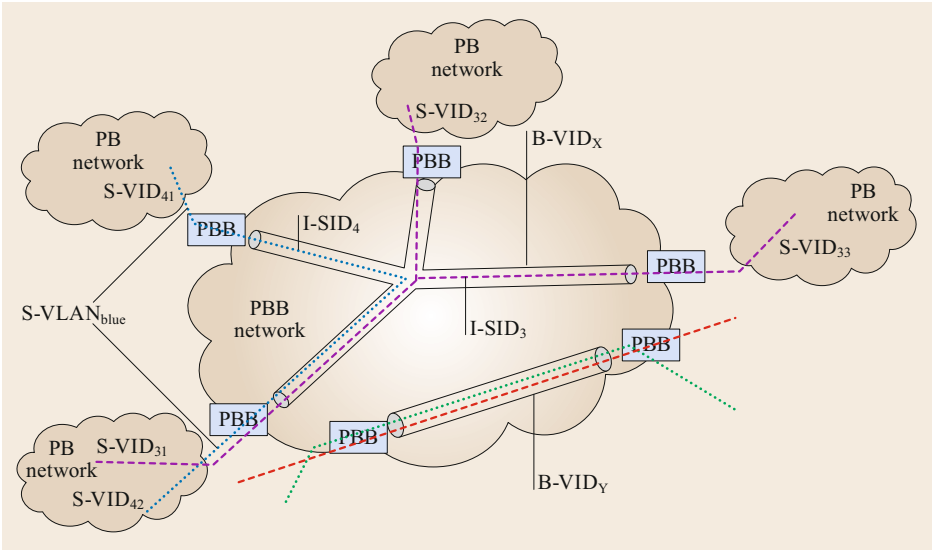


Fig. 18.9 Provider backbone bridging network (PBBN) according to the IEEE 802.1ah standard

and linktrace messages (LTMs) used for service housekeeping.

The premise of PBB [18.22] works as follows: nodes are earmarked as edge nodes and core nodes based on service end points or intermediate points. A node is an edge node for a service starting or terminating at itself, while it is a core node for another service that goes through the node. In some literature, edge and core nodes are also called edge and core bridges to continue to the 802 legacy of bridge based forwarding. In the 802.1ah premise of PBB, a service that enters a PBB domain (also called a provider backbone bridged network (PBBN)), is first mapped on to an intermediate service tag called the ITAG. The ITAG is special—it is unlike other VLAN tags which are four bytes long—the ITAG is in fact six bytes long. The extra length of the ITAG are used to support an ISID of twice the length of a normal VID—to 24 bits. An incoming service into the PBBN, could be a port-based service, a native Ethernet service, a CTAG implementation, a STAG implementation, or a service with CTAG and STAG. The service is first mapped on to an ISID that is unique across the network. The presence of a 24-bit ISID allows for up to 16,777,216—about 16 million services that facilitate significant scalability in terms of services for a provider network. The ISID mapping is an intermediate mapping, whereby the incoming packet flow is converted into a service-specific instance due to the mapping. On one side the ISID is mapped to the incoming service, while on the other side (i.e., into the PBBN), the ISID is mapped on to a backbone tag (called BVID), and a backbone MAC address. This means that the incoming Ethernet frame from the client into the PBBN, is now mapped in its entirety on to another Ethernet frame. This kind of complete encapsulation of a client frame on to a provider frame is also called MAC-in-MAC, as there are two MAC addresses in each frame (the client one and the provider one). The provider now forwards based on a 60-bit unique combination of BVID and BDA (backbone destination address—the providers' MAC). For this forwarding, STP is used for pruning the PBB network and creating a spanning tree. A PBBN implementation is as shown in Fig. 18.9.

Despite all its facilities and nuances, 802.1ah is not carrier class—the latency is dependent on STP performance, which as we know is probabilistic and not deterministic. An improvement over PBB 802.1ah is the first carrier-Ethernet (CE) standard called PBB-TE (provider backbone bridging-traffic engineering). PBB-TE is backward compatible with all the other .1Q standards. The premise of PBB-TE is as follows:

1. Spanning tree protocol is no longer used in the provider domain.
2. MAC learning to create paths in the provider domain is switched OFF.
3. A network management system (NMS) is used as a centralized entity to set up paths (allocate ISIDs, BVIDs and compute work and protection paths).
4. The NMS also supports IEEE802.1ag CFM standard for operation and administration.

In PBB-TE [18.24], EVCs are created by the NMS. Traffic flows only after the NMS has provisioned the work and protection paths by setting the BVIDs and ISIDs. For each service the NMS also sets up the MEPs.

The PBB-TE network is characterized by deterministic delay due to the reliable forwarding nature of the layer-2 forwarding plane. The absence of MAC learning

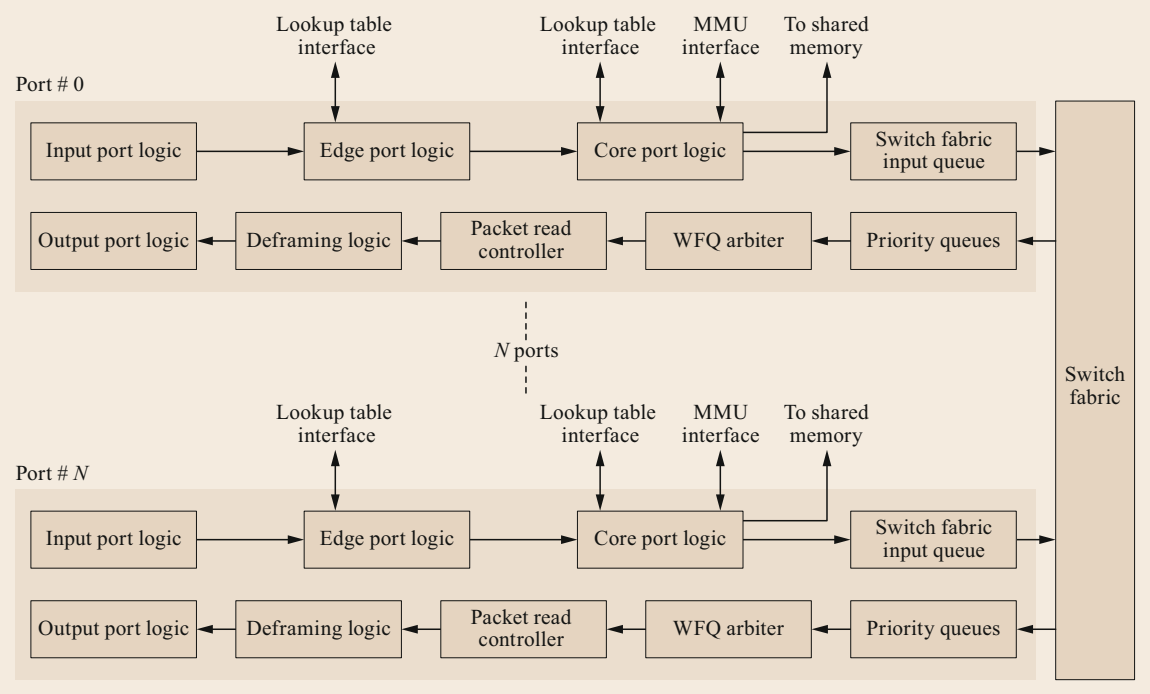


Fig. 18.10 A block diagram of a shared-memory PBB-TE switch (after [18.23])

and STP means that the probabilistic nature of forwarding is now replaced by just the L2 forwarding plane, which leads to bounded delay. The NMS-controlled domain implies that services enter the domain only through the providers' NMS, resulting in a traffic engineered backbone. This facilitates guaranteed service throughputs by advance planning and traffic engineering. Some of the engineering aspects of PBB-TE are vendor specific and hence performance can change from one vendor to another. Shown below is a PBB-TE implementation in Fig. 18.10.

While the Ethernet bodies in the IEEE were standardizing carrier Ethernet, another development was happening in the ITU and IETF bodies. This was the creation of MPLS-TP (transport profile) [18.25] as a carrier-grade Ethernet technology. While one can say that PBB-TE is an enrichment process of native Ethernet, MPLS-TP is a watered down process of MPLS. MPLS as we have discussed before is an improvement over IP forwarding through the use of labels that act as force multipliers on a forwarding equivalence class. MPLS is suited for large networks and heavy traffic granularity. MPLS however, is not fully carrier-grade. The IP-based control plane combined with best-effort behavior of the forwarding plane does not accord the degree of determinism that one would expect in a carrier-grade technology. In some ways it can be said that MPLS due to its feature-rich base needs to be tailored down to become transport-oriented. An effort which initially began in the ITU and soon spread to the IETF, was the MPLS-TP study group (15 in the ITU). The study group defined MPLS-TP as a transport technology that used much of MPLS's forwarding plane, but defined a new control plane. MPLS best effort features such as label merging were abandoned. Also abandoned was penultimate hop forwarding, which implied that a label switched path could be identified by labels from ingress to egress and was supported strictly by the MPLS-TP forwarding plane. The rest of the forwarding plane remained the same as MPLS. As far as the control plane was concerned, the study group recommended the use of a generic associated channel (G-ACh), as defined in RFC (Request for Comments) 5586 (and further updated in RFCs 7214 and 7274). The G-ACh allowed for MPLS-TP OAM techniques to be implemented, similar to the IEEE802.1ag CFM or the ITU equivalent defined in Y.1731.

MPLS-TP defines bidirectional LSPs as the norm—this allows for service oriented communication with a bidirectional control plane that is necessary to provide for necessary carrier-grade performance and OAM. By removing the support for label merging, the forwarding plane is made more deterministic. Finally, MPLS-TP uses the large installed base of MPLS leading to better acceptability in a carrier network. Though MPLS-TP is a certain betterment over MPLS from a carrier-class

performance perspective, the deployment is not clearly in its favor. This is primarily because of the perceived ease of use of MPLS in the operator world. MPLS-TP along with OTN however, continue to be popular in greenfield settings where operators desire a lower-cost and carrier-grade-from-the-beginning option. There is also significant focus on the use of MPLS-TP in backhaul for mobile networks, especially for connectivity from the base station to the switching center. There are however, vendors who are today offering IP/MPLS solutions with the MPLS-TP defined OAM technique. For well planned, traffic engineered networks, this is a good offering resulting in bounded latency. The traffic-engineering aspect takes away to a significant degree the best-effort induced probabilistic latency that is otherwise seen in IP networks.

18.7 SDN and Metro Networks

Software defined networks have been heralded as a game-changer for next generation networks, and SDNs are typically relevant in the metro environment. The concept of SDN engages the use of bringing programmability to networks. The central premise of SDN is the bifurcation of the control and data plane, so that the control plane can be made programmable and the data plane can support a host of protocols. To achieve this objective, the control plane is manifested by a controller. A controller is characterized by its four interfaces: a north bound interface that provides access to APIs, by using this a user can define a service manifestation; a south bound interface that connects the controller to the data plane; east-and-west bound interfaces for hierarchical connection of multiple controllers to scale a network. The data plane is further embodied by nodes that are pure forwarding entities and an implementation of a node is called a whitebox on account of the wide-range of protocols that a node supports. The whitebox architecture is currently getting significant attention. An SDN forwarding plane or whitebox has three key functions: forwarding, modification and dropping packets. The principle of vanilla SDN is to send packets that arrive for the first time to a controller, which then populates forwarding tables in the SDN whitebox [18.26]. The packets thereafter are forwarded based on the controllers' table recommendation. The whitebox typically implements a match-action sequence where by packets are parsed by a parser, which matches protocol fields to actions in a table. These actions could be forwarding to a port, modification of the packet header (like TTL decrement), or dropping a packet altogether.

A key aspect of SDNs is the protocol between the controller and the data plane (or whiteboxes). This protocol, also called south bound protocol is crucial in terms of the support that it offers to the data plane as well as to the north-bound programmable interface. The SDN protocol defines fields that the data plane reads and acts upon in order to forward data. All the processing of protocols happens in the control plane that resides in the centralized entity—the controller. Hence, the shortest-paths are computed by the controller and only the resulting metadata action entity is fed to the data plane such as populating the forwarding table with IP-prefix-output-port matching entries. Similarly, the controller today supports about 40-odd protocols that the data plane must further act upon to be SDN compliant. A leading and popular controller is the open-day-light controller [18.27] that supports one of the leading protocols called open-flow. Currently open-flow supports about 40-odd protocols as shown in the table below. One of the drawbacks of open-flow is the perceived time required for the northbound to add a new protocol, which turns out to be a couple of years of standardization exercises. Another competing protocol is the protocol oblivious forwarding (POF) approach. POF works upon a smaller set of fields and commands and attempts to do away with the rigidity of the open-flow southbound.

An SDN whitebox is characterized by multiple match-table (MT) approaches, such as single match table or SMT, multiple match table (MMT) and reconfigurable match table (RMT) [18.28]. In SMT a single match table is used for deriving a relation between a protocol header field and the action to be done on the packet. In MMT, multiple tables are cascaded together in proprietary formats such that the cascade could be based on round-robin schemes or based on distributed hash table techniques. In the case of RMT, there are multiple tables such that table width and depth are reconfigurable entities. The advantage of the RMT approach is that the data plane in terms of forwarding or modification functionality can be made as programmable as required. To make the data plane further programmable and service oriented, a parsing language has been recently proposed called P4 [18.29]. P4 allows for service definitions, templates and parsing instructions to be passed on to a programmable data plane.

Several operators are going the SDN way. One of the key approaches to SDN is AT&T's ONAP approach. ONAP (open network automation platform) allows creation of services and incorporation of SDN concepts. Several vendors are SDNizing their product portfolio by building controllers and opening APIs on which providers can create new services. The yet another network generator (YANG) language is being considered for service definitions and products. SDN along with network function virtualization can be significant cost-savers and service enhancers in the future network.

18.8 Network Function Virtualization (NFV)

NFV is touted as the next big thing in networks on account of the possibilities in terms of services and cost-savings that are presumed to be associated with NFV. Replacing physical network functions (PNFs) with virtual network functions (VNFs) is the key premise in NFV. By replacing PNFs with VNFs, one can commoditize networks significantly by installing VNFs as software entities on commercial off-the-shelf (COTS) servers. This reduces the need for expensive application-specific hardware that is difficult to maintain, has a vendor lock-in and is quite centralized. In contrast, VNFs can be installed just about anywhere in the network and a provider can negotiate a good pricing policy with the VNF vendor implying the ability to create VNF instances just about anywhere. VNFs can be created in data-centers, at the edge of a network (in central offices—an example of which is the CORD project) [18.30], and in the network core as middle boxes. NFV has been undergoing several discussions in the study groups, and one standardization body in particular—the ETSI has come up with a series of recommendations. An NFV implementation often is said to involve three entities: an NFV infrastructure (NFVI) which provides for the servers, the network fabric etc., the management and network orchestration (MANO) framework [18.31] that facilitates management, lifecycle support and growth of VNFs as well as VNF chains, and finally the VNFs themselves. MANO provides for a detailed mechanism whereby a provider can track, provision, operate and administer the NFV domain and create services, manage customers, and seamlessly operate the network. By connecting multiple VNFs to each other, one can create a service function chain (SFC). Such a service chain can lead to multiple new services as well as additional revenue. Currently it is however, not fully clear as to how much of NFV would be adopted by a service provider [18.32], though it is clear that initial implementations of NFV would all be customer facing. Provisioning VNFs in network cores for core networking functions such as border routers and gateways, still seems far-fetched, but with recent advances in making the x86 processor more amenable to a data plane it is perhaps only a question of time when an entire provider's network could be virtualized. Since most initial NFV deployments are likely to be in the data-center, there is a strong relation between NFV and SDN. While NFV virtualizes the data plane in terms of functionality, SDN facilitates the growth of the NFV virtual topology by routing flows, and managing entities required for NFV especially in a data-center domain. One could envisage VNFs sitting in COTS hardware in a data-center, such that the COTS servers are all connected to each other via an SDN switching system. In such a situation, the SDN controller is made aware of the VNFs and the flow requirements of VNFs. The controller in this case is also assumed to be part of the MANO framework, with distinct communication between the controller and VNF application (run on the NB of the controller).

VXLAN and NVGRE: Two new RFCs in the IETF domain for data-center communication are VXLAN [18.33] and NVGRE [18.34]. Both of these are designed specifically for intra-data-center communication in a virtualized infrastructure. At the crux of these two RFCs is a mechanism that allows a tunnel to set identifiers that are connected to a virtual machine (VM), such that movement of the VM within or even across data-centers allows an associative relationship between the network protocol identifiers (say IP address) to be maintained. Hence, aspects such as load-balancing can easily be carried about whereby moving VMs across the data-center has no bearing on the routing strategy. Though procedurally different, VXLAN and NVGRE aim to achieve a similar set of functionalities.

The future of the metro network involves significant softwarization from the SDN and NFV perspectives. One of the goals is to induce virtual network functions (VNFs) in networks that can be spun up on demand in any corner of the network, and create service chains that can offer new possibilities in terms of the product and service mix. To this end, SDN and NFV has a strong role to play in the future metro network. The issue however, is that of timing—when, where, and how do we induce SDN and NFV in a brown-field metro network. There is obviously no clear answer to this question.

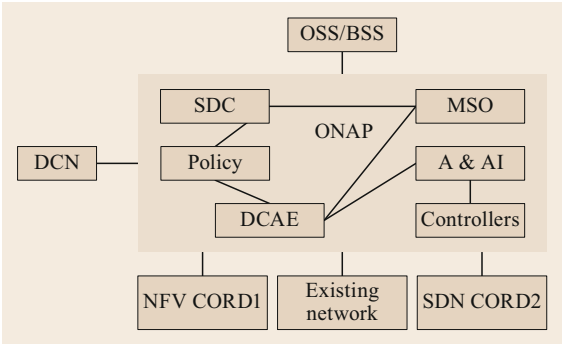


Fig. 18.11 ONAP and its components

One of the leading approaches towards network softwarization is the AT&T induced approach called open network automation platform (ONAP) [18.35]. ONAP is a platform that allows multiple network-related parties, such as developers, application writers, the service provider, VNF vendors, existing PNF vendors and customers (enterprises) to interact. Broadly speaking ONAP consists of the sub-systems shown in Fig. 18.11. The first subsystem of interest is the service design creation (SDC) module. The SDC module facilitates the operator or enterprise customer to create services. SDC creates metadata that is used by other components of ONAP in service definitions. SDC manages different types of assets such as resources and services. Resources include infrastructure, network and applications (VMs), whereas a service is a well-formed object of one or more resources. The SDC contains a catalog that is a repository for assets used in ONAP. Part of the SDC is a design studio that is used to create, modify and add resources and services. Once services are created by the SDC, it certifies them post testing and automation. A distribution studio is used to disseminate certified assets and resources. ONAP uses a policy framework that facilitates the existence of different user driven policies. A second component of ONAP is the MSO (master services orchestrator). The primary function of the MSO is automation of end-to-end service instances. The MSO routinely interrogates the active and available inventory module to find what assets are available in the network. ONAP assumes SDN controllers and other controllers to also be integrated as part of the framework.

There are three types of controllers assumed: infrastructure controllers for creating VMs, provisioning servers etc., network controllers for connectivity and configuration; and application controllers that manage virtual application. To get feedback from the network, ONAP (Fig. 18.11) uses a module called data collection analytics and events (DCAE). This module uses passive and active network probes to monitor the network, collect network data, perform analytics on data and metadata and register events. DCAE is a complex module and is expected to use machine learning to perform network-wide diagnostics. Sub-components of the DCAE module include an analytics framework, a collection framework, a data-distribution bus, a storage medium and access to controllers as well as microservices. The ONAP system has the potential to revolutionize the telecom landscape by inculcating SDN and NFV concepts and allowing small startups to create VNFs as part of a new telecom ecosystem. One key challenge to such as design is to enable interoperability across vendors. Another design challenge is to come up with a high-availability number that would facilitate a carrier-grade network.

18.9 Best Practices for Metro Network Design

We now consider the technologies that we have discussed for best practices in metro network design. The following design considerations are necessary considerations for an efficient metro network design:

1. Cost, vulnerability and energy efficiency are the lowest in lower layers of the data spectrum (Fig. 18.12). It makes sense to keep the data in lower layers—optical if possible, or else data layer, or else SDN etc. in that order.
2. As far as multi layer optimization is concerned, aggregation at the edge is always preferred as compared to aggregation in the core. In the core of the network, it makes sense to route optical circuits of higher granularity than to route LSPs or IP packets.
3. The finer the switching granularity in the core, the more expensive the core. SDN goes a long way in alleviating some of the pain-points in the core of the network.
4. From a telecom design perspective, it makes more sense to have a multihoming architecture. For example if there is a ROADM based metro ring, and an IP overlay, then the corresponding IP router should be multihomed to multiple ROADMs.
5. The same principle of multihoming applies to carrier Ethernet equipment and ROADMs as well.

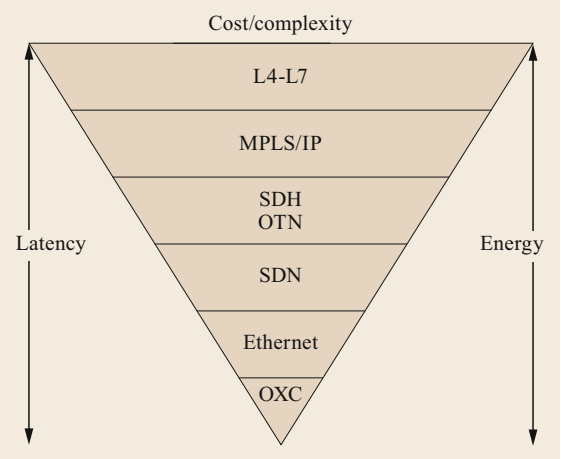


Fig. 18.12 Cost-complexity across the metro layers

6. While designing a metro network, it can safely be assumed that traffic is usually step-wise transient—there is an increase and then a flattening out effect. In such a scenario, at each step it can be assumed that the traffic follows primarily a hub-and-spoke topology. There may be multiple hubs (usually two hubs), and multiple spokes, and this hub-and-spoke topology may be overlaid on a ring or mesh network. The hub-nodes are usually at the core network interconnection points.
7. SDN has the potential to do away with a significant amount of redundant network equipment. The challenge with an SDN based network is the migration plan. Even in the absence of killer-SDN services, an SDN makes more sense than legacy technologies due to the service potential.
8. Providers that implement NFV have the potential to do away with middleboxes and replace these by VNFs. Such a scheme can potentially bring down costs. In such a deployment, the VNF performance parameters need to be benchmarked.
9. MANO/ONAP or other such forays into automation of the network are critical for the future of metro networks to offer a revenue-centric products portfolio.
10. The data-cloud (created due to the network), the application cloud (due to OTT players) and the management cloud are all inter-twined. In such a scenario a unified management system is crucial for end-to-end seamless provisioning as well as effective monetization of the network towards new services.

18.10 Summary

The metro network is perhaps the most important aspect of the network especially when it comes to enterprise and business connectivity. We have described the architecture of contemporary metropolitan networks as well as the different technologies that make up the metro network. To this end, we have considered circuit and packet technologies such as WDM, SONET/SDH, OTN, metro Ethernet, and IP. We have explained the technological processes involved and the required equipment. We have also given an extensive overview of the new metro technology—SDN and highlighted its dominant open-flow protocol. It is expected that the metro network will only grow [18.36] in the years to come and will become a mainstay of enterprises and business and will be used to transport city-wide data.

References

18.1 A. Gumaste, S. Akhtar: Evolution of packet-optical integration in backbone and metropolitan high-speed networks: a standards perspective, IEEE Commun. Mag. **51**, 105–111 (2011)
18.2 J. Donovan, K. Prabhu: *Building the Network of the Future. Getting Smarter, Faster and more Flexible with a Software Centric Approach* (CRC, Boca Raton 2017)
18.3 A. Gumaste, T. Antony: *First Mile Access Networks and Enabling Technologies* (Cisco Pearson, New York)
18.4 Google Cloud: https://cloud.google.com
18.5 D. Kreutz, F.M.V. Ramos, P. Verissimo, C.E. Rothenberg, S. Azodolmolky, S. Uhlig: Software-defined networking: a comprehensive survey, Proceedings IEEE **103**(1), 14–76 (2015)
18.6 Wikipedia: Synchronous optical networking, https://en.wikipedia.org/wiki/Synchronous_optical_networking (2019)
18.7 A. Gupta, R. Jha: A survey of 5G network: architecture and emerging technologies, IEEE Access (2015), https://doi.org/10.1109/access.2015.2461602
18.8 A. Gumaste, N. Krishnaswamy: Proliferation of the optical transport network: a use case based study, IEEE Commun. Mag. **48**(9), 54–61 (2010)

18.9 J. Moy: OSPF Version 2, RFC 2328 (1998)
18.10 Broadcom: Broadcom ships Jericho2, industry's highest bandwidth ethernet switch-router at 10 terabits per second, https://www.broadcom.com/news/product-releases/broadcom-ships-jericho2 (2018)
18.11 K. Pagiamtzis, A. Sheikholeslami: Content addressable memory (CAM) circuits and architectures: a tutorial and survey, IEEE J. Solid-State Circuits **41**(3), 712–727 (2006)
18.12 E. Rosen, A. Viswanathan, R. Callon: Multiprotocol Label Switching Architecture, RFC 3031 (2001)
18.13 P. Pan, G. Swallow, A. Atlas: Fast Reroute Extensions to RSVP-TE for LSP Tunnels, RFC 4090 (2005)
18.14 D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G.Swallow: RSVP-TE: Extensions to RSVP for LSP Tunnels, RFC 3209 (2001)
18.15 E. Rosen, D. Tappan, G. Fedorkow, Y. Rekhter, D. Farinacci, T. Li: MPLS Label Stack Encoding, RFC 3032 (2001)
18.16 M. Vigourex, D. Ward, M Betts: Requirements for OAM in MPLS Transport Networks, RFC 5860 (2010)
18.17 A. Gumaste, N. Ghani: Reach optimized architecture for multi-rate transport system (ROAMTS): One size does not fit all. In: *OSA Opt. Fiber Conf. (OFC)* (2009), Paper OMQ3
18.18 IEEE Standard document Virtual LAN 802.1Q
18.19 IEEE Standard Document MSTP 802.1s
18.20 IEEE: Standard Document 802.1ad—Provider Bridges, http://www.ieee802.org/1/pages/802.1ad.html (2006)
18.21 MEF: Ethernet services, http://www.mef.net/carrier-Ethernet-services/ce-2-0-overview (2019)
18.22 IEEE 802 standard document Provider Backbone Bridging PBB (2008)
18.23 S. Mehta, A. Upadhyaya, S. Bidkar, A. Gumaste: Design of a shared memory Carrier Ethernet switch compliant to provider backbone bridging-traffic engineering (IEEE802.1Qay). In: *2012 IEEE 13th Int. Conf. High Perform. Switch. Routing, Belgrade* (2012) pp. 29–35
18.24 IEEE 802 standard document Provider Backbone Bridging and Traffic Engineering (2009)
18.25 S. Bryant, L. Andersson: Joint Working Team (JWT) Report on MPLS Architectural Considerations for a Transport Profile, RFC 5317 (2009)
18.26 A. Kushwaha, S. Sharma, N. Bazard, T. Das, A. Gumaste: A 400Gbps carrier-class SDN white-box design and demonstration: the bitstream approach, J. Lightwave Technol. **36**(15), 3115–3130 (2018)
18.27 SDN Controller Open Daylight: http://www.opendaylight.org
18.28 P. Bosshart, G. Gibb, H.-S. Kim, G. Varghese, N. Mc Keown, M. Izzard, F. Mujica, M. Horowitz: Forwarding metamorphosis: Fast programmable match-action processing in hardware for SDN, ACM Sigcomm Comp. Commun. Rev. **44**(4), 99–110 (2013)
18.29 P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. Mc Keown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, G. Varghese, D. Walker: P4: Programming Protocol-Independent Packet Processors, ACM Sigcomm Comp. Commun. Rev. **44**(3), 87–95 (2014)
18.30 L. Peterson, A. Shabibi, T. Anshutz, S. Baker, A. Bavier, S. Das, J. Hart, G. Parulkar, W. Snow: Central office re-architected as a data center, IEEE Commun. Mag. **54**(10), 96–101 (2016)
18.31 ETSI: *Network Function Virtualization: Management and Orchestration, GS NFV MAN 001, Group Specification* (ETSI, Sophia Antipolis 2014)
18.32 A. Gumaste, S. Sharma, A. Kushwaha, T. Das: How much NFV should a service provider adopt?, J. Lightwave Technol. **35**(13), 2598–2611 (2017)
18.33 M. Mahalingam, D. Dutt, K. Duda, P. Agarwal, L. Kreeger, T. Sridhar, M. Bursell, C. Wright: Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks, RFC 7348 (2014)
18.34 P. Garg, Y. Wang, Network Virtualization using Generic Routing Encapsulation (NVGRE), RFC 7637 (2015)
18.35 Open Network Automation Platform: https://wiki.onap.org/
18.36 AT&T: Domain 2.0 vision white paper, https://www.att.com/Common/about_us/pdf/AT&T%20Domain%202.0%20Vision%20White%20Paper.pdf (2013)

Ashwin Gumaste

Computer Science and Engineering
Indian Institute of Technology, Bombay
Mumbai, India
ashwing@ieee.org

Ashwin Gumaste, Ph.D. 2003, FINAE, is an Institute Chair Professor at IIT Bombay. He has built products in the Carrier Ethernet and SDN space, transferred to industry. He has published 185 papers and 25 US patents granted. He was awarded Bhatnagar award 2018 and Swarnajayanti Fellowship 2013.

19. Energy Efficiency in Optical Networks

Daniel C. Kilper

Energy efficiency is important for optical networks in terms of scalability, low-cost operation, and sustainability. At the same time, optical networks play an important role in enabling energy efficiency for the Internet as a whole and for information and communication technologies more generally. Understanding energy in the context of optical networks begins with an understanding of the constituent components and equipment that make up an optical network and how their energy use is evolving over time. The network architecture, describing how these parts are put together and operated, will also have a considerable impact on the energy efficiency of optical networks. This impact can be modeled and understood in different ways. Perhaps the largest energy impact, however, comes from including optical networks in cross-layer design and survivability strategies. These aspects of energy-efficient optical network design are examined, along with issues related to mobile and optical network convergence, nonlinear optics and optical processing, and computer and optical network cross-optimization. An introduction to resources for recommendations and guidance from standards bodies and other organizations on the energy efficiency of optical networks is also provided.

19.1 Energy and Optical Networks

Communication and data networks often play a key role in strategies for improving energy efficiency overall across society and for addressing climate change. Energy efficiency improvements and smart use of energy account for one-third to one-half of the reductions in greenhouse gas (GHG) emissions in published strategies for controlling climate change [19.1]. They are essential for so-called *smart* technologies, which use networks to monitor and control resources and energy. Recent industry reports estimate that 12 Gt of CO_{2e} GHG will be saved by 2030 through the use of information and communication technologies (ICT), which is nearly 10 times the GHG that ICT will produce, either directly or indirectly [19.2]. Figure 19.1 shows the different smart technology areas in which ICT may have an impact and their relative potential for carbon footprint abatement.

Optical networks, as part of this smart infrastructure, typically provide the greatest efficiency on a per-bit basis, i.e., the energy per bit of optical communication is lower than in other forms of communication or data transport [19.4–7]. Therefore, greater use of optical systems can enable lower-energy and more scalable smart infrastructure. Note, however, that in one study,

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_19

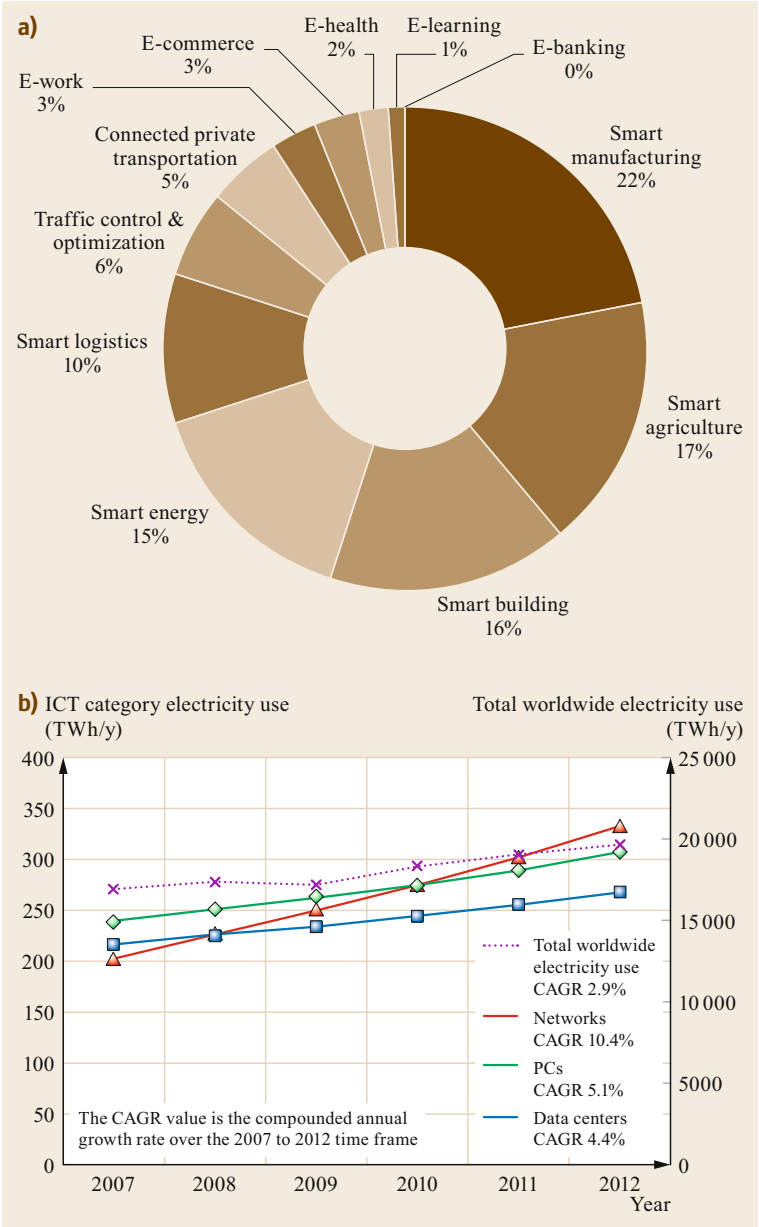


Fig. 19.1 **(a)** CO_{2e} abatement potential by use case [19.2] (Source: *SMARTer2030 report* © GESI 2015) and **(b)** electricity use of ICT over time compared with worldwide growth in electricity consumption [19.3] (© Elsevier 2013)

the embodied energy of the Internet was estimated, and fiber optic and copper cabling had the largest contributions [19.8]. Optical systems were not included in the study, but telecom switches were included, which are typically dominated by legacy hardware such as electronic circuit switches [19.9]. In fact, as network operators retire this legacy telephone equipment and replace it with IP and optical hardware, they have the potential for a short-term reduction in their energy use. Continued traffic growth and expansion of network hardware will result in long-term growth in network energy usage.

Research and development in the area of optical networks has historically been driven by efforts to increase performance and capacity, with little consideration for energy efficiency. Starting in roughly 2007, however, the energy efficiency of optical systems, along with other ICT technologies, began receiving increased attention [19.10]. Since that time there have been a number of review papers and tutorials that address the energy efficiency of optical networks [19.10–15], including a Best Readings list (http://www.comsoc.org/best-readings/topics/green-communications).

19.1.1 Global Impact and Scalability

Energy efficiency is the aspect of sustainability in which communication networks—and research and development of communication networks in particular—will typically play a role. Other aspects of sustainability in which communication networks have less impact include public policy, reduced use of fossil fuels, or direct efficiency improvements [19.1, 16]. Several studies, however, have pointed out that the use of Internet and data technologies can lead to waste and greater consumption [19.17]. In general, increased economic activity has shown a strong historical correlation with energy use, and the use of ICT correlates with increased economic activity [19.18]. This has drawn concern that greater use of ICT will lead to increased energy use. Furthermore, there are well-known economic models whereby efficiency improvements or efficacy measures reduce pricing or increase demand, which can drive up energy use or resource depletion as consumers use more. These models are often referred to as rebound effects. An early example of a rebound effect is the Jevons paradox, which comes from the observation that efficiency improvements in the use of coal in nineteenth-century England led to low costs and increased use as many households and factories were able to take advantage of coal-based heating and machinery [19.19]. A more recent generalization of the Jevons paradox is the Khazzoom–Brookes postulate, which attributes macroscale increases in energy use to microscale efficiency improvements [19.18].

Optical networks typically make ICT networks more efficient and therefore enable greater use of ICT. In fact, the Internet itself is made possible by the availability of high-capacity long-distance transmission in optical networks. Thus far, however, ICT networks, including optical networks, have not been associated with a rebound effect. Recent evidence on a global scale shows no clear impact on energy use or GHG emissions since the introduction of the Internet and optical networks [19.20]. Furthermore, world economies as a whole are becoming steadily more energy-efficient, as shown by various measures such as the gross domestic product (GDP) per unit of energy use or GHG emissions (e.g., GDP per unit of energy use, Google Public Data Explorer, www.google.com/publicdata), particularly for developed countries. GHG emissions are heavily impacted by industries such as concrete production, which have little connection to ICT. Likewise, electricity use includes sizable contributions from heating, lighting, and generation, which have seen large efficiency improvements from sources other than control through ICT. As technologies across all sectors become intrinsically more efficient, gains will increasingly need to come from how these technologies are used, which is where ICT often can have an impact. Thus there is the potential for ICT to play a more important role [19.1, 16].

A 2012 study, which includes comparisons against results from similar studies, places the worldwide use-phase electricity consumption for ICT at 909 TWh/year, including 334, 307, and 268 TWh/year for communication networks, personal computers, and data centers, respectively (see Fig. 19.1b) [19.3]. Note that ICT has no standard classification system, and therefore any estimates regarding energy efficiency must indicate the technologies under consideration as part of ICT. Worldwide electricity use is roughly 20 000 TWh/year, placing this classification of ICT at 5% of the total. Communication networks are estimated to represent 1.7% of worldwide electricity consumption, which includes both wireless and wire-line networks. It is difficult to extract the specific contribution of optical networks, given that deployed commercial networks, which these numbers represent, include a mix of legacy and state-of-the art technologies. Estimates considering the relative energy efficiency of current-generation optical systems place the core network contribution at 1–2 orders of magnitude below other communication network technologies; however, edge and access optical networks are closer to 10% of the total for communication networks [19.7]. Thus, from an overall energy use standpoint, optical networks have a negligible direct impact. Furthermore, the telecommunications network equipment as a whole is dominated by use-phase energy and GHG impact. Embodied energy from other sustainability life-cycle phases such as manufacturing, deployment, and disposal are below 10% [19.10]. Nevertheless, optical systems play an important role in the overall network architecture and can enable energy savings in other, more energy-consuming technologies.

While optical networks are used throughout core networks, they are in an adoption transition for access and home networks. Since access networks tend to use the most power, this makes it difficult to estimate the optical network contribution unless the relative deployment of optical access networks is known. Networks tend to be hierarchical, and decrease in energy utilization per user or consumer by roughly an order of magnitude or more from the edge to the core [19.7]. At the same time, the capacity of the network equipment increases by many orders of magnitude going from the edge to the core, with regard to consumer networks. Therefore, the efficiency of the equipment on an energy-per-bit basis is improving by orders of magnitude from the edge to the core of the network, benefiting from the efficiency of the aggregation of traffic in the core.

Despite the low energy use of optical networks compared with other communication technologies, the energy efficiency of optical networks is still an important parameter to improve, particularly from an optical network scalability perspective [19.15]. As optical networks find greater application and as network capacity increases at near exponential rates, the energy efficiency of optical networks must improve proportionately in order for them to scale—i.e., achieving similar or smaller footprint and cost as the capacity increases. Studies have shown that, like other communication hardware, the energy efficiency of optical systems is not keeping up with traffic growth [19.7, 21, 22]. The consequence is that optical system racks have reached thermal density limits for air cooling in central offices and data centers [19.10]. Increased capacity therefore requires more floor space and drives up cost. As higher bandwidth is needed near the edge of the network, optical networks will be used in higher-volume applications, which again, have smaller footprint, power, and cost requirements. This trend is already apparent in data centers, where thermal density and form factor are critical design criteria. This data center focus on energy efficiency has been influencing other networks where optics play a role, such as metro and inter-data-center (IDC) networks.

Key points are as follows:

- ICT energy use is a small portion of global energy use and GHG emissions, and optical network energy use is a fraction of ICT energy use.
- While rebound effects can occur, energy efficiency improvements from ICT and optical systems have shown no evidence of a rebound effect, although more data is needed.
- Optical networks are more efficient in terms of energy per bit in the core network system than in metro or access network systems.
- Hardware in optical systems is reaching thermal density limits for data centers and central offices, and energy efficiency has become an important metric in the design and scalability of optical systems.

19.1.2 Fundamental Energy Considerations

Although data transmission is highly efficient using optics, optical devices themselves can have a large device power requirement relative to other ICT devices, such that they become energy-efficient on a per-bit basis only when transporting large volumes of data. This characteristic is not fundamental for optical systems, and as photonic devices mature and take advantage of photonic integration, they can be expected to achieve very low-energy operation. In fact, a major challenge for optical devices is the ability to reach power levels that will enable the use of optical interconnects in computer backplanes and even at the chip scale [19.23, 24]. However, the energy of a photon, $E = h\nu$ (where h is Planck's constant and ν is the optical frequency), is several orders of magnitude larger than thermal, kT, or electronic junction energies, qV (where k is Boltzmann's constant, T is temperature, q is the charge on an electron, and V is voltage) [19.25–27]. Therefore, electronics are fundamentally lower-energy than optics per computational unit/operation. It is only when the data is transported over some distance that optics become more efficient. In current systems, the transition point tends to be around 100 (Gb/s) m, i.e., transporting data at 10 Gb/s over more than 10 m with optical interconnects is more efficient (and cost-effective) than with electronic interconnects [19.21]. Furthermore, modern advanced integrated electronics also enables electronics to be far more efficient than optics in computational tasks.

The energy use required for communication has no fundamental lower bound [19.26]. In principle, one can create an adiabatic or regenerative system that recovers information from the communication medium, e.g., photons, without destroying or changing the energetic state of the photon. Initial energy is required to set up the system (create the photons), but once the link is established, there is no fundamental lower limit on energy use. However, communication systems operate irreversibly, removing energy from the medium in order to read the information. Furthermore, practical systems are not perfectly isolated from the environment and exhibit loss, requiring energy to overcome the loss. In such a destructive or irreversible measurement-based communication system, the minimum energy per bit is set by thermodynamics at $kT \ln 2$. Also note that this adiabatic limit corresponds to the minimum entropy of a digital bit, $k \ln 2$, which provides a physical connection to the Shannon channel capacity limit. This limit can be written as the maximum communication rate at which error-free communication can be achieved as a function of the energy per bit [19.28, 29]. In this way, the details of the communication channel can be accounted for, including the modulation format and statistics of the noise source, and further related to a minimum required energy per bit, for both quantum and classical channels [19.30].

Perhaps the most well-known result is the minimum number of photons per bit for a bit error rate of 10^{-9} for on–off keyed transmission at 10 photons/bit, often referred to as the semiclassical receiver sensitivity limit. It is also notable that in the low communication rate limit, the quantum and classical channel capacities converge and decrease without bound [19.30]. Low-rate communication can be achieved with modulation formats such

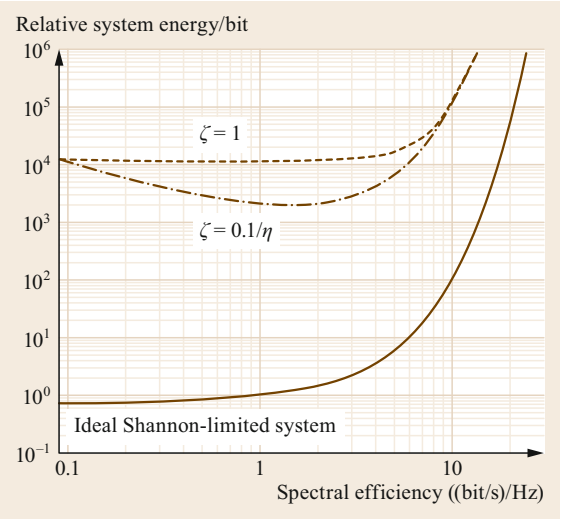


Fig. 19.2 System energy efficiency versus spectral efficiency relative to the Shannon channel capacity limit [19.31] (© IEEE 2011)

as pulse-position modulation, which sends a large pulse in a particular time slot, often used for space communication. Note that use of error correction codes on a channel operating at the sensitivity limit would appear to allow for even better performance. In this case, however, the error correction coding needs to be included in the channel model. Once the error correction codes are properly accounted for, the correct limit can be found, which may be lower than the corresponding limit for the physical channel by itself. For the case of on–off keyed modulation, the lower bound for error-free transmission with coding is in fact 1 photon/bit [19.28].

The minimum energy per information bit can be derived from the channel capacity, $E_b = (2^\eta - 1)/\eta$, where η is the channel capacity and E_b is the energy per information bit measured relative to the received energy per bit for a signal-to-noise ratio of 1 [19.31]. The total system energy can be decomposed into a component that is proportional to the channel energy per information bit and another component that is independent of the signal energy, $E_{bsys} = \gamma E_{b^*} + \zeta E_0$, where γ is the bit energy proportionality factor, E_0 is the fixed energy component, and $\zeta = F/\eta$ is a factor accounting for the efficiency benefit of parallelism. Figure 19.2 assumes that γ is 10^3 and E_0 is four orders of magnitude larger than E_{b^*}. In this regime, the system energy is independent of the channel capacity up until very high spectral efficiencies. Increasing system capacity through multiple transceivers in parallel results in a constant energy per bit and $F = \eta$. However, if these parallel transceivers can be integrated into a common platform or line card, then there is an integration benefit corresponding to a reduction in the fixed energy component by F/η. For example, if 10 transceivers can share the fixed component equivalent of one transceiver through integration, then $F = 0.1$, and optimal spectral efficiency appears. This analysis can be expanded to consider the benefits of parallelism for different system reach and spectral efficiency, which will be described in more detail in Sect. 19.3.1.

Key points:

- While there is no fundamental lower limit to energy use in communication, practical systems are limited by thermodynamic principles related to the Shannon channel capacity limit.
- In computation, optical systems require orders of magnitude more energy per bit than electronic switch (transistor)-based systems unless transmission over distance is involved.
- Communication systems require much larger energy per bit at high spectral efficiencies.
- In systems limited by the bit-rate-independent components, an energy optimal spectral efficiency can be found through the integration of multiple communication channels.

19.1.3 Commercial Systems and Networks

Commercial systems and networks operate at efficiencies that are many orders of magnitude worse than the fundamental limits, for many practical reasons, similar to the scenario depicted in Fig. 19.2. An optical communication system is typically configured in a chassis with pluggable line cards. Line cards, similar to pluggable electronic boards in a computer, carry electronics and modules for each of the major components: transceivers, amplifiers, multiplexers, and switches, along with management cards for interfacing and controlling the system. Historically, the energy use of optical line cards was dominated by the control electronics, which can be thought of as similar to a laptop computer. The optical transceiver hardware itself used much less energy. As a result, as the line-card data rate increased, the energy use of the line card remained largely unchanged, resulting in excellent energy scaling. In recent years, the heavy processing required for signal recovery has changed this situation such that the communication-related electronics use more power than the control electronics [19.10]. The energy use of commercial systems and networks is considered in detail below.

19.1.4 Energy Metrics for Optical Networks

Energy in electronic equipment is measured through electricity consumption in units of kilowatt-hours (kWh), where 1 kWh corresponds to 1 kW of power de-

livered over 1 h or 3600 kJ of energy. In metrics that use the actual energy use or energy bill-related accounting, the energy can be expressed as a power by including a time period for the measurement such as kWh/year. The optical communication equipment in optical networks consumes electricity while delivering data between ingress and egress points. Therefore, the energy efficiency is often specified in terms of the energy per bit of data delivered. For the transmission of data, the energy per bit can be written as the power P (in $\mathrm{W} = \mathrm{J/s}$) required to support a given data rate D (in bits/s), $\varepsilon = P/D$. Measuring efficiency using the power and data rate is often more convenient for optical network equipment because the equipment is specified for a certain operating power and the network interfaces deliver a specified data rate. Detailed descriptions of common energy metrics are provided by energy standards groups, scientific communities, and governmental agencies (e.g., in the US: www.NERC.com, https://www.aps.org/policy/reports/popa-reports/energy/, www.eia.gov).

For an individual device or component, determining the energy efficiency is a straightforward calculation of the device power and speed or data rate. For a network, energy efficiency is far more complicated, because both the power and data rate have contributions from many different sources that may or may not be relevant to the network efficiency of interest. For example, the data carried in a network may include protocol or forward error correction information that is intrinsic to the network design and operation, not the data delivered to the end user. Thus, using the data rate of the transceivers in the network for the energy efficiency computation would overestimate the actual end-user data that the network is handling. Also consider that many networks are operated at low levels of utilization in order to avoid delays due to congestion. This low utilization would further reduce the network efficiency, and represents a design trade-off of reliability in favor of energy efficiency. The term *goodput* is used to refer to data delivered to the end user. While goodput is useful for evaluating the overall efficiency of a data network, it can be a difficult quantity to obtain and is therefore rarely used. The energy per information bit is usually used for encoded data to account for the coding overhead, where *information bit* is the actual payload data or useful information that is transmitted on the channel.

Determining the relevant power specifications to use in energy efficiency calculations is also complicated for networks. The power specified for commercial equipment may be the peak power, while the actual operating levels can be significantly lower. This power specification may also include inefficiencies due to legacy feature support or implementation design choices that may not reflect the actual power used for transporting the data. Thus the equipment power specification alone may not be useful for understanding the minimum power required for a network or the power due to the communication components alone. Furthermore, some equipment can operate using power management modes such as sleep mode or voltage and frequency scaling on the electronics, which would further reduce the equipment power from the maximum rating and render the power a complex function of time and traffic in the network. Time variations can be dealt with either explicitly or by considering averages over a period of time.

In addition to measuring the energy efficiency of an optical network using the energy-per-bit metric, one can consider energy efficiency from a relative improvement over a reference case. The reference case can be an instance in time or a particular architecture. This approach is particularly important when it is necessary or desirable to work with energy or power rather than energy per bit. Reporting the total power of the equipment in a particular network or architecture is generally not very meaningful, as there is no good reference for comparison.

Network energy use can also be specified in terms of carbon footprint or greenhouse gas emissions. Optical network equipment runs entirely off of electricity, and therefore these metrics require the electricity source to be known or given/assumed. Network equipment at a location where the electricity is generated entirely by renewable energy sources, such as hydropower, will have a zero-carbon-emission use-phase contribution. In some cases, the network can be used strategically based on the source of electricity and availability of renewable energy. Several proposals have been made to use networks to move data to processing locations where renewable energy is available [19.32]. When dealing with sustainability metrics, the entire sustainability life cycle is relevant. This accounts for embodied energy such as the energy required to manufacture, deploy, and finally dispose of the optical networking equipment, including the fiber in the ground.

Key points:

- Optical communication energy efficiency is often characterized using energy-per-bit or power-per-bit rate. Absolute power or energy measures need to be compared against a reference system or network. Appropriate averaging over time should be taken depending on the application and operating conditions of the respective system.
- Data on both the power and the data rate for optical networks can include contributions from many factors and can be measured in many different ways.

- Equipment and component power ratings should specify whether peak, mean, or typical power is used and what features are included in the rating.
- Data rates should account for overhead, protocol messages, and other data that is not part of the user or application data, which is often referred to as the network goodput.
- Carbon footprint metrics need to account for the electrical energy source and specify the phase of the sustainability life cycle.

19.2 Optical Components and Equipment

The energy efficiency of the individual optical components and equipment is a major element of the energy efficiency of optical systems and networks. The architecture, determining how the equipment is organized and used, is another major element of the overall efficiency and will be described in Sect. 19.3. A network can be decomposed into switching and transmission hardware. The basic building block of the transmission hardware is the communication link, which is composed of a transmitter, transmission conditioning equipment, and a receiver. The transmission conditioning equipment can include optical amplifiers, dispersion compensators, optical power attenuators, and other devices. Optical switches can operate on multiple dimensions, mostly space and wavelength, but they can also operate on properties such as polarization or spatial mode. A space switch connects a fiber at one spatial location to another fiber at a different spatial location, which are referred to as space switch ports. A wavelength switch can have single or multiple input and output fibers and switch or block signals of different wavelengths between different fibers. These components and their energy efficiency properties are described below.

Energy use can be analyzed through either a top-down or a bottom-up approach [19.33]. In the top-down approach for optical network equipment, the equipment specifications of the system (transmission system) or subsystem (optical line card) are used, and the specific contributions of the underlying components are either ignored or inferred. This is useful for obtaining the actual efficiencies that are seen in practice, but there may be implementation factors that have little to do with the optical communication function that will influence the results. Implementation factors can include board control electronics or standards for electrical power regulation and delivery. In the bottom-up approach, the energy efficiency of the specific optical communication components such as modulators and lasers are accounted for and used to build up and calculate the energy efficiency of a reference system. This approach underestimates the energy use of a commercial implementation, but provides insight into what might be achieved if the commercial system were specifically designed for maximum energy efficiency, and it provides the relative contributions of the individual components. The choice of top-down or bottom-up analysis depends on the goals of the analysis and is particularly important when considering different architectures (see Sect. 19.3).

The main contribution to energy use in both optical transmission and switching equipment is often from complementary metal–oxide–semiconductor (CMOS) electronics. Optical devices such as signal lasers, pump lasers, photodetectors, and microelectromechanical systems (MEMS) actuators can be small in comparison. Temperature control hardware, both within the optical components themselves and in the subsystem (e.g., erbium-doped fiber amplifier EDFA module) packaging will often dominate the electricity consumption for the optical devices and subsystems. The CMOS controller electronics on an optical system line card can be similar to a laptop computer—roughly 30–50 W peak power. Transceiver line cards can include framing electronics as well as digital signal processing (DSP) electronics for signal recovery and forward error correction. These components can be 100 W or more, particularly since they are running at high speeds. For this reason, the optical system energy efficiency is strongly dependent on the efficiency trends of CMOS electronics [19.34]. Figure 19.3a shows the evolution of silicon functional scaling over time for different electronic platforms, and Fig. 19.3b shows the energy efficiency trend in CMOS processors over time [19.35]. Moore's law trends are tracked by the International Technology Roadmap for Semiconductors (ITRS), which was recently revised (www.itrs2.net). Each generation shows a jump in the number of functions, whereas the power of the individual chips remains relatively constant, resulting in greater efficiency. Optical networks can be made more efficient simply by upgrading to the latest generation of electronics. However, energy efficiency improvement for CMOS is seen to be slowing [19.36]. This is due in part to the fact that in the past, CMOS voltage was steadily reduced with gate size. At voltages of around 1 V, leakage current becomes an obstacle to further voltage reduction [19.37]. The main contribution to the power is the CV^2 (where C is the transistor

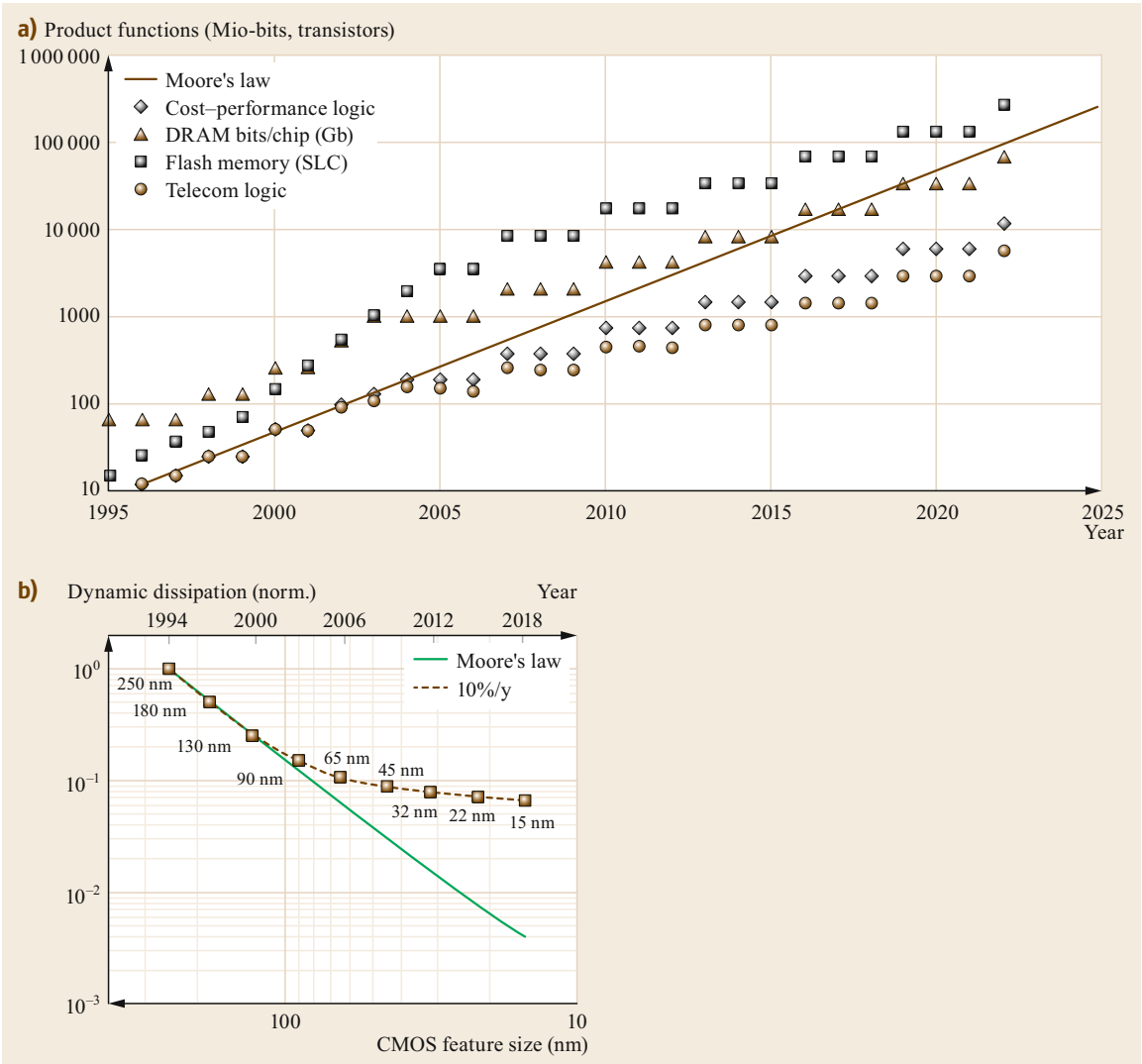


Fig. 19.3a,b Silicon scaling over time **(a)** by feature size [19.34] (© Wiley & Sons 2010) and **(b)** with respect to relative energy use of transistor switching [19.35] (DRAM: dynamic random access memory; SLL: single-level call)

capacitance and V is the voltage) losses, and thus the squared voltage reduction has a large effect. The physical dimensions of CMOS chips have also increased, leading to higher on-chip transmission losses, which is now limiting the chip size. These changes have resulted in single-chip processor power leveling off at 100–200 W [19.24, 36]. If further processing is needed beyond what can be delivered at these power levels and available footprint, then multiple chips are used, resulting in overall larger footprint, power, and cost.

Key points:

- Integrated electronics use the most energy in commercial optical communication systems, and therefore the energy efficiency of these systems is tied to Moore's law trends.
- Energy efficiency improvement due to Moore's law scaling has been slowing, and therefore efficiency improvement in optical systems over time has been slowing.

19.2.1 Optical Transmission Equipment

The basic transmission building block of an optical network is the optical link. In its simplest form, it is just a transmitter and receiver with fiber, waveguide, or free space in between. Several studies have looked at the minimum achievable energy per bit for an op-

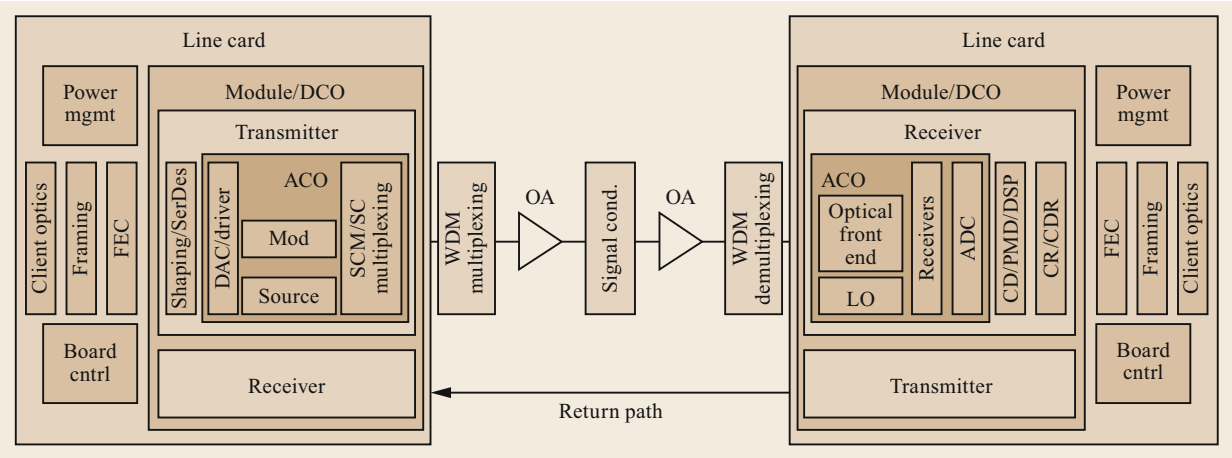


Fig. 19.4 An optical transmission link with subsystems and components

tical link [19.23, 38], and femtojoule-per-bit operation has been demonstrated [19.39]. Fundamentally, the minimum energy per bit is determined by the Shannon channel capacity (i.e., for a destructive communication system—see Sect. 19.1.2). Commercial links are composed of modules and components shown in Fig. 19.4. Typically, a line card or shelf will provide support for one or more line-side transceivers, which include a transmitter and receiver for use in the communication link. Electrical or optical client interfaces provide data to the line card (here *client optics*). Today, transceivers are either externally pluggable modules or modules that can be board-mounted at the time of system assembly. The various subcomponents shown here may be found on the line card or the modules, depending on the particular module type and system requirements. For systems using advanced modulation and coherent receivers, the modules can be organized as analog coherent optics (ACO) or digital coherent optics (DCO), as shown. The ACO configuration is often used because the power associated with the digital signal processing electronics can exceed the maximum allowed for the desired pluggable form factor. In some cases, the system vendors also prefer to use their own custom DSP implemented on their line cards. The line hardware including optical amplifiers (OA), signal conditioning (e.g., dynamic gain-equalizing filters), and optical multiplexing and switching hardware will be housed in a separate line card or shelf.

In accounting for the energy efficiency of an optical link, there is wide variation in which of the components in Fig. 19.4 are included. In optical device studies, it is common to report the power of the components in the ACO block (not including the digital-to-analog/analog-to-digital converters and drivers), as these correspond to the opto-electronic components. Sometimes *serdes-to-serdes* energy efficiency is reported.

Table 19.1 DSP Contributions broken down by function and CMOS generation [19.40] (© IEEE 2014). Energy per information bit consumed by various DSP component (pJ/bit)

CMOS process technology	40 nm	28 nm	20 nm
Encoding	36	25	18
Pulse shaping	13	9	7
CD compensation	162	113	81
Timing recovery	52	37	26
PMD compensation	123	86	61
Carrier recovery	23	16	12
Decoding	86	60	43
Total	495	346	248

State-of-the-art commercial systems in core networks primarily use advanced modulation formats with coherent detection [19.41]. Figure 19.5 shows the energy per information bit for different transmission configurations, and Table 19.1 provides the digital signal processing contributions considering different CMOS generations [19.40]. Note that 100 Gb/s at 1000 pJ/bit corresponds to 100 W. Dual-polarization quadrature phase-shift keying (DP-QPSK) and DP 16-quadrature amplitude modulation (QAM) are widely used in commercial systems. The subcomponents in the bar graphs correspond to different elements shown in Fig. 19.4 and include (for Fig. 19.5a): I–laser, II–encoding, III–pulse shaping, IV–p/s conversion, V–DAC, VI–modulator, VII–booster amplifier, a–CD compensation, b–timing recovery, c–PMD compensation, d–carrier recovery, e–decoding; (for Fig. 19.5b): I–transmitter, II–link EDFAs, III–receiver (excluding DSP), IV–DSP. Note that stronger codes are used for the longer distances.

Note that this analysis uses a bottom-up approach in calculating the energy efficiency and thus does not account for inefficiencies associated with partic-

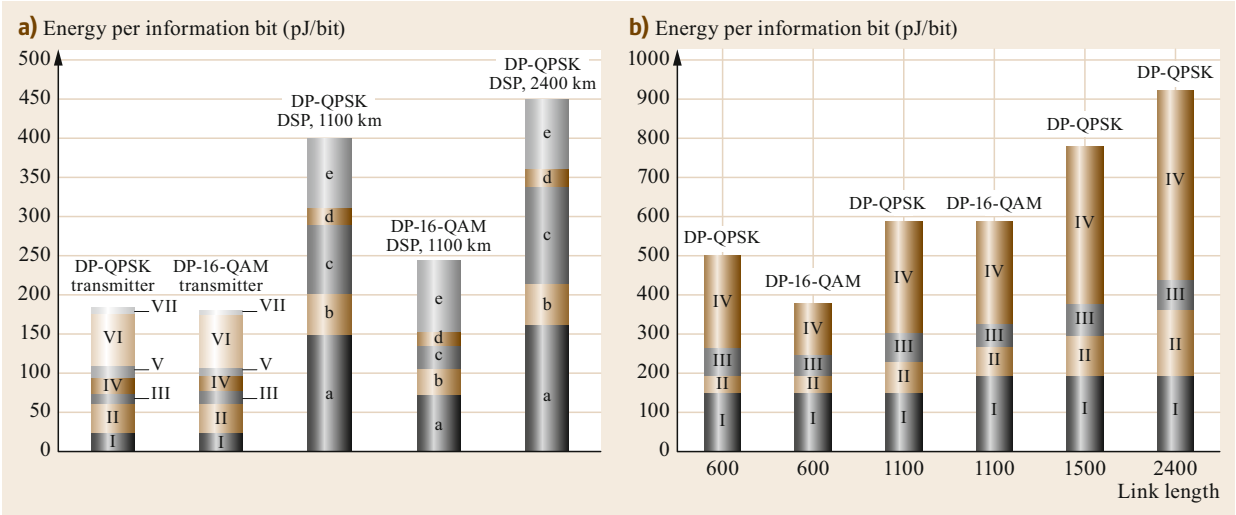


Fig. 19.5a,b Energy-per-information-bit contribution for different transmission system configurations using coherent detection with either QPSK or 16-QAM modulation for **(a)** transceiver and **(b)** full transmission system [19.40] (© IEEE 2014)

ular commercial implementations. In general, a given commercial implementation is likely to be more energy-consuming by some amount than shown here. The energy efficiency of the CMOS components changes by a factor of 2 going from 40 to 20 nm, and thus it is important to account for the CMOS generation when considering the DSP components.

The industry is working to reduce the footprint and power of advanced modulation format transceivers [19.42]. DCO modules have recently become available in the same CFP2 (C form-factor pluggable 2) package used for ACO modules, through the use of integrated photonics and other advances [19.43]. Note that CFP2 is a standard pluggable optical transceiver package, sometimes referred to as the transceiver form factor; other common standards include small form-factor pluggable, SFP and quad-SFP, QSFP, each of which have power limitations based on the corresponding multisource agreement specification.

On–off keying modulation is still found in many long-haul systems, often in combination with advanced modulation format signals, and is widely used in metropolitan and access networks. Dense wavelength-division multiplexed (DWDM) tunable SFP+ (+ stands for 10 Gb/s) modules with tunable lasers designed for long-distance transmission can be found with power ratings of 2 W and therefore an energy per bit of 200 pJ/bit (at 10 Gb/s data rate). Note that when on–off keying modulation is used with advanced modulation formats, a guard band is required due to the impact of cross-phase modulation from the on–off keyed signals to the phase-modulated signals. Furthermore, 10 Gb/s signals may require a dispersion map and other dispersion compensation along the path and in some cases at the receiver, reducing the energy efficiency [19.44].

The energy efficiency of on–off keyed modulation-based transmission systems will depend on the modulation technology and the set of technologies needed for the particular distance or application. For example, data center links include only a transmitter and receiver, with a strand of fiber in between. Forward error correction (FEC) coding is not used, and direct modulation of the transmitter laser is possible. This results in highly energy-efficient links, in the 10 pJ/bit range [19.38]. For longer-distance transmission, amplifiers are used between the transceivers to make up for transmission losses, FEC is used, and external modulators are needed to reduce chirp from direct modulation. An example of the design trade-offs for energy efficiency in short-reach links is shown in Fig. 19.6 [19.45, 46]. Increased modulation amplitude increases the transmitter energy use while reducing the receiver energy use. This shows the benefit of moving to low optical signal power in order to minimize the transmitter power, which is dominant. As the receiver sensitivity limit is approached, however, the energy use rises dramatically due to the additional processing and electronics that are required to compensate, such as high-gain amplifiers and error correction coding.

For longer-reach on–off keyed systems, the energy use includes contributions from the link components such as optical amplifiers and dispersion compensation. One study examined the trade-off between the energy due to electronic dispersion compensation and the additional amplifier energy needed to overcome the losses of

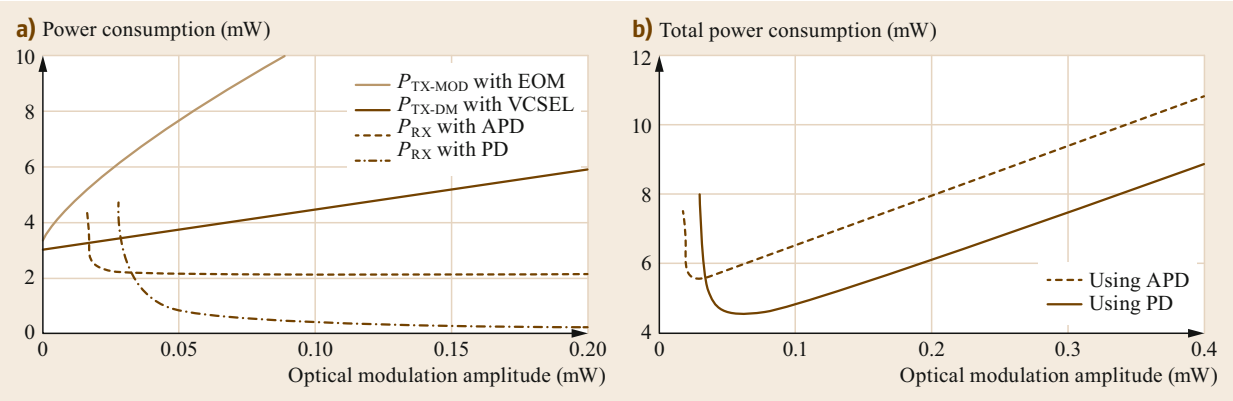


Fig. 19.6a,b Power consumption for a 1 Gb/s point to point link across 12 dB loss, **(a)** transmitter power with external optical modulator (EOM) and direct modulation (VCSEL) and receiver with avalanche photodiode (APD) and PIN photodiode (PD); **(b)** total power of directly modulated link (after [19.45])

fiber-based (passive) dispersion compensation [19.44]. Numerous other studies looking at the energy efficiency of optical systems are described in Sect. 19.3. In general, similar to coherent systems, the transceivers are the high-power-consuming elements, as might be expected since the amplifiers are shared among the 80 or more DWDM optical channels.

Key points:

- Transceivers use the most energy in typical optical communication systems, and digital signal processing is responsible for the largest portion of the energy use for coherent systems.
- Transceiver energy use is divided into optical module and line-card components, which may differ depending on the implementation and standardized module availability.
- For unamplified links, transmit power can be reduced to improve efficiency until the sensitivity limits of the receiver are approached.

19.2.2 Optical Switching Equipment

Optical switching equipment is used at optical nodes to add and drop wavelengths and set the light path for wavelengths passing through a node. Optical switches used in core networks are primarily wavelength-selective switches (WSS), although space switches are also used in some architectures [19.21, 47, 48]. These switches operate in a *set-and-forget* mode and do not switch continuously. Therefore, the optical switch energy is associated primarily with holding the state, which can be quite small. For MEMS-based switches, active control, including dithering, may be used to hold the mirror position for optimal output power. Latching mirrors can also be used, which results in negligible holding energy. Recent work on integrated photonic switches is promising for reducing energy use along with cost and footprint. For example, a MEMS-based silicon photonic crossbar switch uses latched switching for very low-power operation [19.49]. Therefore, thermal management or control electronics will often dominate the energy use of the optical switches.

The node architecture for core networks today uses a reconfigurable optical add-drop multiplexer (ROADM) structure (see Fig. 19.7; [19.47, 48]). ROADMs can be designed for colorless (i.e., add and drop ports can support all wavelengths), directionless (i.e., any transceiver can add or drop signals on any input/output line-side fiber), and/or contentionless (i.e., any transceiver can add or drop signals at the node without blocking another transceiver from adding or dropping signals at the node) operation. A colorless/directionless/contentionless (CDC) ROADM supports all three of these attributes. The node in Fig. 19.7 can be used for CDC operation. The switching-tier port-counts ($k \times 1$ where $k = 2$ in Fig. 19.7) determine the higher-degree scalability of the node. High-port-count WSS switches can also use ports for aggregation, eliminating the need for aggregation-tier switches. Note that a passive splitter/coupler can also be used for colorless multiplexing, replacing some switches but potentially adding loss and reducing functionality. An optical channel monitor (OCM) can be shared among different fibers in the node and is used to detect the presence of the signals and their optical power levels—needed when the attenuator function in the WSS switches is used to adjust the signal power.

Optical switching equipment is often used together with electronic switches. Figure 19.8 shows the rela-

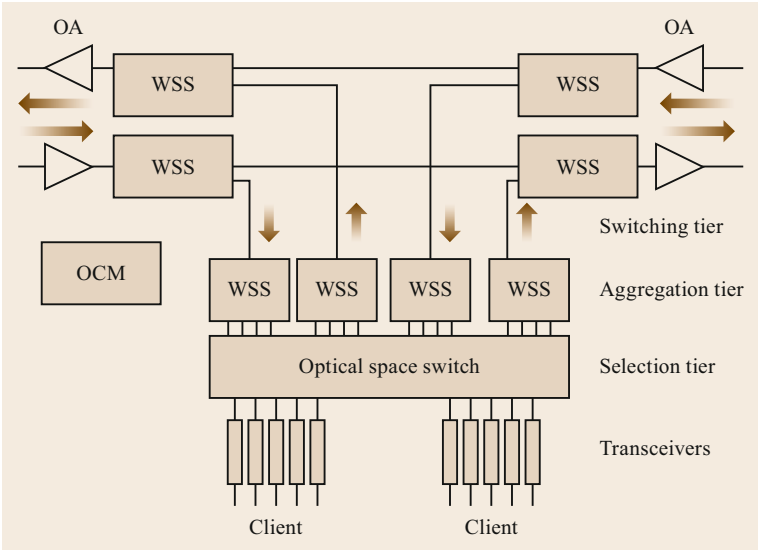


Fig. 19.7 Reconfigurable optical add-drop module (ROADM) switching node, degree 2

tive energy efficiency of different commercial switching technologies used in core networks in 2008 numbers [19.34]. Routers are roughly five times less efficient than optical switching hardware, and electronic circuit switches, SONET/SDH and ODU/OTN (currently replacing SONET), are 50–100% less efficient. Each of these electronic platforms provides significantly greater processing functionality than an optical switch, and in fact the optical switch in this analysis includes transceivers, which account for the largest share of the power. This analysis demonstrates the cost of different levels of advanced switching functionality, from optical switching without processing, to circuit, packet, and deep packet inspection. Note that the analysis in Fig. 19.8 was performed by taking a survey of publicly available data on commercial equipment and decomposing the total power into respective components using a reference model. For the switching components that are dependent on the CMOS technology generation, conversion factors are used in order to compare equipment from different generations; thus, here the equipment has been adjusted to 2008 CMOS (see Fig. 19.3).

Numerous studies have considered active optical switching, such that the optical switches operate through a control plane that responds to time-dependent

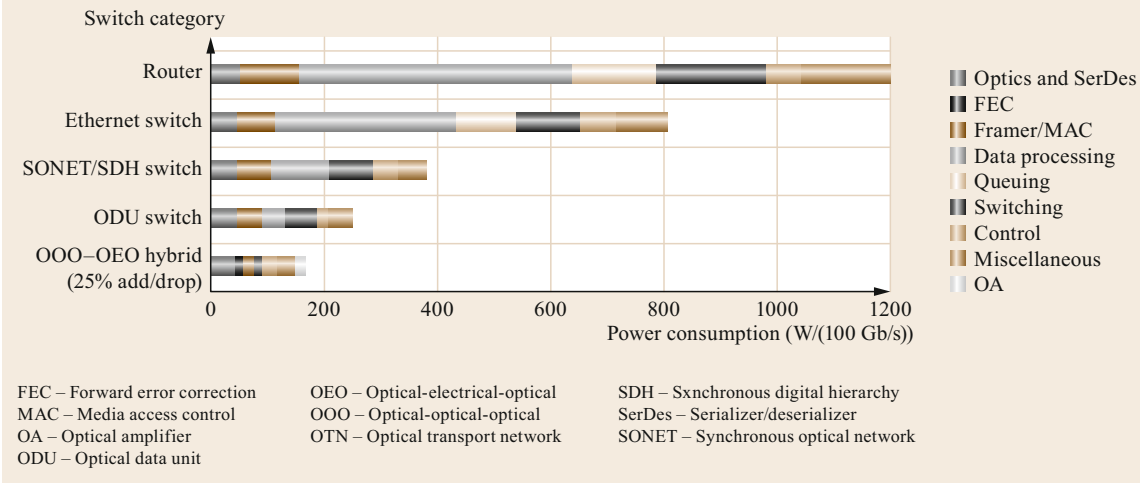


Fig. 19.8 Energy efficiency of optical and electronic switching node hardware [19.34] (© Wiley & Sons 2010); average power consumption normalized to 2008

data traffic [19.50]. In fact, some approaches, such as subwavelength switching, have been commercialized but have not seen widespread adoption. Much of the value and energy use of switching in communication systems comes from the control plane and data processing in the switches, such as that derived from deep packet inspection. These are computing-intensive functions that cannot be addressed with optical switches [19.51, 52]. However, optical switches can be used to bypass this processing when it is not needed. Studies have shown that IP packets go through 10–20 routers on their way to their destination [19.5, 7]. In many cases these routers may be performing unnecessary processing or may be associated with the exchange of data through different network domains, i.e., operator infrastructure. Methods to keep signals in their native optical form for as much of the path as possible can reduce the power by factors even greater than is shown in Fig. 19.8, as the optical switching in Fig. 19.8 includes the transceiver energy. This approach requires new overall architectures including both the electronic and optical switching elements, which will be discussed in Sects. 19.3 and 19.4.

Key points:

- Commercial optical switches do not actively switch, and therefore the holding or latching energy provides the main switching power requirements.
- Optical switch energy is dominated by the power of the switch control and thermal management electronics.
- Commercial optical switches are typically more energy-efficient than electronic switches, but provide significantly less functionality.

19.3 Energy-Efficient System and Network Architectures

This section examines how the organization of components and network elements into systems and networks affects energy efficiency. Particularly important at this level of analysis are the different methods for modeling energy use. Different modeling methods answer different questions and must be understood well in order to properly interpret the results. Therefore, this section reviews the main modeling approaches, including both the optical systems themselves and the traffic characteristics and modulation used to transfer data over these systems. The following Sect. 19.4 looks at important energy efficiency studies and trade-off analyses that use these different modeling approaches.

As discussed above with regard to components, understanding energy use in optical systems and networks likewise strongly depends on whether a top-down or a bottom-up analysis is used. The top-down approach [19.3, 6, 8, 53] in this case may extend to considering the energy bills of network operators and their published network traffic volume to obtain overall network energy efficiency. Equipment can be further subdivided into different categories based on the assumed or known composition of network and service offerings. The top-down approach is frequently used when analyzing commercial systems for which global information such as electricity consumption may be publicly available but the specific details of the network are proprietary. Top-down studies can also use information such as the number of units of certain types of equipment sold or delivered. A top-down approach is typically more accurate for determining the energy efficiency of deployed network services, as it incorporates all aspects of the business that may go into the service delivery. Combining public energy figures with total traffic data for the network or service running on the network can also give the appropriate energy per bit. Such methods have been proposed for use in *green* service-level agreements or Energy Star ratings for network services [19.54]. On the other hand, top-down analyses may not be useful or accurate for understanding how or why the network is more or less efficient. Top-down information on one network operator will reflect the business priorities and implementation choices for that network operator, which might not be generalizable. For example, one provider's network might have a large amount of legacy equipment in order to support particular services that only work with that equipment, which has very low energy efficiency [19.9]. Other factors such as the building and power facilities may also play a role [19.55].

On the other hand, as with the case of the communication equipment described above, understanding the effects of different network architectures and system designs on the network energy efficiency often requires a bottom-up analysis. In Sect. 19.3.1, bottom-up energy analysis methods are described, accounting for how system architectural considerations affect overall network energy efficiency. Network-element energy efficiency considerations are described in Sect. 19.3.2, including traffic-dependent energy use, power management, and node common equipment. How the network is structured or layered and how each layer handles traffic will also have an important impact on energy

efficiency, as described in Sect. 19.3.3. Traffic models, described in Sect. 19.3.4, are another dimension to understanding energy use in networks, and need to be chosen carefully when evaluating the energy efficiency of a particular system or network. Finally, the coding and modulation used within a system will affect the system performance and energy efficiency. The energy efficiency of coding and modulation, including modeling methods, is reviewed in Sect. 19.3.5.

19.3.1 Modeling Energy Efficiency of Optical Systems

In general, two main approaches are used in bottom-up modeling of network energy efficiency. The first is a network-based model, which involves summing up all of the contributions from the components that make up the network. In this approach, a reference network is needed to account for the entire network (Fig. 19.9a). Optimization studies will typically take this approach, incorporating the appropriate network energy models within the objective functions and constraints. The second approach is a path- or transaction-based model, which breaks the energy down by service, considering the path that a service transaction takes through the network, for example the path that a consumer Internet transaction takes from a home to a data center (Fig. 19.9b). The transaction model is useful for modeling the dependence of the network energy of different services on the constituent networks and equipment, and their technology evolution [19.5, 7, 56, 57]. Note that the parameter values in these equations can reflect unique instances or averages over time. As discussed in Sect. 19.1.4, how the power and traffic quantities are defined depends on the specific objectives of the analysis and should be carefully considered.

Network-Based Models

$$P_{\mathrm{Net}} = \sum_i P_{ci} \,, \tag{19.1}$$

$$P_{\mathrm{Net}} = \sum_j P_{\mathrm{N}cj} + \sum_k P_{\mathrm{L}ck} \,, \tag{19.2}$$

$$P_{\mathrm{Net}} = \sum_j P_{\mathrm{N}0j}\left(\delta_j\right) + P_{\mathrm{T}j}\left(B_j\right) + P_{\mathrm{N}\delta j}\left(\delta_j\right) + \sum_k P_{\mathrm{L}0k}\left(d_k\right) + P_{\mathrm{A}k}(d_k, B_k) \,, \tag{19.3}$$

$$\varepsilon_{\mathrm{Net}} = \frac{P_{\mathrm{Net}}}{D_{\mathrm{Net}}} \,, \tag{19.4}$$

$$\varepsilon_{\mathrm{NetMD}} = \varepsilon_{\mathrm{NetCore}} + \varepsilon_{\mathrm{NetMetro}} + \varepsilon_{\mathrm{NetAccess}} \,, \tag{19.5}$$

where summations are taken over all i components, j nodes, and k links. The total network power P_{Net} is the sum of all components in the network P_{ci}, which can be grouped into node components $P_{\mathrm{N}cj}$ and link components $P_{\mathrm{L}ck}$. Node components will have contributions from the total ingress/egress traffic at the node B_j and the node degree δ_j. Link components depend on the distance of the link d_k (e.g., through the number of amplifiers). Nodes and links will also include so-called *common* equipment which is independent of or weakly dependent on traffic, such as the cooling, electric power delivery, and control electronics, $P_{\mathrm{N}0j}$ and $P_{\mathrm{L}0k}$, for nodes and links, respectively. The energy per bit can be determined by dividing the network power by the traffic that it handles. When different network domains are included, it may be necessary to decompose the energy per bit by domain to account for the different traffic loads in each domain. Also note that, given the network energy per bit, the power associated with a given data transaction or service can be found from the product of the number of bits

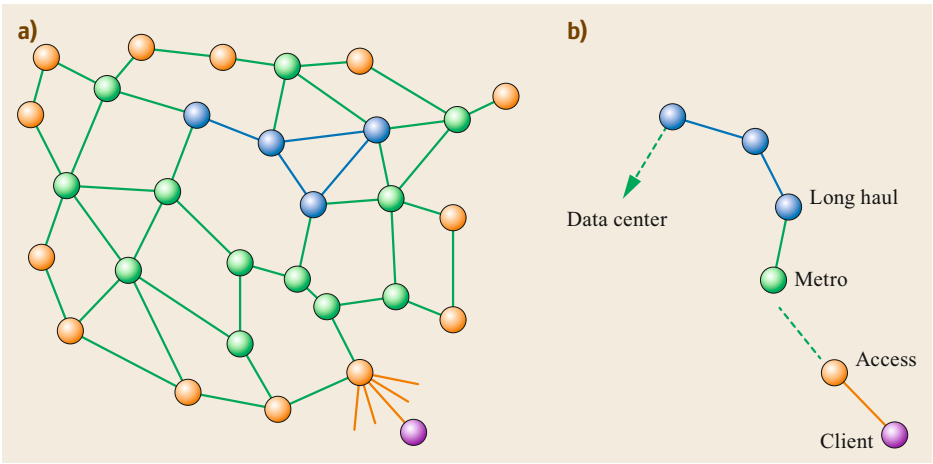


Fig. 19.9 **(a)** Network-based model accounts for an entire reference network and **(b)** a transaction-based model accounts for a data or service path through the network

and the energy efficiency of the respective networks that handle the service.

Transaction-Based Models

$$\varepsilon_S = \sum_p \frac{P_{Ncp}}{C_{Ncp}} + \sum_q \frac{P_{Lcq}}{C_{Lcq}} , \tag{19.6}$$

$$\varepsilon_S = \frac{N_R P_R}{C_R} + \frac{N_{Eth} P_{Eth}}{C_{Eth}} + \frac{N_{ON} P_{ON}}{C_{ON}} + \frac{N_{OL} P_{OL}}{C_{OL}} , \tag{19.7}$$

$$\varepsilon_S = N_R \varepsilon_R + N_{Eth} \varepsilon_{Eth} + N_{ON} \varepsilon_{ON} + N_{OL} \varepsilon_{OL} , \tag{19.8}$$

$$\varepsilon_S = \varepsilon_{SCore} + \varepsilon_{SMetro} + \varepsilon_{SAccess} , \tag{19.9}$$

$$P_S = \varepsilon_S D_S , \tag{19.10}$$

where the service energy per bit ε_S is the sum of the energy per bit of the nodes and links that handle the service traffic. The energy per bit in this case is the power of the respective equipment divided by its capacity. Additional multiplicative factors can be included to account for equipment that is used below capacity and for overhead factors such as rack or facility cooling. The node and link equipment power is often separated by equipment type, including routers P_R, Ethernet switches P_{Eth}, optical node P_{ON} equipment, and optical link P_{OL} equipment, and their respective capacities C. The constituent equipment energy per bit is multiplied by the number of each equipment type along the path, and these can be organized according to the network domains through which service traffic is transmitted. The power per user or transaction for the service is the product of the service energy per bit and the data rate of the service D_S.

It is instructive to understand how the power depends on the different node models (e.g., Fig. 19.7) and architectural assumptions behind the models. For example, an optical system built on a 1×4 WSS ROADM will experience a jump in power going from degree 4 to degree 5 because extra amplification and switching components are needed to extend the node to a higher degree at that point. This effect will create a capacity dependence once the maximum number of channels on a link (typically in the range of 80–128 channels) is reached and additional fibers are used—effectively increasing the node degree. Figure 19.10a shows the power of a network as a function of the capacity for different WSS configurations (1×9 and 1×20), channel data rate (50 and 400 Gb/s, fiber type G.652 and G.655), and amplifier type (EDFA and hybrid Raman/EDFA, HRA), utilizing the most efficient technology for each capacity [19.58] (19.1). Here, up to 120 channels per fiber are allowed using flex-grid packing for the EDFA case, and 240 channels for the hybrid Raman/EDFA case, using a wider amplification bandwidth. Flex-grid refers to the case where optical channels are not fixed to a 50 or 100 GHz spaced grid, but instead the channel width and center wavelength can be adapted in steps given by the wavelength granularity of the switches, typically 6.25 GHz or smaller. A mesh traffic pattern was used to route wavelengths for each case. The analysis using the network approach (19.1–19.5) was run on an Italian national network of 44 nodes and 71 links.

Figure 19.10b shows the results of a transaction-based model analysis, which breaks the energy-per-bit contributions down for routers, fiber-to-the home access

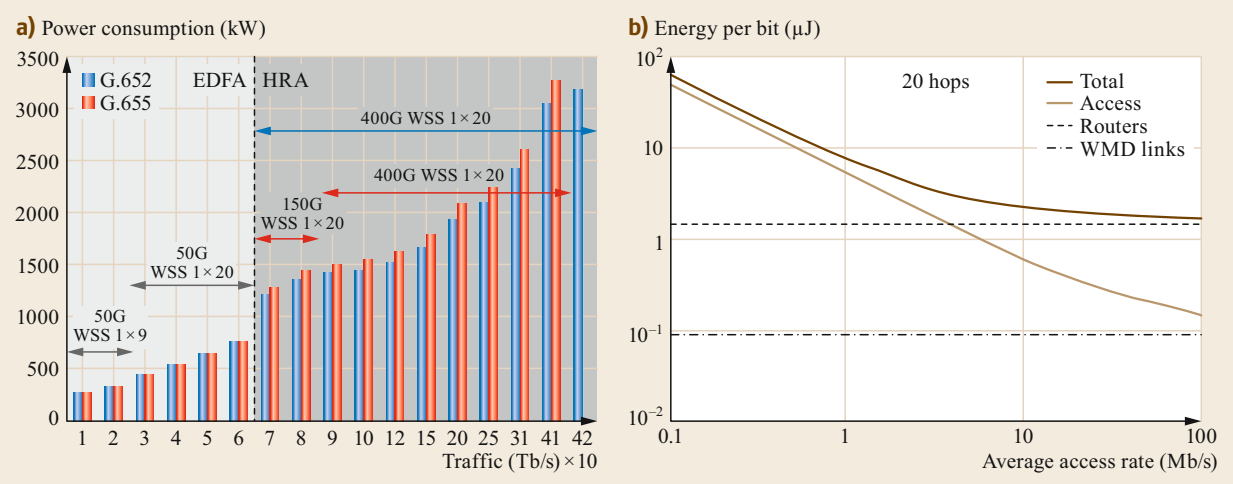


Fig. 19.10 **(a)** Power requirement for Italian national network for different traffic volumes and optical system technologies [19.58] (© IEEE 2014). **(b)** Energy efficiency of telecommunication networks using a transaction-based model, broken down by technology area using 2008 technology [19.59] (© IEEE 2009)

technology, and WDM links. The analysis assumes that the service takes 20 hops through the core network and uses 2008 technology efficiency values [19.59].

Routing and wavelength assignment (RWA) involves identifying an optical path or light path through a transparent network or series of optical switching nodes and choosing the wavelength on which the signal will be carried along that path [19.60]. RWA can be performed as part of the optical network planning using optimization or heuristic algorithms. Optical networks often use some form of impairment-aware RWA, IA-RWA, which takes into account the optical impairments or other performance criteria in selecting the path. RWA can also be performed as part of a real-time control plane in an optical network that performs real-time optical switching. In this case the algorithms often involve heuristics to reduce the computational complexity.

Designing for maximum performance in an optical system may not lead to energy-efficient networks. Shortest-path routing, minimizing the number of nodes and links that handle traffic, tends to improve performance while also saving energy. However, the optical impairments can complicate this association [19.61]. Energy awareness has also been considered for RWA (EA-RWA) [19.62], as well as impairment and energy awareness (IEA-RWA) [19.63]. Typically, performance margins are introduced to account for uncertainty, and larger margins will result in increased regeneration of the optical signals. This will increase the energy use of the network. Regenerator placement algorithms can be designed to take network energy efficiency into account [19.64].

Key points:

- Optical network energy use can be determined using a network-based approach or service transaction-based approach.
- A network-based approach is often used for optimization studies and may include routing and wavelength assignment to determine the network utilization.
- Transaction-based models are used for determining the energy cost of different services and network technologies, including evolution over time.

19.3.2 Energy Efficiency of Network Elements

The network modeling approaches described above often benefit from considering network element-level modeling that may represent the different component combinations described in Sect. 19.2, or represent specific architectural and/or operational design aspects such as power management. Below is a description of optical network energy efficiency considerations related to node common equipment, traffic-dependent equipment energy use, and power management such as sleep mode operation.

Node Common Equipment

An important consideration in practical models for energy-efficient systems is accounting for the system chassis and hardware implementation, including power delivery and cooling systems. Figure 19.11 shows ex-

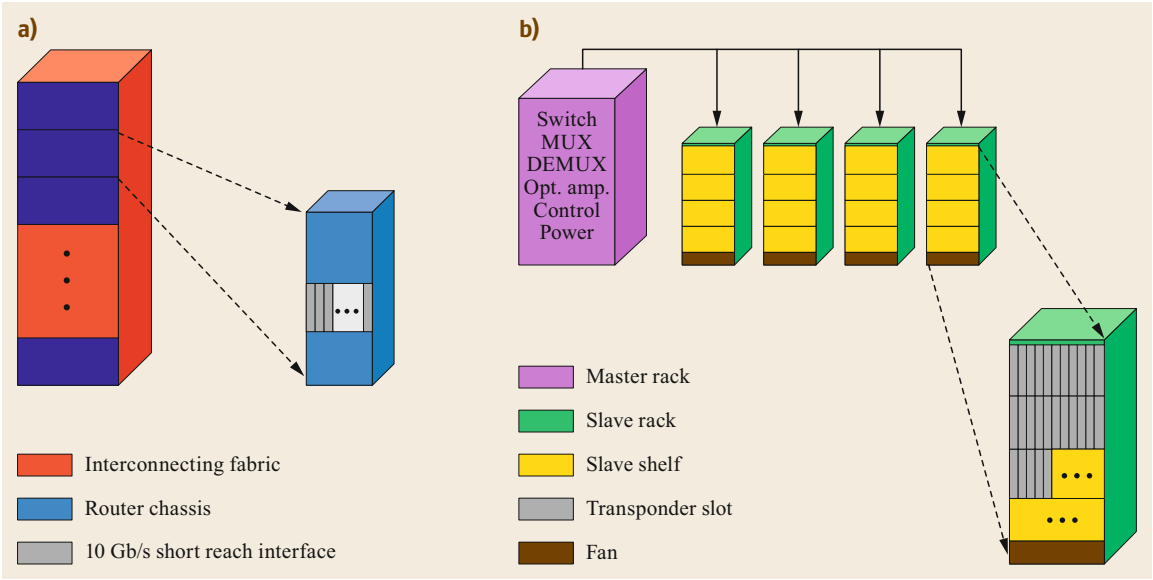


Fig. 19.11 **(a)** Rack configuration for a router and **(b)** rack configuration for an optical system (after [19.65])

amples of router and optical system rack implementations [19.65]. In most systems today, the power of the hardware is largely independent of the traffic that is being handled. Thus, even for no traffic flow, the hardware delivers the same or slightly lower power compared with full traffic load. This can be improved through the use of traffic-proportional designs, including frequency and voltage scaling in the CMOS processors and sleep mode operation (see below under power management). Nevertheless, in this relatively constant power draw scenario, the power or electricity consumed will depend on the number of racks and line cards powered up. Thus, the power-versus-capacity curve will be a stepwise increasing function as new racks or line cards are deployed in order to handle more traffic. In this light, energy efficiency is directly proportional to resource efficiency in terms of number of transceiver line cards, amplifier line cards, and other deployed equipment required to handle a given amount of traffic and/or number of nodes. Note that network operators will upgrade their node hardware before reaching full capacity with respect to the average traffic load. This strategy allows for traffic variations over time, gives some lead time to allow for the deployment of the new equipment, and finally avoids excessive network congestion that can reduce performance. For optical networks in the core, traffic utilization is in the range of 10–50% [19.7]. This applies to the optical transceivers and the subwavelength switching hardware (i.e., OTN, Ethernet, IP). One-plus-one protection will use double the capacity for a given traffic load and will require wavelength routing on disjoint paths. The optical node hardware will be upgraded when the number of wavelengths on a link is above 50% of the maximum—provisioning a new fiber to allow for additional growth.

The common equipment in an optical system includes cooling and electric power delivery. These are often major components of the overall energy efficiency—as much as 50%—which makes these practical considerations a target of many network operators [19.66].

Traffic-Dependent Equipment Energy Use

The energy efficiency of individual components or entire subsystems or nodes can be modeled in terms of the fixed energy use and the traffic (i.e., load)-dependent energy use. Figure 19.12 shows an example of the traffic load-dependent power for electronic and optical switching nodes and aggregate nodes of different size [19.67], indicating current typical commercial models and state-of-the-art best models. The particular dependence will vary based on how the traffic load proportionality is implemented—turning on and off subcomponents, changing the clock speed or electronic drive voltage (clock and voltage scaling), or other techniques.

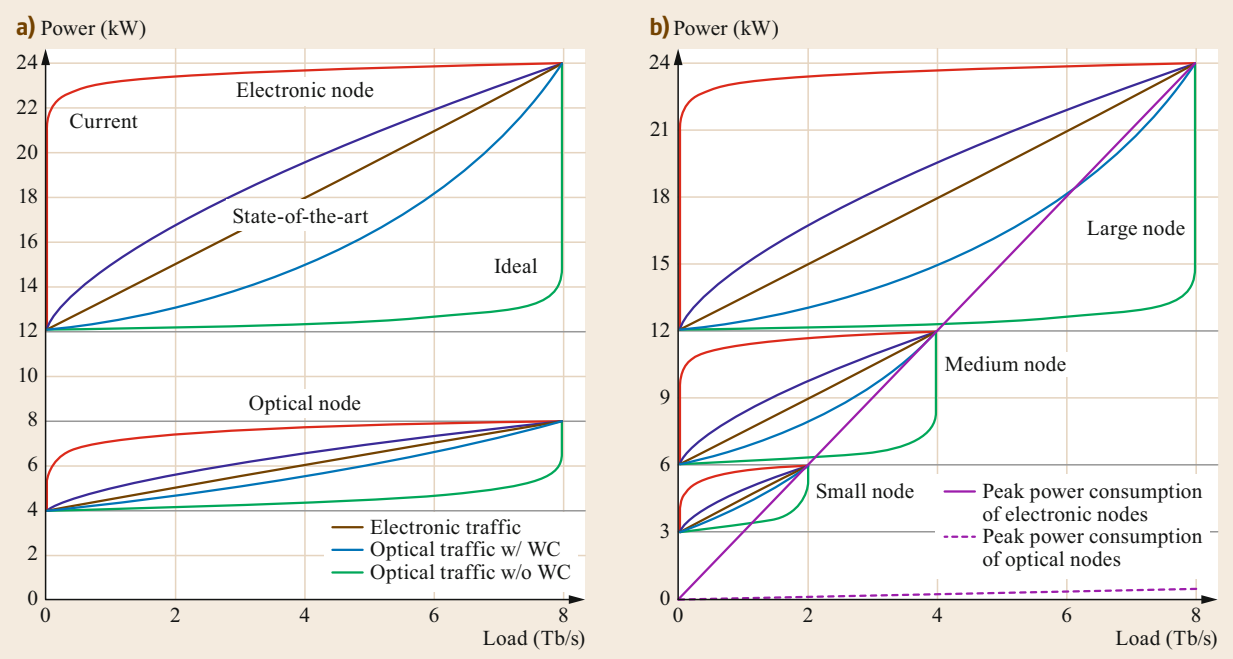


Fig. 19.12a,b Node power dependence on load including fixed and load-dependent operation; **(a)** electronic and optical node types and **(b)** different node sizes [19.67]

Power Management

Optical network models can include a range of different power management strategies. Power management can either take advantage of traffic-dependent equipment power as described above, or put equipment to sleep when not processing traffic. Whereas a core network optical switch is likely to be carrying traffic at all times, subsystems within the switch might be turned on and off. As optical networks are used closer to the edge of the network, where less aggregation occurs, sleep modes will have greater benefit. An energy-efficient Ethernet standard was developed for sleep mode operation in Ethernet switches [19.69]. As power management such as this finds greater use in the upper-network-layer equipment such as Ethernet switches, there will be greater motivation for power management in optical networks. Note that energy efficiency benefits of power management will be limited by additional energy associated with the control hardware and software and the energy use during the transitions when traffic is not transmitted. Transition times, therefore, can affect the efficiency of power management as well as the system performance. For example, if during a shutdown transition new traffic arrives requiring the equipment to turn back on [19.70], the latency for this new traffic may increase. Thermal management or thermal-dependent components can lead to long transition times.

Assuming that the individual technologies and systems are able to efficiently perform power management, there are other factors that can affect the potential energy efficiency gains. The electrical power delivery system and supply chain may not run efficiently over the range of operating conditions of the equipment [19.55]. The traffic must also be sufficiently time-varying that energy savings can be realized [19.13, 71]. Figure 19.13 shows time-dependent network power behavior based on traffic models and the corresponding energy savings based on the minimum and maximum power variations over time. There is also the difficulty of working in a multidomain, multioperator environment and coordinating these power management actions [19.72].

Optical transceivers and amplifiers have been considered for power management in optical networks. However, neither of these technologies makes use of power management in commercial systems, and significant system performance challenges must be overcome in order for these to find application. Turning optical channels on and off today is a time-consuming process that can take minutes or longer [19.73]. One strategy to contend with this is to leave the laser on at all times and just power-manage the electronics in the transceivers [19.74]. This avoids the complexity of setting up the optical channel, and given that most of the power of the line card is associated with the electronics, this provides most of the energy benefit. Furthermore, tunable lasers are often thermally stabilized, which can further increase the transition times. A number of studies have looked at the potential benefits of power management in optical systems [19.74–76] including transceivers both with and without transition delays. Optical amplifiers usually involve thermal management with similar transmission engineering complications in their control as the transceivers.

Power management has also been considered for use with optical protection switching. In the case of $1 + 1$ protection, the protection path channel is always active and using energy along with the working path. These protection channels can be placed into sleep mode instead [19.77]. The impact of protection on the energy efficiency of optical networks with dynamic

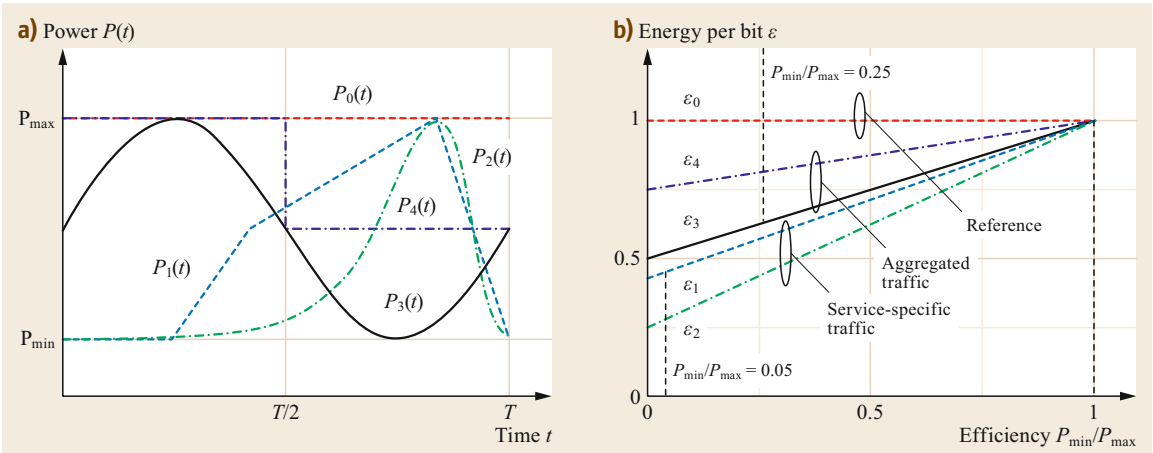


Fig. 19.13 (a) Time-dependent power variations based on time-dependent traffic for service-specific traffic and aggregated (diurnal) traffic, and (b) the corresponding relative energy efficiency improvement based on the maximum and minimum power of the network [19.68] (© IEEE 2010)

sleep mode capabilities was also studied for shared protection using a concept of differentiated reliability [19.78].

One of the difficulties with sleep mode operation is that it can involve power cycling of the equipment, which can lead to increased equipment failures, particularly when the thermal state of the components changes in the process. Models for the reduced lifetime of optical components have been developed [19.79]. RWA algorithms can be designed to take this into account. In particular, components often have a maximum allowable failure rate. The RWA algorithms that aggregate light paths onto common links so that unused links can be put to sleep, can be adapted so that power cycling is less frequent, reducing the failure rate. Note that aggregating more traffic onto particular links can also reduce the component lifetime, because the links are running at a heavy load, although this effect was found to be weaker than the power cycling for EDFAs. Nevertheless, maintaining component reliability through less aggressive energy-aware light path routing results in a reduction in energy efficiency gains for the network [19.80, 81].

In other studies related to power management, the general theory of network optimization has been studied for the case of network elements that scale their energy efficiency by varying the transmission speed [19.82] and as a multicommodity flow problem for core and edge networks [19.83].

Key points:

- Network elements can be broadly characterized using models for common node equipment, traffic- or load-dependent equipment, and equipment power management.
- Accounting for common node equipment leads to resource-dependent energy use, with stepwise increasing power with traffic load.
- Typical commercial equipment is characterized by a large fixed power requirement and a small traffic load dependence, although many other dependencies are possible and are being investigated.
- The benefit of power management such as sleep mode operation will depend on the traffic variability and other constraints on power control, including mode transition times.
- Power management may lead to higher component failure rates due to thermal cycling and can be minimized through cycling-aware control design.

19.3.3 Multilayer Network Architectures

Commercial networks use a layered network architecture with IP packet switching (L3, layer 3) and circuit switching (L2, layer 2), both using electronic switches, and a ROADM-based optical network (L1, layer 1) (Fig. 19.14). Service data flows are passed between different network domains (i.e., access network, metro network, long-haul network, and different operator networks) at central offices (CO) and Internet exchange points (IXP). These transactions take place in layer 3 IP packet switches, although increasing layer 2 IXP capabilities is being investigated. Layer 2 includes Ethernet, SONET, and OTN circuit switching technologies. Note however that most optical systems today include OTN framing and switching capabilities. Also, many L2 and L3 switches today can accept DWDM optical network line-side transceivers and therefore connect directly to the optical switching hardware. When layer 2 switching is not included, the architecture is often referred to as *IP over WDM*. The layer 2 and 3 equipment at a node can be *bypassed* by using optical switches in the ROADM nodes to *express* the signal through without add/drop, shown by the dashed red line in Fig. 19.14. The extent to which equipment in different layers is used will have a large impact on the overall energy efficiency of the network, as will be considered in detail in Sect. 19.4.

The relative contributions of the traffic load-independent and traffic load-dependent energy will af-

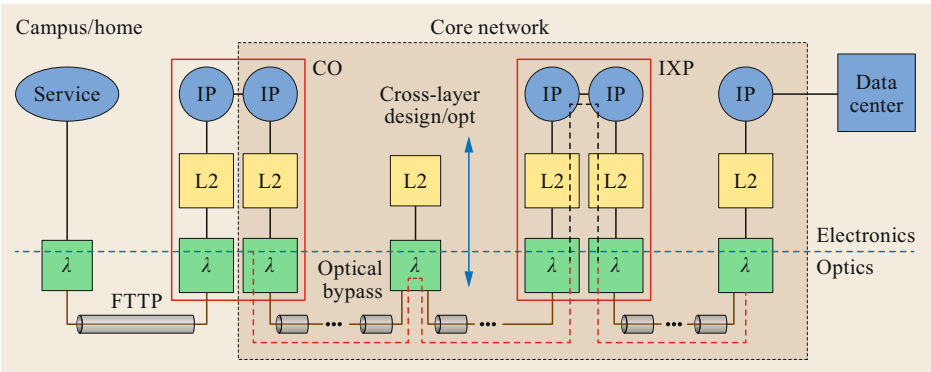


Fig. 19.14 Multilayer architecture with optical bypass

fect the efficiency of the network architecture. If the load-independent energy use is zero, then there is no energy cost to using more equipment, some of which may be lightly utilized. Idle equipment in this case does not lead to inefficiency. The traffic-dependent energy function (see Fig. 19.12) and the relative energy use of different network elements, therefore, determine the best architecture for energy-efficient operation. For example, if the nodes use more energy than the links and the load-independent power is small, then a flat, full mesh network that minimizes the transmission distances and reduces node hops may be preferred. Conversely, high load-independent energy use will lead to architectures that minimize the amount of equipment in the network, preferring more centralized or hierarchical architectures [19.84]. These considerations include not only the optical equipment, but also the electronic switching and routing equipment, which is often characterized by large traffic load-independent power.

Subwavelength switching architectures generally involve setting up overlapping wavelength routes in a network and then switching the signals within these routes to achieve time-division multiplexing (TDM). A common approach is to use wavelength-based broadcast and select, such that each node is set to receive a certain wavelength. By selecting the appropriate wavelength, data can be sent to that node. Other variants include different medium access control methods and wavelength switching strategies, including slotted or TDM switching [19.85, 86]. The ability to perform real-time wavelength switching or subwavelength switching has been shown to be energy-efficient for metro applications [19.87]. Flexibility in the node equipment provides further potential for energy savings. The architecture-on-demand concept allows for maximum flexibility in the node hardware by switching in components only when needed [19.88].

Key points:

- The flow of traffic to higher-layer equipment can lead to larger energy use.
- Depending on the relative power of the equipment in each layer and the traffic load dependence, the architecture and topology of the network can be configured for best energy efficiency.
- Architectures that employ greater use of optical switching, including time-division multiplexing, have been shown to improve network energy efficiency.

19.3.4 Network Traffic Models

At the system and network levels, the traffic model becomes a critical factor in determining the energy efficiency. The time dependence of the traffic will determine the efficiency gains from power management [19.68], as shown above. Diurnal traffic variations are commonly considered for optical networks due to the relatively long time scale at which optical systems can be adapted [19.68], although various packet- and circuit-based patterns can be used as relevant. The size of traffic demands can be modeled with a Pareto or heavy-tail distribution as well as a Zipf function that is often used for content distribution networks [19.89]. The actual traffic distribution will depend on the service and aggregation, and may vary considerably [19.13].

Traffic growth over time is often considered in the analysis of network energy efficiency. The Cisco Visual Networking Index provides estimates of total traffic by year for different regions of the world, broken down according to the different types of networks, including fixed and wired networks, consumer Internet, and business and managed IP [19.90]. This study uses numbers of users, adoption rates, minutes of use, and bit rates to arrive at traffic volume estimates. Other studies actively monitor traffic on the network and derive estimates from these measurements, often cross-referenced with other sources [19.91, 92].

The volume of traffic generally determines the number of transceivers or wavelengths in a network. It can also determine the number of fibers on a link, if the capacity of a single fiber is exceeded, which impacts the effective node degree and amount of link hardware (i.e., amplifiers). Additional components will be needed depending on various performance or reliability requirements that may be introduced. Furthermore, in an optical network, a single wavelength can occupy multiple links, blocking that wavelength for use by other signals. This wavelength blocking can result in spectrum fragmentation and lower utilization of the fiber spectrum. To account for wavelength blocking, wavelengths must be routed in the network based on a traffic matrix. In practical situations, the traffic matrix is a set of subwavelength demands that can be groomed onto different wavelengths, resulting in a multilayer traffic optimization problem to determine the energy efficiency (see Sect. 19.4.1).

19.3.5 Coding and Modulation

Coding and modulation are two of the most significant transceiver functions in terms of energy use. A wide range of modulation formats can be used in optical systems, each of which has both an implementation cost and a system impact in terms of spectral efficiency and transmission performance. A modulation format that allows for a signal to be transmitted farther will save energy by reducing regeneration. Channel coding plays

a role in this as well. Forward error correction coding enables more data to be transmitted over longer distances and through more nodes, but requires the additional energy of the FEC processing (Fig. 19.5). There is an important trade-off between using parallelism with simple coding and modulation and using fewer channels with more spectrally efficient and higher-performing transmission [19.94]. There is also a trade-off with source coding or data compression. Source coding reduces the amount of data that is transmitted in an optical network but requires more energy to compress the data [19.95].

Energy-efficient coding and modulation for optical systems has been evaluated in terms of comparing well-known methods as well as through the introduction of new, more energy-efficient methods. For coding and modulation, both optical power efficiency and overall energy efficiency have been studied. Optical power efficiency tends to be more relevant for extreme low-power or direct modulation applications for which the optical signal power can be an important factor. In most communication systems, the processing associated with the modulation and coding tends to use far more energy than the optical power itself. Therefore, energy-efficient modulation and coding is more concerned with the amount of processing or the power of the opto-electronics (e.g., modulators and modulator drivers) than with the optical signal power. Error correction codes can be designed to minimize the amount of processing—to provide the maximum coding gain for the minimum processing energy [19.96]. Note however that the method of implementing the coding in the processor and the choice of electronics can also have a significant impact, as described in Sect. 19.2 above.

Communication systems use M-bit binary coding schemes with 2^M constellation points or states represented in I-Q quadrature space, and modulation on two orthogonal optical polarizations when coherent receivers are used. These modulation formats can be made more power-efficient by jointly optimizing the two quadratures and two polarizations [19.97]. Pulse-position modulation is also known to be highly power-efficient and can be used with advanced modulation formats for greater spectral efficiency [19.98]. Subcarrier modulation in intensity-modulated direct detection systems was shown to achieve both optical and electrical power efficiency [19.99].

The spectral efficiency and energy efficiency of coding can be improved by going to multinary signaling with p^M constellation points [19.93], an example of which is shown in Fig. 19.15a. The nonuniform clustering of constellation points near the origin contributes to power efficiency. Energy-efficient coded modulation was also realized using statistical physics-inspired coding employing an entropy function for the bit constellation spacing [19.100].

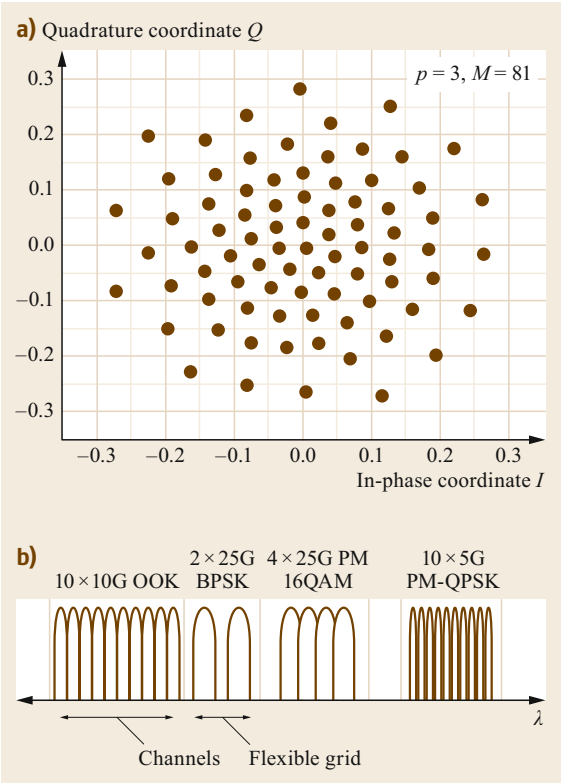


Fig. 19.15 **(a)** Power- and energy-efficient multinary modulation diagram [19.93]; **(b)** flexible-grid spectrum with mixed line rates and modulation formats, and multicarrier and banded channels

Mixed-Rate and Elastic Modulation

The choice of modulation data rate can also improve the energy efficiency of optical systems. The use of a high-data-rate transceiver to carry a small amount of traffic can lead to inefficient network operation. To realize the full benefit of optical transmission efficiency, the transceiver data rate should be matched to the traffic volume. Optical systems can use a single-line-rate transceiver, mixed- or multiple-line-rate transceivers, or rate-adaptive transceivers. The energy efficiency of optical networks has been studied taking into account the line rate in the optical routing and wavelength assignment [19.101–106]—in this case the traffic dependence is expressed through the network traffic matrix. The energy efficiency of the interfaces will also be affected by the time-dependent traffic offered to the interface or a set of interfaces. Rate adaptation is in fact a form of power management. Fixed- and adaptive-line-rate transceivers have been compared for energy efficiency

in handling time-dependent traffic including the effects of sleep mode power management [19.70]. The effects of wavelength tear-down and setup time on energy efficiency and delay in a transparent link was also examined for this case.

If the rate adaptation is carried over to the channel bandwidths in the network, this is referred to as elastic bandwidth with the use of a flexible grid (flex-grid) as opposed to a fixed grid (Fig. 19.15b). Traditionally, wavelength-division multiplexing has been used with a 50 GHz-bandwidth grid DWDM (*dense* WDM), 100 GHz grid, or larger grid CWDM (*coarse* WDM). With flex-grid, the signals are assigned spectral bands, and RWA becomes a process of routing and spectrum assignment (RSA). Flex-grid wavelength switching is achieved by using a fine-granularity switching band (e.g., 6.25 or 12 GHz) that can be combined through neighboring switch elements to form wider bands. The spectral efficiency improvement from elastic-bandwidth networks has also been studied with respect to energy efficiency [19.107, 108] including protection schemes [19.109]. In a comparison of orthogonal frequency-division multiplexing (OFDM) and mixed-line-rate modulation using bandwidth-variable transceivers in an elastic-bandwidth system, the OFDM modulation was found to be more energy-efficient [19.110].

Key points:

- Energy-efficient network coding and modulation saves energy by reducing the digital processing and/or modulation electronics.
- Optical power-efficient coding and modulation is used for very low-power applications and includes novel modulation constellations and pulse-position modulation.
- Flexible spectrum methods and adaptive modulation can lead to more efficient network utilization and reduced optical regeneration for energy savings.

19.4 Energy-Efficient Network Design

Network design for energy efficiency involves different trade-offs, which due to the relative energy benefit of optical transmission compared with higher-layer network technologies may involve trade-offs across multiple network layers. Many design features including reliability, flexibility, resilience, and security come at a cost in terms of energy use or efficiency. Spectral efficiency, capacity, and latency (due to processing) are exceptions that tend to benefit with greater energy efficiency, although there can also be exceptions to this.

19.4.1 Multilayer Traffic Engineering

The energy efficiency of optical networks can be exploited in order to achieve a more energy-efficient overall data network including the higher layers of the network. If data is more efficiently packed into wavelengths, then the data will require less processing in the higher layers, thus improving the energy efficiency. As shown in Fig. 19.8, the energy per bit of router processing can be five times that of the energy per bit of the optical switches. This difference is even larger if the signal is not regenerated in the optical node and instead passes through transparently (i.e., optical bypass). Multilayer traffic optimization, including both grooming and resource allocation, is usually considered from a cost or network utilization perspective, so energy adds a new dimension to this well-known problem space.

Optical Bypass

Optical bypass, in which the optical signal is passed transparently through a node, can be optimized based on the power of the higher-layer equipment or the energy use of layers 1–3. In this way, the virtual topology planning in the IP layer can be optimized for minimum energy use through optical bypass, including route selection for grooming traffic from multiple nodes [19.111–113]. Virtual topology here refers to the topology of the connections between routers or other higher-layer network elements, which may be different from the actual physical fiber connection topology. Thus, this virtual topology needs to be mapped to the physical topology or fiber connections, and that choice of routing is what can be optimized for energy efficiency. Energy-aware routing and wavelength assignment together with energy-aware multilayer traffic engineering, taking into account the traffic proportionality of the equipment at all layers, can show significant improvements over network management that is not energy-aware [19.67, 114].

For the case of dynamic optical operation, in which the wavelengths are rerouted or put to sleep over time for energy efficiency, the power management or sleep mode strategy of IP router ports can be optimized together with the optical route reconfiguration for enhanced energy savings [19.115]. In this case, the traffic proportionality of the higher-layer equipment together with the traffic patterns served will have a large im-

pact on the potential energy savings [19.68, 116]. With heavy traffic, few interfaces can be put to sleep, and therefore there is little benefit. For lower levels of traffic, when some interfaces can be put to sleep, if the idle or sleep power of the interface is high, then again there is little benefit. Such optical circuit switching can actually lead to less efficient networks if either the optical or IP resources end up poorly utilized. Hybrid switching refers to the use of optical circuit switching together with packet switching, in which a data flow size threshold is used to separate small *mice* flows from the large *elephant* flows. The ability to separate elephant flows has important consequences for the energy efficiency and resulting latency of this hybrid switching approach [19.117]. The impact of the protection strategy has also been examined for energy-efficient multilayer traffic engineering [19.77, 109].

Using commercial architectures and detailed equipment energy models, mixed-line-rate and elastic transceivers were multilayer-optimized for energy efficiency in an IP-over-WDM network [19.65, 118]. The design of energy efficiency-optimized IP-over-WDM networks was also studied for single line rates and mixed line rates (10−100 Gb/s) for networks that allow single-hop, few-hop, or full end-to-end transmission of optical signals, often referred to as opaque, translucent, and fully transparent networks, respectively [19.119]. Elastic optical networking was also used for energy-efficient many-cast routing in cloud computing applications [19.120]. Other studies have considered issues such as the upgrade strategy for best energy efficiency using mixed line rates in a network [19.121].

Renewable Energy

Multilayer optimization can be used to minimize the use of nonrenewable energy by organizing network and computing resources according to the cost and availability of renewable energy sources. This applies to the virtual network topology of the higher-network-layer equipment as well as the computational resources in data centers. Optimization can be done in network planning [19.67] or in real-time dynamic network control. Real-time network control based on the availability of renewable energy sources such as solar and wind is referred to as *follow the sun, follow the wind* [19.32, 122, 123]. This approach relies on either high-capacity optical connections or the ability to modify the optical routes in real time, which is not available in systems today. For this reason, field trials have utilized layer 2 switching for this purpose [19.32, 123]. The results shown in Fig. 19.16 take into account both time-varying traffic demand and time-varying renewable energy sources (solar variations shown in Fig. 19.16a), accounting for different traffic-proportional energy use models in the equipment. In other work, stochastic optimization of service provisioning is used to minimize the use of nonrenewable energy in data centers interconnected through an optical network [19.124].

Optimizing a network to make maximum use of available renewable energy can lead to frequent wavelength reconfiguration and high resource utilization as

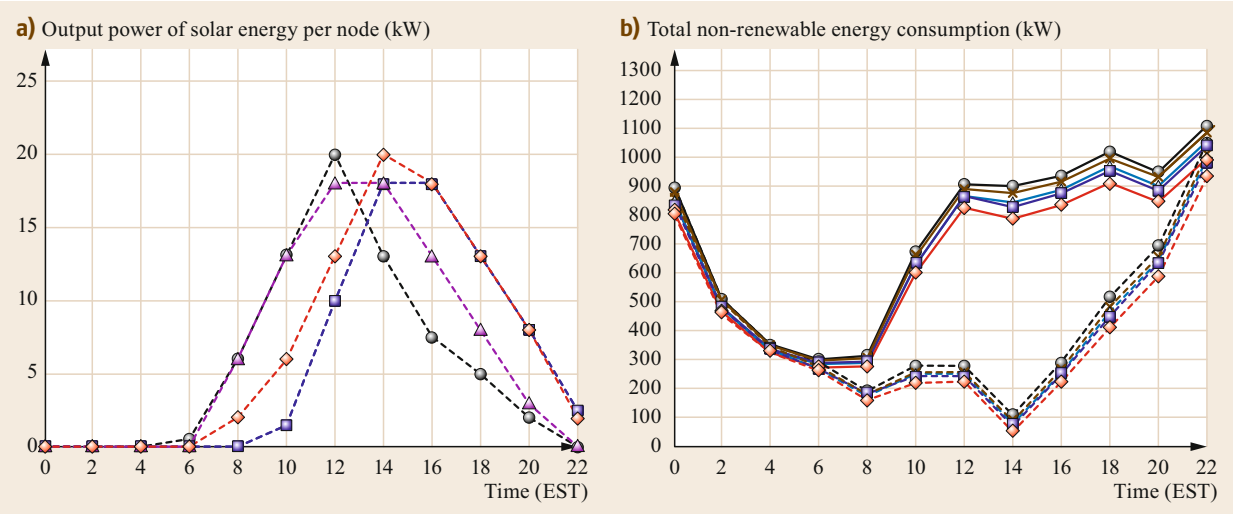


Fig. 19.16a,b Minimization of nonrenewable energy use following diurnal traffic and renewable energy profiles on different numbers of nodes [19.122] (courtesy of J. Elmirghani © IEEE 2010). **(a)** Solar energy at four representative nodes as a function of time; **(b)** additional total nonrenewable energy minimized by optimizing optical bypass routes with 80 KW (*solid*) and 20 kW (*dashed*) total solar power and considering different load-dependent energy models

signals are routed around to where the renewable energy is available. Introducing energy storage such as batteries was shown to decouple the network planning and traffic engineering from the renewable energy availability, improving the resource utilization throughout the network [19.126]. The impact of time-of-use pricing for renewable energy with storage on the operational expense for IP-over-WDM networks was also examined for different energy management policies [19.127]

19.4.2 Nonlinear Optics and Optical Processing

Nonlinear optics tends to work against energy efficiency, as nonlinear phenomena are nonlinear in optical intensity and therefore are only observed as the optical power is increased. The use of photonic integration and highly nonlinear materials can improve the situation both by increasing the optical intensity and medium response in the interaction region, and by reducing the impact of coupling losses between discrete devices. Nevertheless, as mentioned in Sect. 19.1.2, electronic devices are both fundamentally and practically more energy-efficient than optical devices for computational operations.

At very high speeds, the electronic devices become less efficient and often need to exploit parallelism, which still reduces energy efficiency although more slowly. Terminating an optical signal in order to perform electronic processing functions can also be costly from an efficiency perspective. As shown in Fig. 19.5, minimum power levels for a 100 Gb/s transceiver are approaching 100 W, and two transceivers would be needed in addition to the high-speed processing electronics itself. Since high-power optical sources and amplifiers needed for nonlinear optical processing often use only 1–10 W, there is a potential opportunity to achieve energy savings through the use of nonlinear optics-based signal processing. Figure 19.17a shows the relative efficiency of different nonlinear optical processing techniques compared with CMOS electronics [19.125, 128]. In Fig. 19.17b, the energy per bit as a function of the number of operations per bit is shown for both CMOS electronics and the best-case energy-efficient nonlinear optical techniques (see reference for full description of the different nonlinear optical techniques), considering different data rates for the optical signals. Also shown here is the energy associated with terminating the optical signals (O/E/O only), which is a lower bound on the opto-electronic approach, and the power associated with (de-)multiplexing the high-speed optical signals down to the lower-speed electronic processing clock speeds (O/E/O+(DE)MUX). The optical curves consider channel bandwidths of 100 GHz or 1 THz. Nonlinear optics becomes efficient for a small number of operations per bit N_{op}.

One example of processing on optical signals that involves a small number of operations is the opti-

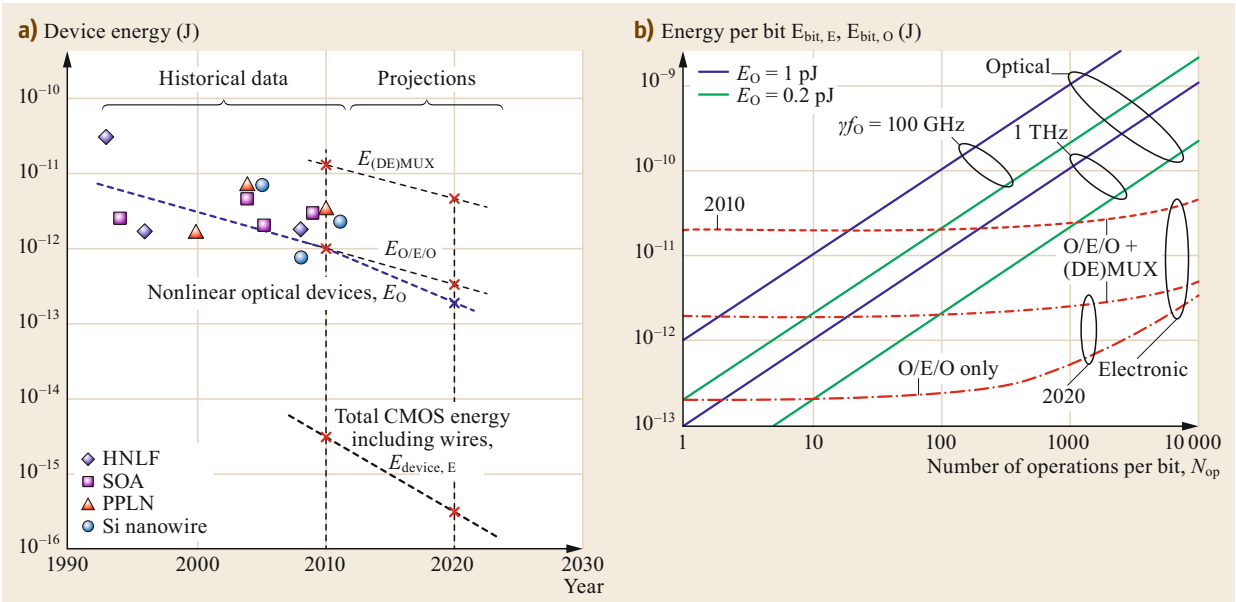


Fig. 19.17 **(a)** Energy efficiency of nonlinear optical switching devices and electronic switching and processing devices (incl. CMOS); **(b)** the energy per bit for optical and electronic processing as a function of the number of processing operations per bit [19.125] (© IEEE 2011)

cal grooming of multiple lower-data-rate signals into a single higher-data-rate signal. For QAM signals, this can be carried out through a series of nonlinear mixing operations that use power either due to a large pump signal or due to temperature stabilization. Nonlinear optical networking techniques were found to be energy-efficient at high baud rates and when integrated photonic solutions were used [19.129]. Optical time-division multiplexing has also been studied in terms of the energy efficiency of nonlinear optical techniques. Among three different techniques used for time multiplexing of on–off keyed modulation, the nonlinear optics-based methods were not able to achieve the energy efficiency of conventional WDM techniques [19.130]. Wavelength conversion is perhaps the simplest nonlinear optical networking technique, and was shown to provide energy efficiency benefits for systems with $1+1$ protection at 10 and 40 Gb/s with on–off keying modulation [19.131].

19.4.3 Cross-Optimized Computer and Optical Networks

Computer networks refers to the computers in a data center, office, or home and their communication in order to perform various computing tasks. While multilayer network optimization involves the joint optimization of the networks and their respective equipment operating at the different network layers, computer and optical network optimization refers to the trade-off between the computer processing functions and the optical network operations. Often multiple network layers in between are also involved in the analysis. Examples of processing operations that have been studied in regard to optical networks from an energy efficiency perspective include cloud processing, content delivery networks, and video processing. In general, the high capacity and low latency of optical networks are utilized to move data between processing resources to achieve greater efficiency or access to renewable energy sources.

Cloud Processing

Cloud computing involves computing and storage virtualized in data centers and executed over the network. The network plays an important role both in providing access to the computing and storage resources and in providing data continuity within the cloud across multiple data centers. Depending on the frequency with which these resources are accessed and the amount of data in that access, the relative contributions of transport, storage, and computing to the total energy use will vary [19.132, 133]. Figure 19.18 shows the relative contributions in the case of private cloud storage as a service. A key question is whether it is more efficient to perform a computational task on a personal computer or in the cloud. The answer to this question depends on the frequency of downloads/transactions or frames per second involved in the remote computation. Cloud computing is more efficient for infrequently accessed tasks, but local computing becomes more efficient under conditions of heavy use. Relative energy efficiency contributions were examined for storage as a service, software as a service, and processing as a service [19.132], and the network optimization problem was addressed for time-varying traffic demands [19.134]. The minimum bounds were determined for energy use in cloud computing implemented over optical networks [19.135].

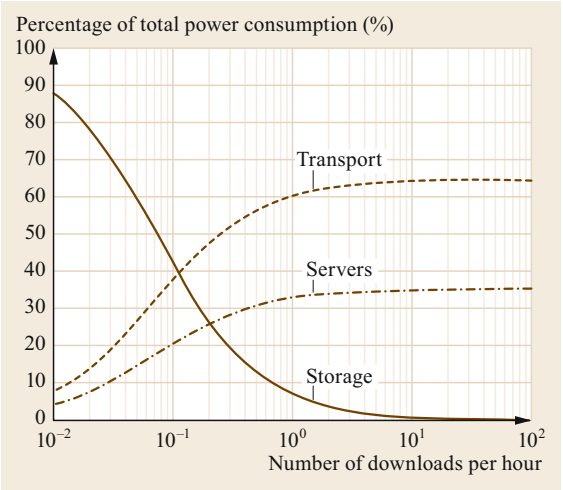


Fig. 19.18 Relative energy use for storage, transport, and servers in a private cloud storage as a service [19.132] (© IEEE 2010)

Other studies have looked at the energy efficiency of moving compute jobs between data centers for distributed cloud processing. Since the router energy use dominates over optical network energy use, the core network is often modeled without taking into account the optical network energy use contribution. However, an important role that optical networks can play is in reducing the latency and increasing the capacity between data centers. High-capacity connections between data centers have been shown to reduce job completion times and reduce energy use for distributed computing jobs [19.136].

Hybrid optical and electronic switching has been proposed for data center networks as a scalable means of introducing optical switching to data centers. The basic idea is to introduce optical switches in parallel with electronic switches (i.e., IP routers, Ethernet switches) to provide direct end-to-end connectivity for certain

traffic flows or operations. By placing large flows on the optical switching fabric, the electronic switching load is reduced and the overall efficiency is improved [19.139]. Hybrid switching was also proposed for core and integrated intra-data-center networks, and hybrid switching was shown to improve energy efficiency [19.140].

Content Delivery Networks

An important energy trade-off for content delivery networks involves the efficient replication and placement of content. More replication requires more storage and computing energy, but less energy in the delivery of the content through the network [19.133, 141, 142], similar to the case for cloud computing. In one study, the relative efficiency of over-the-top versus Internet service provider (ISP) content delivery was compared, and ISP-based video on demand was found to be more energy-efficient [19.143].

For content delivery, the critical parameter for energy-efficient optical bypass is for the content file/transaction size to be greater than $B_S > \gamma C_\lambda T^0 P_{wdm}/P_r$, where C_λ is the capacity of the optical wavelength channel, T^0 is the transfer time including optical switching and transmission tuning, P_{wdm} is the power or energy per bit of the optical network, P_r is the energy per bit of the routers or higher-layer equipment being bypassed, and γ is a proportionality constant that depends on the network. This metric was used to study the energy efficiency benefits of content delivery in a US national network including optical bypass and different content download frequency models, shown in Fig. 19.19 [19.137, 144]. When large files are more frequently downloaded, then the optical architecture will result in energy efficiency improving with normalized file size and in proportion to the relative power of the routing and optical equipment.

Information (or content)-centric networking is a new research approach to content delivery that can have consequences for the network energy efficiency. In this architecture, data flows are routed based on *named data* rather than an IP address. As content is transferred through the network, it is cached at each router. When a new request for the same content arrives at a router, it will first check its cache to see whether a copy already exists before forwarding the request to the data center. As long as the router continues to receive requests for the content, it will keep the content in its cache. In this way, popular content is stored near the edge of the network, reducing the traffic load and transmission energy, and enabling more efficient content distribution [19.145]. Similar to the case for content distribution networks, as the average number of hops to the content increases, the transmission energy per bit also increases. For information-centric networking, however, the cache size and energy decrease with the number of hops. In this case, the overall energy efficiency becomes a complex function of the cache size and efficiency and the algorithms for maintaining content in the caches [19.146], relative to the optical transmission energy.

Virtual- and augmented-reality video can involve multiview processing that can take place in the

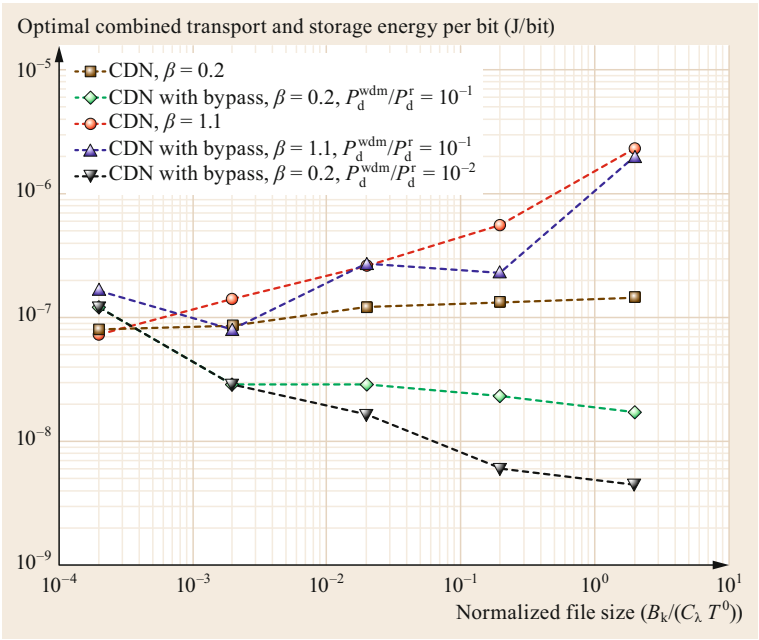


Fig. 19.19 Optimized transport and storage energy for content distribution with and without optical bypass and considering different relative energy efficiency of the optical and routing equipment, as a function of the file size normalized to the bandwidth delay product. Small β values correspond to more frequent download of large files [19.137] (© IEEE 2011)

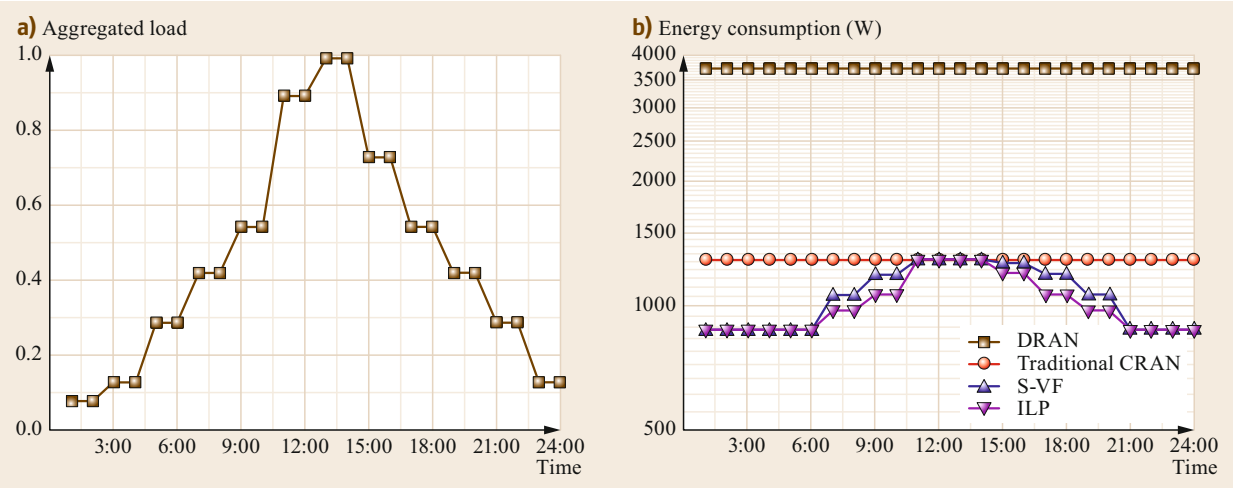


Fig. 19.20 (a) Time-dependent load variation for a converged wireless-optical network; (b) energy consumption using conventional, cloud, variable, and optimal radio access network architectures [19.138] (Courtesy of B. Mukherjee, © IEEE 2016)

user equipment, in the data center, or at locations within the network. The relative energy efficiency of these different cases has been considered in detail [19.147].

19.4.4 Mobile and Fixed Optical Convergence

Optical networks offer opportunities to improve the energy efficiency of wireless networks. The potential for energy savings in this regard is particularly important for future 5G wireless networks [19.148], which are targeting user access rates of 1 Gb/s, peak urban bandwidth densities of 10 (Tb/s)/km^2, and low latency (see ngmn.org). To achieve these targets, the number of access locations needs to be increased dramatically, given that the wireless access rates are strongly dependent on wireless transmission distance. To contend with this large growth in the number of base stations, cloud radio access networks (C-RAN) have been proposed, which pool baseband processing into data center and hoteling locations to gain the scalability of cloud processing. Figure 19.20b shows the relative power of C-RAN compared with traditional networks for a time-varying traffic load as shown in Fig. 19.20a. Heuristic algorithms using a dynamic virtual base station formation technique achieve traffic proportionality [19.138]. The energy use and performance trade-offs were determined for several converged optical and wireless 5G architectures [19.149–151].

19.5 Standards, Guidance, and Data Sources

Standards bodies and industry organizations have developed various recommendations and requirements for communication networks, including optical networks [19.152]. Many public sources of energy-related data are also available. Table 19.2 provides an overview of a number of these.

Several studies have considered the impact of such standards and guidance on the energy efficiency of communication networks. Service-level agreements (SLAs) are used to specify contractual obligations with respect to the network operation, such as the reliability or outage time per year. The use of energy efficiency metrics in SLAs has been proposed as well [19.53, 153]. Information on the amount of data handled by telecommunications equipment is often recorded and reported by each operating equipment unit. If similar data were provided for the electricity consumed while processing or transmitting this data, then this information could be used to form *green* SLAs based on the energy per bit or energy efficiency of the equipment. A network operator could either ensure that their entire network meets the green SLA requirement or provide separate equipment at each node that meets the efficiency criteria for this equipment based on the corresponding SLA requirement. Energy Star rating systems can be used to simply report the energy efficiency, and this could be done for individual applications or user data either by tracing the data route or by including energy effi-

Table 19.2 Organizations providing standards, guidance, and data relevant to energy-efficient networks

Organization/initiative	Description	Reference/website
Global e-Sustainability Initiative	Industry consortium providing guidance and recommendations	www.gesi.org
Greenpeace	Data and reports on sustainability	www.greenpeace.org
IEEE/IEEE Green ICT Initiative	Professional and standards organization, Green ICT Initiative provides touch point across all divisions	www.greenict.org
International Telecommunications Union (ITU-T)	Standards body	www.itu.int/en/ITU-T
Internet Engineering Task Force (IETF)	Standards body	www.ietf.org
Google	Public data search and visualization tool of human and environmental indicators	https://www.google.com/publicdata/directory
Greenhouse Gas Protocol	Greenhouse gas accounting standards and tools	www.ghgprotocol.org
World Resources Institute	Source of data on energy and sustainability	www.wri.org
World Wildlife Fund	Conservation organization providing reports and guidance on ICT and sustainability	https://wwf.panda.org

ciency information in a data header field that can then be updated at each node during transmission through the network. The coordination of the network energy monitoring with the end-user applications/equipment has been studied in a general context [19.154]. Progress in this area could help shape the behaviors of both the network operators and the network users toward more energy-efficient practices.

19.6 Conclusions

Energy plays an important role in optical systems and networks today and can be expected to increase in importance into the future. Technologies such as 5G mobile communications are expected to create greater demands for high speed and low latency on wired networks, further motivating the use of optics in energy-sensitive applications. With integrated photonics driving down the cost of optical equipment, everything is pointing to more widespread use of optical communications. As described in this chapter, the efficiency of optics in transporting data will contribute to the efficiency of the overall ICT sector and enable further use of ICT for smart, sustainable solutions across society. Much progress has been made in terms of understanding the main fundamental and applied technological issues around energy use in optical systems, and different modeling approaches have been developed in order to investigate energy use in a wide variety of optical networks. More research is needed as both the systems and components evolve and find new applications. As integrated photonics becomes intertwined with integrated electronics and computing, and optical and wireless technologies converge, new challenges will certainly arise with respect to the role of energy in optical communications.

Acknowledgments. The author acknowledges support from the National Science Foundation research grants #CNS-1737453, CNS-1650669, and EEC-0812072.

References

19.1 S. Teske, A. Zervos, C. Lins, J. Muth: *Energy [R]evolution: A Sustainable World Energy Outlook* (Greenpeace International, EREC, Amsterdam, Brussels 2010)

19.2 Global e-Sustainability Initiative: *#SMARTer2030: ICT Solutions for 21st Century Challenges* (GeSI, Brussels 2015)

19.3 W. Van Heddeghem, S. Lambert, B. Lannoo, D. Colle, M. Pickavet, P. Demeester: 2014 trends in worldwide ICT electricity consumption from 2007 to 2012, Comp. Commun. **50**(1), 64–76 (2014)

19.4 The Climate Group: *SMART 2020: Enabling the Low Carbon Economy in the Information Age* (GeSI, Brussels 2008)

19.5 J. Baliga, R. Ayre, K. Hinton, W. Sorin, R. Tucker: Energy consumption in optical IP networks, J. Lightwave Technol. **27**(13), 2391–2403 (2009)

19.6 C. Lange, D. Kosiankowski, R. Hulsermann, R. Weidmann, A. Gladisch: Energy footprint of telecommunication networks. In: *Eur. Conf. Opt. Commun.* (2010), Paper Mo1D2

19.7 D. Kilper, G. Atkinson, S.K. Korotky, S. Goyal, P. Vetter, D. Suvakovic, O. Blume: Power trends in communication networks, IEEE J. Sel. Top. Quantum Electron. **17**(2), 275 (2011)

19.8 B. Raghavan, J. Ma: The energy and emergy of the Internet. In: *Proc. Hotnets'11* (ACM, Cambridge 2011)

19.9 S. Phillips, S.L. Woodward, M.D. Feuer, P.D. Magill: A regression approach to infer electricity consumption of legacy telecom equipment, ACM SIGMETRICS Performance Evaluation Review (2010)

19.10 D. Kilper, K. Guan, K. Hinton, R. Ayre: Energy challenges in current and future optical transmission networks, Proc. IEEE **100**(5), 1168–1187 (2012)

19.11 Y. Zhang, P. Chowdhury, M. Tornatore, B. Mukherjee: Energy efficiency in Telecom optical networks, IEEE Commun. Surv. Tutor. **12**(4), 441–458 (2010)

19.12 D.C. Kilper, R. Tucker: Energy efficient telecommunications. In: *Optical Fiber Communications VI*, ed. by I.P. Kaminow, T. Li, A.E. Willner (Academic Press, Burlington 2013)

19.13 C. Lange, D. Kosiankowski, A. Betker, H. Simon, N. Bayer, D. von Hugo, H. Lehmann, A. Gladisch: Energy efficiency of load-adaptively operated telecommunication networks, J. Lightwave Technol. **32**, 571–590 (2014)

19.14 Y. Ye, F.J. Arribas, J. Elmirghani, F. Idzikowski, J.L. Vizcaino, P. Monti, F. Musumeci, A. Pattavina, W. Van Heddeghem: Energy-efficient resilient optical networks: Challenges and trade-offs, IEEE Commun. Mag. **53**(2), 144–150 (2015)

19.15 D.C. Kilper, H. Rastegarfar: Energy challenges in access and aggregation networks, Philos. Trans. R. Soc. A **374**, 20140435 (2016)

19.16 J. Romm: The Internet and the new energy economy. In: *Sustainability at the Speed of Light*, ed. by D. Pamlin (WWF, Sweden 2002) pp. 30–51

19.17 G. Cook, J.V. Horn: *How Dirty is Your Data?* (Greenpeace International, Amsterdam 2011)

19.18 H. Saunders: The Khazzoom-Brookes postulate and neoclassical growth, Energy J. **13**(4), 131–148 (1992)

19.19 S. Jevons: *The Coal Question—Can Britain Survive?* (MacMillan, London 1865), reprinted by MacMillan in 1906

19.20 J. Romm, A. Rosenfeld, S. Herrmann: *The Internet Economy and Global Warming* (CECS, GETF, Arlington 1999)

19.21 D.T. Neilson: Photonics for switching and routing, IEEE J. Sel. Top. Quantum Electron. **12**, 669–678 (2006)

19.22 E. Bonetto, L. Chiaraviglio, D. Cuda, G.A.G. Castillo, F. Neri: Optical technologies can improve the energy efficiency of networks. In: *Eur. Conf. Opt. Commun.* (2009), Paper 5.5.1

19.23 J. Orcutt, R. Ram, V. Stojanovic: CMOS photonics for high performance interconnects. In: *Optical Fiber Telecommunications VIA*, ed. by I. Kaminow, T. Li, A.E. Willner (Elsevier, Amsterdam 2013) pp. 419–460

19.24 D.A.B. Miller: Device requirements for optical interconnects to silicon chips, Proc. IEEE **97**, 1166–1185 (2009)

19.25 J.D. Meindl: Low power microelectronics: Retrospect and prospect, Proc. IEEE **83**, 619 (1995)

19.26 S. Lloyd: Ultimate physical limits to computation, Nature **406**, 1047–1054 (2000)

19.27 D.C. Kilper, D. Neilson, D. Stiliadis, D. Suvakovic, S. Goyal: Fundamental limits on energy use in optical networks. In: *Eur. Conf. Opt. Commun. (ECOC)* (2010), https://doi.org/10.1109/ECOC.2010.5621584

19.28 D.C. Kilper, S. Chandrasekhar, G. Kramer, X. Wei: Quantum limits for receiver sensitivity with coding. In: *Opt. Fiber Commun. Conf.* (2004), Paper WM1

19.29 C. Antonelli, A. Mecozzi, M. Shtaif, P.J. Winzer: Quantum limits on the energy consumption of optical transmission systems, J. Lightwave Technol. **32**, 1853–1860 (2014)

19.30 Y. Yamamoto, H. Haus: Preparation, measurement, and information capacity of optical quantum states, Rev. Mod. Phys. **58**, 1001–1020 (1986)

19.31 D.C. Kilper, K. Guan, J. Llorca, G. Atkinson, R.S. Tucker: Coding and capacity in efficient optical networks. In: *16th Opto-Electron. Commun. Conf.* (2011) pp. 32–33

19.32 S. Figuerola, M. Lemay, V. Reijs, M. Savoie, B. St. Arnaud: Converged optical network infrastructures in support of future internet and grid services using IaaS to reduce GHG emissions, J. Lightwave Technol. **27**, 1941–1946 (2009)

19.33 C.A. Chan, A.F. Gygax, E. Wong, C.A. Leckie, A. Nirmalathas, D.C. Kilper: Methodologies for assessing the use-phas power consumption and greenhouse gas emissions of telecommunications network services, Environ. Sci. Technol. **47**, 485–492 (2013)

19.34 O. Tamm, C. Hermsmeyer, A.M. Rush: Eco-sustainable system and network architectures for future transport networks, Bell Labs Tech. J. **14**, 311–328 (2010)

19.35 D. Kilper: Energy efficient networks. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC)* (2011), https://doi.org/10.1364/OFC.2011.OWI5

19.36 A. Danowitz, K. Kelley, J. Mao, J.P. Stevenson, M. Horowitz: CPU DB: Recording microprocessor history, Queue (2012), https://doi.org/10.1145/2181796.2181798

19.37 M. Horowitz, E. Alon, D. Patil, S. Naffziger, R. Kumar, K. Bernstein: Scaling, power, and the future of CMOS. In: *Electron Devices Meet.* (2005), https://doi.org/10.1109/IEDM.2005.1609253

19.38 A.V. Krishnamoorty, K.W. Goossen, W. Jan, X. Zheng, R. Ho, G. Li, R. Rozier, F. Liu, D. Patil, J. Lexau, H. Schwetman, D. Feng, M. Asghari, T. Pinguet, J.E. Cunningham: Progress in low-power switched optical interconnects, IEEE J. Spec. Top. Quantum Electron. **17**(2), 357–376 (2011)

19.39 E. Timurdogan, Z. Su, K. Settaluri, S. Lin, S. Moazeni, C. Sun, G. Leake, D.D. Coolbaugh, B.R. Moss, M. Moresco, V. Stojanović, M.R. Watts: An ultra-low power 3D integrated intra-chip silicon electronic photonic link. In: *Opt. Fiber Commun. Conf. (OFC)* (2015), https://doi.org/10.1364/OFC.2015.Th5B.8

19.40 B.S. Pillai, B. Sedighi, K. Guan, N.P. Anthapadmanabhan, W. Shieh, K.J. Hinton, R.S. Tucker: End-to-end energy modeling and analysis of long-haul coherent transmission systems, J. Lightwave Technol. **32**(18), 3093–3111 (2014)

19.41 S. Chandrasekhar, X. Liu: Terabit superchannels for high spectral efficiency transmission. In: *36th Eur. Conf. Opt. Commun. (ECOC)* (2010), https://doi.org/10.1109/ECOC.2010.5621580

19.42 W. Forysiak, D.S. Govan: Progress toward 100-G digital coherent pluggables using InP-based photonics, J. Lightwave Technol. **32**, 2925–2934 (2014)

19.43 Gazettabyte: Acacia looks to co-package its coherent PIC and DSP-ASIC, http://www.gazettabyte.com/home/2017/2/8/acacia-looks-to-co-package-its-coherent-pic-and-dsp-asic.html (2017)

19.44 B.S.G. Pillai, B. Sedighi, W. Shieh, R.S. Tucker: Chromatic dispersion compensation—An energy consumption perspective. In: *Opt. Fiber Commun. Conf.* (2012), https://doi.org/10.1364/OFC.2012.OM3A.8

19.45 B. Sedighi, J. Li, K.-L. Lee, S. Gambini, H. Chow, R.S. Tucker: Energy-efficient optical links: Optimal launch power, IEEE Photonics Technol. Lett. **25**, 1715–1718 (2013)

19.46 K.-L. Lee, B. Sedighi, R.S. Tucker, H. Chow, P. Vetter: Energy efficiency of optical transceivers in fiber access networks, J. Opt. Commun. Netw. **4**(9), A59–A68 (2012)

19.47 M.D. Feuer, D.C. Kilper, S. Woodward: ROADMs and their system applications. In: *Optical Fiber Communications VB*, ed. by I.P. Kaminow, T. Li, A.E. Willner (Academic Press, San Diego 2008)

19.48 D. Marom, P.D. Colbourne, A. D'Errico, N.K. Fontaine, Y. Ikuma, R. Proietti, L. Zong, J.M. Rivas-Moscoso, I. Tomkos: Survey of photonic switching architectures and technologies in support of spatially and spectrally flexible optical networking, J. Opt. Commun. Netw. **9**, 1–26 (2017)

19.49 T.J. Seok, N. Quack, S. Han, R.S. Muller, M.C. Wu: Highly scalable digital silicon photonic MEMS switches, J. Lightwave Technol. **34**, 365–371 (2016)

19.50 S. Gringeri, N. Bitar, T.J. Xia: Extending software defined network principles to include optical transport, IEEE Commun. Mag. **51**(3), 32–40 (2013)

19.51 R.S. Tucker: The role of optics and electronics in high-capacity routers, J. Lightwave Technol. **24**, 4655–4673 (2006)

19.52 S. Aleksic: Analysis of power consumption in future high-capacity network nodes, J. Opt. Commun. Netw. **1**, 245–258 (2009)

19.53 K. Kawamoto, J.G. Koomey, B. Nordman, R.E. Brown, M. Piette, M. Ting, A.K. Meier: Electricity used by office equipment and network equipment in the US, Energy **27**(3), 255–269 (2002)

19.54 R. Basmadjian, G. Lovasz, M. Beck, H. De Meer, X. Hesselbach-Serra, J.F. Botero, S. Klingert, M.P. Ortega, J.C. Lopez, A. Stam, R. Van Krevelen, M. Di Girolamo: A generic architecture for demand response: The ALL4Green approach. In: *Int. Conf. Cloud Green Comput.* (2013), https://doi.org/10.1109/CGC.2013.79

19.55 C. Lange, D. Kosiankowski: Energy efficiency improvements in telecommunication networks—Optimizing the power supply chain. In: *Int. Conf. Photonics Switch. (PS)* (2012), A3811492

19.56 R. Tucker: Green optical communications—Part 1: Energy limitations in transport, IEEE J. Sel. Top. Quantum Electron. **17**(2), 245–260 (2011)

19.57 R. Tucker: Green optical communications—Part 2: Energy limitations on networks, IEEE J. Sel. Top. Quantum Electron. **17**(2), 261–274 (2011)

19.58 G. Rizzelli, G. Maier, M. Quagliotti, M. Schiano, A. Pattavina: Asssessing the scalability of next-generation wavelength switched optical networks, J. Lightwave Technol. **32**(12), 2263–2270 (2014)

19.59 R.S. Tucker, R. Parthiban, J. Baliga, K. Hinton, R.W.A. Ayre, W.V. Sorin: Evolution of WDM optical IP networks: A cost and energy perspective, J. Lightwave Technol. **27**, 243–252 (2009)

19.60 I. Tomkos, S. Azodolmolky, J. Sole-Pareta, D. Careglio, E. Palkopoulou: A tutorial on the flexible optical networking paradigm: State-of-the-art, trends, and research challenges, Proc. IEEE **102**, 1317–1337 (2014)

19.61 C. Bastos-Filho, D. Araujo, E. Barboza, D. Chaves, J. Martins-Filho: Design of transparent optical networks considering physical impairments, capex and energy consumption. In: *13th Int. Conf. Transpar. Opt. Netw.* (2011), https://doi.org/10.1109/ICTON.2011.5971080

19.62 Y. Wu, L. Chiaraviglio, M. Mellia, F. Neri: Power-aware routing and wavelength assignment in optical networks. In: *35th Eur. Conf. Opt. Commun.* (2009), Paper P5-12

19.63 C. Cavdar, M. Ruiz, P. Monti, L. Velasco, L. Wosinska: Design of green optical networks with signal quality guarantee. In: *IEEE Int. Conf. Commun. (ICC)* (2012), https://doi.org/10.1109/ICC.2012.6364041

19.64 Z. Zhu, X. Chen, F. Ji, L. Zhang, F. Farahmand, J. Jue: Energy efficient translucent optical transport networks with mixed regenerator placement, J. Lightwave Technol. **30**(19), 3147–3156 (2012)

19.65 G. Rizzelli, A. Morea, M. Tornatore, O. Rival: Energy efficient traffic-aware design of on–off multilayer translucent optical networks, Comput. Netw. **56**(10), 2443–2445 (2012)

19.66 S. Roy: Energy logic: A road map to reducing energy consumption in telecommunications networks. In: *IEEE Int. Telecommun. Energy Conf.* (2008), https://doi.org/10.1109/INTLEC.2008.4664025

19.67 S. Ricciardi, F. Palmieri, U. Fiore, D. Careglio, G. Santos-Boada, J. Sole-Pareta: An energy-aware dynamic RWA framework for next-generation wavelength-routed networks, Comput. Netw. **56**, 2420–2442 (2012)

19.68 C. Lange, A. Gladisch: Energy efficiency limits of load adaptive networks. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC)* (2010), https://doi.org/10.1364/OFC.2010.OWY2

19.69 X. Pan, T. Ye, T.T. Lee, W. Hu: Power efficiency and delay tradeoff of 10GBase-T energy efficient ethernet protocol, IEEE/ACM Trans. Netw. **25**(5), 2773–2787 (2017)

19.70 K. Guan, D.C. Kilper, Y. Pan, O. Rival, A. Morea: Energy efficient file transfer over rate adaptive optical networks. In: *IEEE Online Conf. Green Commun. (GreenCom)* (2012), https://doi.org/10.1109/GreenCom.2012.6519613

19.71 C. Lange, D. Kosiankowski, A. Gladisch: Effect of load-proportional systems on the energy efficiency of fixed telecom operator networks. In: *40th Eur. Conf. Opt. Commun. (ECOC)* (2014), https://doi.org/10.1109/ECOC.2014.6963836

19.72 C. Lange, D. Kosiankowski, A. Gladisch: Realistic energy-saving potential of load-adaptive operation in conventional and platform-consolidated operator networks. In: *39th Eur. Conf. Opt. Commun. (ECOC)* (2013), https://doi.org/10.1049/cp.2013.1370

19.73 L.E. Nelson, G. Zhang, N. Padi, C. Skolnick, K. Benson, T. Kaylor, S. Iwamatsu, R. Inderst, F. Marques, D. Fonseca, M. Du, T. Downs, T. Scherer, C. Cole, Y. Zhou, P. Brooks, A. Schubert: SDN-Controlled 400GbE end-to-end service using a CFP8 client over a deployed, commercial flexible ROADM system. In: *Opt. Fiber Commun. Conf.* (2017), https://doi.org/10.1364/OFC.2017.Th5A.1

19.74 A. Morea, S. Spardo, O. Rival, J. Perello, F.A. Bujan, D. Verchere: Power management of optoelectronic interfaces for dynamic optical networks. In: *Eur. Conf. Opt. Commun.* (2011), https://doi.org/10.1364/ECOC.2011.We.8.K.3

19.75 Y. Nomura, H. Yonezu, D. Ishii, S. Okamoto, N. Yamanaka: Dynamic topology reconfiguration for energy efficient multi-layer network using extended GMPLS with link power control. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC)* (2012), https://doi.org/10.1364/NFOEC.2012.NTu2J.6

19.76 L. Chiaraviglio, D. Ciullo, M. Mellia, M. Meo: Modeling sleep modes gains with random graphs. In: *IEEE Conf. Comput. Commun. Worksh. (INFOCOM WKSHPS)* (2011), https://doi.org/10.1109/INFCOMW.2011.5928837

19.77 A. Muhammad, P. Monti, I. Cerutti, L. Wosinska, P. Castoldi, A. Tzanakaki: Energy-efficient WDM network planning with dedicated protection resources in sleep mode. In: *IEEE Glob. Telecommun. Conf.* (2010), https://doi.org/10.1109/GLOCOM.2010.5683205

19.78 A. Muhammad, P. Monti, I. Cerutti, L. Wosinska, P. Castoldi: Reliability differentiation in energy efficient optical networks with shared path protection. In: *IEEE Online Conf. Green Commun.* (2013), https://doi.org/10.1109/OnlineGreenCom.2013.6731030

19.79 L. Chiaraviglio, P. Wiatr, P. Monti, J. Chen, J. Lorincz, F. Idzikowski, M. Listanti, L. Wosinska: Is green networking beneficial in terms of device lifetime?, IEEE Commun. Mag. **53**(5), 232–240 (2015)

19.80 P. Wiatr, J. Chen, P. Monti, L. Wosinska: Energy efficiency and reliability tradeoff in optical core networks. In: *Opt. Fiber Commun. Conf. (OFC)* (2014), https://doi.org/10.1364/OFC.2014.Th4E.4

19.81 P. Wiatr, J. Chen, P. Monti, L. Wosinska: Energy efficiency versus reliability performance in optical backbone networks, J. Opt. Commun. Netw. **7**, A482–A491 (2015)

19.82 M. Andrews, A.F. Anta, L. Zhang, W. Zhao: Routing for energy minimization in the speed scaling model, IEEE/ACM Trans. Netw. **20**, 285–294 (2012)

19.83 L. Chiaraviglio, M. Mellia, F. Neri: Reducing power consumption in backbone networks. In: *IEEE Int. Conf. Commun.* (2009), https://doi.org/10.1109/ICC.2009.5199404

19.84 A. Ahmad, A. Bianco, E. Bonetto, D. Cuda, G. Gavilanes Castillo, F. Neri: Power-aware logical topology design heuristics in wavelength routing networks. In: *15th Int. Conf. Opt. Netw. Design Model.* (2011)

19.85 D. Chiaroni, G. Buforn Santamaria, C. Simonneau, S. Etienne, J.C. Antona, S. Bigo, J. Simsarian: Packet OADM for the next generation of ring networks, Bell Labs Tech. J. **14**(4), 265–283 (2010)

19.86 A. Bianco, T. Bonald, D. Cuda, R.-M. Indre: Cost, power consumption and performance evaluation of metro networks, J. Opt. Commun. Netw. **5**, 81–91 (2013)

19.87 E. Bonetto, E. Le Rouzic, L. Sadeghioon, P. Gavignet, B. Arzur, O. Renais: Facing the traffic explosion in metro transport networks with energy-sustainable architectures, Photonic Netw. Commun. **30**, 29–42 (2015)

19.88 M. Garrich, N. Amaya, G.S. Zervas, P. Giaccone, D. Simeonidou: Power consumption analysis of architecture on demand. In: *38th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2012), https://doi.org/10.1364/ECEOC.2012.P5.06

19.89 M. Cha, P. Rodriguez, J. Crowcroft, S. Moon, X. Amatriain: Watching television over an IP network. In: *Proc. 8th ACM SIGCOMM Conf. Internet Meas.* (2008), https://doi.org/10.1145/1452520.1452529

19.90 Cisco: *The Zettabyte Era: Trends and Analysis* (Cisco, San Jose 2017)

19.91 A.M. Odlyzko: Internet traffic growth: Sources and implications, Proc. SPIE (2003), https://doi.org/10.1117/12.512942

19.92 S.K. Korotky: Semi-empirical description and projections of Internet traffic trends using a hyperbolic compound annual growth rate, Bell Labs Tech. J. **18**, 5–21 (2013)

19.93 I.B. Djordjevic, T. Liu, T. Wang: Multinary-signaling-based coded modulation for ultrahigh-speed

optical transport, IEEE Photonics J. **7**(1), 7024098 (2015)
19.94 P.J. Winzer: Energy-efficient optical transport capacity scaling through spatial multiplexing, IEEE Photonics Technol. Lett. **23**, 851–853 (2011)
19.95 X. Dong, A. Lawey, T.E.H. El-Gorashi, J.M.H. Elmirghani: Energy-efficient core networks. In: *16th Conf. Opt. Netw. Design Model.* (2012), https://doi.org/10.1109/ONDM.2012.6210196
19.96 I.B. Djordjevic: On the energy-efficient multidimensional coded modulation for optical transport networks. In: *IEEE Online Conf. Green Commun.* (2014), https://doi.org/10.1109/OnlineGreenCom.2014.7114428
19.97 E. Agrell, M. Karlsson: Power-efficient modulation formats in coherent transmission systems, J. Lightwave Technol. **27**, 5115–5126 (2009)
19.98 M. Karlsson, E. Agrell: Generalized pulse-position modulation for optical power-efficient communication. In: *37th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2011), https://doi.org/10.1364/ECOC.2011.Tu.6.B.6
19.99 J. Karout, K. Szczerba, M. Karlsson: Designing power-efficient modulation formats for noncoherent optical systems. In: *IEEE Glob. Telecommun. Conf* (2011), https://doi.org/10.1109/GLOCOM.2011.6133546
19.100 I.B. Djordjevic, L. Xu, T. Wang: Statistical physics inspired energy-efficient coded-modulation for optical communications, Opt. Lett. **37**, 1340–1342 (2012)
19.101 O. Rival, A. Morea, J.-C. Antona: Optical network planning with rate-tunable NRZ transponders. In: *35th Eur. Conf. Opt. Commun. (ECOC)* (2009)
19.102 A. Nag, M. Tornatore, B. Mukherjee: Optical network design with mixed line rates and multiple modulation formats, J. Lightwave Technol. **28**(4), 466–475 (2010)
19.103 P. Chowdhury, M. Tornatore, B. Mukherjee: On the energy efficiency of mixed-line-rate networks. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC)* (2010), https://doi.org/10.1364/OFC.2010.OWY3
19.104 K. Christopdoulopoulos, K. Manousakis, E. Varvarigos: Adapting the transmision reach in mixed line rate WDM transport networks. In: *Int. Conf. Opt. Netw. Design Model.* (2011)
19.105 A. Morea, O. Rival, N. Brochier, E.L. Rouzic: Datarate adaptation for night-time energy saving in core networks, J. Lightwave Technol. **31**(5), 779–785 (2013)
19.106 A. Udalcovs, P. Monti, V. Bobrovs, R. Schatz, L. Wosinska, G. Ivanovs: Spectral and energy efficiency consideration in mixed-line rate WDM networks with signal quality guarantee. In: *15th Int. Conf. Transp. Opt. Netw.* (2013), https://doi.org/10.1109/ICTON.2013.6602845
19.107 A. Fallahpour, H. Beyranvand, S.A. Nezamalhosseini, J.A. Salehi: Energy efficient routing and spectrum assignment with regenerator placement in elastic optical networks, J. Lightwave Technol. **32**, 2019–2027 (2014)
19.108 P. Papanikolaou, P. Soumplis, K. Manousakis, G. Papadimitriou, G. Ellinas, K. Christodoulopoulos, E. Varvarigos: Minimizing energy and cost in fixed-grid and flex-grid networks, J. Opt. Commun. Netw. **7**(4), 337–351 (2015)
19.109 M. Ju, F. Zhou, S. Xiao, Z. Zhu: Power-efficient protection with directed p-cycles for asymmetric traffic in elastic optical networks, J. Lightwave Technol. **34**, 4053–4065 (2016)
19.110 A. Nag, T. Wang, B. Mukherjee: Robust design of spectrum-efficient green optical backbone networks, J. Lightwave Technol. **31**, 1138–1144 (2013)
19.111 G. Shen, R.S. Tucker: Energy-minimized design for IP over WDM networks, J. Opt. Commun. Netw. **1**, 176–186 (2009)
19.112 G.J. Eilenberger, S. Bunse, L. Dembeck, U. Gebhard, F. Ilchmann, W. Lautenschlaeger, J. Milbrandt: Energy-efficient transport for the future Internet, Bell Labs Tech. J. **15**, 149–170 (2010)
19.113 W. Van Heddeghem, M. De Groote, W. Vereecken, D. Colle, M. Pickavet, P. Demeester: Energy-efficiency in telecommunications networks: Link-by-link versus end-to-end grooming. In: *14th Conf. Opt. Netw. Design Model.* (2010), https://doi.org/10.1109/ONDM.2010.5431570
19.114 C. Lee, J.-K.K. Rhee: Traffic grooming for IP-over-WDM networks: Energy and delay perspectives, J. Opt. Commun. Netw. **6**(2), 96–103 (2014)
19.115 Y. Lui, G. Shen, W. Shao: Design for energy-efficient IP over WDM networks with joint light-path bypass and router-card sleeping strategies, J. Opt. Commun. Netw. **5**, 1122–1138 (2013)
19.116 B. Puype, D. Colle, M. Pickavet, P. Demeester: Energy efficient multilayer traffic engineering. In: *35th Eur. Conf. Opt. Commun. (ECOC)* (2009)
19.117 T. Nandagopal, K. Guan, D.C. Kilper: Energy efficiency and delay performance of data transfer using dynamic optical switching. In: *IEEE Int. Conf. Commun. (ICC)* (2012), https://doi.org/10.1109/ICC.2012.6364964
19.118 A. Klekamp, U. Gebhard: Performance of elastic and mixed line-rate scenarios for a real IP over DWDM network with more than 1000 nodes, J. Opt. Commun. Netw. **5**, A28–A36 (2013)
19.119 P. Chowdhury, M. Tornatore, A. Nag, E. Ip, T. Wang, B. Mukherjee: On the design of energy-efficient mixed-line-rate (MLR) optical networks, J. Lightwave Technol. **30**, 130–139 (2012)
19.120 A. Fallahpour, H. Beyranvand, J.A. Salehi: Energy-efficient manycast routing and spectrum assignment in elastic optical networks for cloud computing environment, J. Lightwave Technol. **33**, 4008–4018 (2015)
19.121 A. Nag, M. Tornatore, B. Mukherjee: Energy-efficient and cost-efficient capacity upgrade in lixed-line-rate optical networks, J. Opt. Commun. Netw. **4**, 1018–1025 (2012)
19.122 X. Dong, T. El-Gorashi, J. Elmirghani: IP over WDM networks employing renewable energy sources, J. Lightwave Technol. **29**(1), 3–14 (2010)
19.123 M. Lemay, K.-K. Nguyen, B. St. Arnaud, M. Cheriet: Toward a zero-carbon network: Converging cloud

computing and network virtualization, IEEE Internet Comput. **16**(6), 51–59 (2012)
19.124 M. Anastasopoulos, A. Tzanakaki, D. Simeonidou: Stochastic energy efficient cloud service provisioning deploying renewable energy sources, IEEE J. Sel. Areas Commun. **34**, 3927–3940 (2016)
19.125 R.S. Tucker, K. Hinton: Energy consumption and energy density in optical and electronic signal processing, IEEE Photonics J. **5**(3), 821–833 (2011)
19.126 N. Abji, A. Tizghadam, A. Leon-Garcia: Energy storage for IP over WDM networks with renewable energy. In: *Proc. 26th Int. Teletraffic Congr. (ITC)* (2014), https://doi.org/10.1109/ITC.2014.6932960
19.127 N. Abji, A. Tizghadam, A. Leon-Garcia: Energy storage management in core networks with renewable energy in time-of-use pricing environments. In: *IEEE Int. Conf. Commun. (ICC)* (2015), https://doi.org/10.1109/ICC.2015.7248311
19.128 K. Hinton, G. Raskutti, P.M. Farrell, R.S. Tucker: Switching energy and device size limits on digital photonic signal processing technologies, IEEE J. Sel. Top. Quantum Electron. **14**(3), 938 (2008)
19.129 M. Bhopalwala, H. Rastegarfar, D.C. Kilper, M. Wang, K. Bergman: Energy efficiency of optical grooming of QAM optical transmission channels, Opt. Express **24**, 2749–2764 (2016)
19.130 J. Xu, C. Peucheret, P. Jeppesen: Power consumption comparison between point-to-point WDM systems and OTDM systems. In: *12th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2010), https://doi.org/10.1109/ICTON.2010.5549296
19.131 K. Georgakilas, A. Tzanakaki: The impact of optical wavelength conversion on the energy efficiency of resilient WDM optical networks. In: *13th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2011), https://doi.org/10.1109/ICTON.2011.5971078
19.132 J. Baliga, R.W.A. Ayre, K. Hinton, R.S. Tucker: Green cloud computing: Balancing energy in processing, storage, and transport, Proc. IEEE **99**, 149–167 (2010)
19.133 J. Baliga, R. Ayre, K. Hinton, R. Tucker: Architectures for energy-efficient IPTV networks. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC)* (2009), https://doi.org/10.1364/OFC.2009.OThQ5
19.134 A.Q. Lawey, T.E.H. El-Gorashi, J.M.H. Elmirghani: Distributed energy-efficient clouds over core networks, J. Lightwave Technol. **32**, 1261–1281 (2014)
19.135 J. Buysse, K. Georgakilas, A. Tzanakaki, M. De Leenheer, B. Dhoedt, C. Develder, P. Demeester: Calculating the minimum bounds of energy consumption for cloud networks. In: *Proc. 20th Int. Conf. Comput. Commun. Netw. (ICCCN)* (2011), https://doi.org/10.1109/ICCCN.2011.6005778
19.136 B. Aksanli, T.S. Rosing, I. Monga: Benefits of green energy and proportionality in high speed wide area networks connecting data centers. In: *Design Automat. Test Eur. Conf. Exhib. (DATE)* (2012), https://doi.org/10.1109/DATE.2012.6176458
19.137 K. Guan, D. Kilper, G. Atkinson: Evaluating the energy benefit of dynamic optical bypass for content delivery. In: *IEEE Conf. Comput. Commun. Worksh.* (2011), https://doi.org/10.1109/INFCOMW.2011.5928830
19.138 X. Wang, S. Thota, M. Tornatore, H.-S. Chung, H.-H. Lee, S. Park, B. Mukherjee: Energy-efficient virtual base station formation in optical-access-enabled cloud-RAN, J. Sel. Areas Commun. **34**, 1130–1139 (2016)
19.139 H.H. Bazzaz, M. Tewari, G. Wang, G. Porter, T.S.E. Ng, D.G. Anderson, M. Kaminsky, M.A. Kozuch, A. Vahdat: Switching the optical divide: Fundamental challenges for hybrid electrical/optical datacenter networks. In: *2nd ACM Symp. Cloud Comput. (SOCC)* (2011), https://doi.org/10.1145/2038916.2038946
19.140 M. Fiorani, S. Aleksic, P. Monti, J. Chen, M. Casoni, L. Wosinska: Energy-efficiency of an integrated intra-data-center and core network with edge caching, J. Opt. Commun. Netw. **6**, 421–432 (2014)
19.141 U. Mandal, P. Chowdhury, C. Lange, A. Gladisch, B. Mukherjee: Energy-efficient networking for content distribution over telecom network infrastructure, Opt. Switch. Netw. **10**(4), 393–405 (2013)
19.142 S. Namiki, T. Kurosu, K. Tanizawa, J. Kurumida, T. Hasama, H. Ishikawa, T. Nakatogawa, M. Nakamura, K. Oyamada: Ultrahigh-definition video transmission and extremely green optical networks for future, IEEE J. Sel. Top. Quantum Electron. **17**(2), 446–457 (2011)
19.143 N. Abji, A. Tizghadam, A. Leon-Garcia: Energy efficient content delivery in service provider networks with content caching. In: *IEEE Online Conf. Green Commun.* (2015), https://doi.org/10.1109/OnlineGreenCom.2015.7387374
19.144 K. Guan, G. Atkinson, D.C. Kilper, E. Gulsen: On the energy efficiency of content delivery architectures. In: *IEEE Int. Conf. Commun. Worksh. (ICC)* (2011), https://doi.org/10.1109/iccw.2011.5963557
19.145 U. Lee, I. Rimac, D. Kilper, V. Hilt: Toward energy-efficient content dissemination, IEEE Network **25**, 14–19 (2011)
19.146 N. Choi, K. Guan, D.C. Kilper, G. Atkinson: In-network caching effect on optimal energy consumption in content centric networking. In: *IEEE Int. Conf. Commun.* (2012) pp. 2889–2894
19.147 J. Llorca, K. Guan, G. Atkinson, D.C. Kilper: Energy efficient delivery of immersive video centric services. In: *IEEE Conf. Comput. Commun.* (2012), https://doi.org/10.1109/INFCOM.2012.6195536
19.148 C.-L. I, C. Rowell, S. Han, Z. Xu, G. Li, Z. Pan: Toward green and soft: A 5G perspective, IEEE Commun. Mag. **52**(2), 77–73 (2014)
19.149 A. Tzanakaki, M.P. Anastasopoulos, S. Peng, B. Rofoee, Y. Yan, D. Simeonidou, G. Landi, G. Bernini, N. Ciulli, J.F. Riera, E. Escalona, J.A. Garcia-Espin, K. Katsalis, T. Korakis: A converged network architecture for energy efficient mobile cloud computing. In: *Int. Conf. Opt. Netw. Design Model.* (2014) pp. 120–125
19.150 A. Tzanakaki, M. Anastasopoulos, I. Berberana, D. Syrivelis, P. Flegkas, T. Korakis, D. Camps Mur, I. Demirkol, J. Gutierrez, E. Grass, Q. Wei, E. Pa-

teromichelakis, N. Vucic, A. Fehske, M. Grieger, M. Eiselt, J. Bartelt, G. Fettweis, G. Lyberopoulos, E. Theodoropoulou, D. Simeonidou: Wireless-optical network convergence enabling the 5G architecture to support operational and end-user services, IEEE Commun. Mag. **55**(19), 184–192 (2017)

19.151 X. Gog, Z. Ning, L. Guo, X. Wei, Q. Song: Location-recommendation-aware virtual network embedding in energy-efficient optical-wireless hybrid networks supporting 5G models, IEEE Access **4**, 3065–3075 (2016)

19.152 C. Despins, F. Labeau, T. Le Ngoc, R. Labelle, M. Cheriet, C. Thibeault, F. Gagnon, A. Leon-Garcia, O. Cherkaoui, B. St. Arnaud, J. Mcneill, Y. Lemieux, M. Lemay: Leveraging green communications for carbon emission reductions: Techniques, testbeds, and emerging carbon footprint standards, IEEE Commun. Mag. **49**(8), 101–109 (2011)

19.153 M.S. Hasan, Y. Kouki, T. Ledoux, J.-L. Pazat: Exploiting renewable sources: When green SLA becomes a possible reality in cloud computing, IEEE Trans. Cloud Comp. **5**, 249–262 (2017)

19.154 K. Hatakeyama, Y. Osana, S.-I. Kuribayashi: Reducing total energy consumption with collaboration between network and end systems. In: *Int. Conf. Netw.-Based Inf. Syst.* (2009), https://doi.org/10.1109/NBiS.2009.15

Daniel C. Kilper

Optical Sciences
University of Arizona
Tucson, AZ, USA
dkilper@optics.arizona.edu

Dan Kilper is Research Professor in the College of Optical Sciences and Electrical and Computer Engineering at the University of Arizona and holds an adjunct faculty position in the Data Science Institute at Columbia University. From 2000–2013, he was a Member-of-Technical-Staff at Bell Labs and he has made contributions primarily spanning energy-efficient communication networks, optical performance monitoring, and dynamic optical networks.

20. Optical Packet Switching and Optical Burst Switching

Pablo Jesus Argibay-Losada, Dominique Chiaroni, Chunming Qiao

Optical transmission has long been the established choice for nonwireless data transmission spanning distances longer than a few tens of meters, due to its high bandwidth and electromagnetic noise immunity. Most current high-bandwidth networks are essentially a group of fiber-optic links connected by nodes whose function is to forward incoming data to the appropriate output. The input to these nodes is data in optical form, and their output is also data in optical form. Thus, it makes sense to also process the data in the optical domain in order to have simple node architectures that improve reliability, performance, and cost. However, optical node technology cannot yet match the flexibility of electronic technology: The main roadblocks are the lack of random-access optical memories and of optical processors. Industry has addressed this lack by creating two kinds of nodes that achieve opposite extremes in the trade-off between efficiency and complexity: electronic packet switching (EPS) nodes and optical circuit switching (OCS) nodes.

EPS nodes transform all optical data (payloads and headers) to the electronic domain and back in order to process the header, wasting lots of power in the conversion process and leading to complex node architectures. OCS nodes do not process passing optical data at all, and instead optically transfer the optical signal from one input to its output using mirrors, prisms, or equivalent means. As OCS nodes do not process data, they cannot dynamically know which output corresponds to each entering packet, and as a result the OCS network is a collection of end-to-end circuits. Thus, OCS nodes must be configured before data transmission begins in a flow, and the network edge must perform access control to only allow relevant traffic to enter a given circuit. This mode of operation is obviously highly inefficient in terms of bandwidth usage but slow to create connections: OCS nodes can be configured either statically (timescales of years), manually (timescales of weeks), or automatically (timescales of seconds).

Optical burst switching (OBS) and optical packet switching (OPS) are solutions for bridging the gap between the power demands and flexibility of EPS and the inefficiency and simplicity of OCS. Their nodes allow electronic processing of packet headers but keeping payloads in the optical domain; as payloads are usually much larger than headers in most applications, their energy savings are large compared with EPS. They allow the processing of packets one by one, avoiding bandwidth wastage and the startup delay characteristic of OCS circuit setup. OPS tries to closely emulate EPS by keeping headers and payloads in the same channel; OBS takes headers to an out-of-band chan-

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_20

nel that can be processed in a conventional, EPS way, and also offers the possibility to group, in larger packets called bursts, packets that share a route. OBS is thus more amenable to implementation with current technology than OPS. This chapter describes in detail both technologies and provides an overview of the main milestones in their evolution and perspectives for the future.

20.1 Technology Basics

Optical burst switching (OBS) and optical packet switching (OPS) were identified a long time ago as promising technologies to replace or complement electronic packet switching (EPS) systems. This chapter describes their fundamentals, in order to understand the main directions explored by the research community.

OPS was proposed in the late 1980s to describe a new technology able to directly switch data in the optical domain [20.1]. Control was done in the electronic domain due to the advantages and maturity of electronic processing technology; the objective was to maintain data in the optical domain. However, early OPS designs relied on the availability of optical memories to limit packet loss. These memories could be random access or simpler ones such as fiber delay lines; in any case, these memories were hard to manufacture and control. As a result, OPS networks were only able to be created in very limited prototypes that did not lead to practical deployments.

OBS was then proposed in the mid-1990s [20.2] to overcome the need for optical memories. Its basic principle was to avoid the need for optical memories by sending, along an out-of-band channel, the minimal amount of information needed to perform forwarding of the optical packets across the network; this information is transported in special packets called control packets. This out-of-band channel is managed with conventional EPS technology and determines the configuration of the all-optical switches that switch the data. Optical packets typically follow a short time after the respective control packets. As the control packets only need to contain forwarding information, typically a destination address, OBS allows most of the data that needs to traverse the network to be left in the optical domain. In addition, OBS naturally allows packets that share input and output nodes to be grouped into large switching units, or bursts, as they share forwarding information, i.e., the control packet. Bursts thus further increase the proportion of data that is kept in the optical domain, avoiding complex and energy-hungry electrooptical conversion.

OBS and OPS have been mainly proposed for the implementation of datagram networks. Datagram networks send data without performing end-to-end resource reservation. Thus, they avoid the need to block connections, relying on congestion control algorithms to dynamically adapt the sending rate of the sources to the state of the network. As many popular applications have been developed for datagram networks (e.g., anything running on the Internet), datagram networks are the most straightforward option for OPS and OBS. Compared with EPS datagram networks, however, OPS and OBS typically present higher losses, as their nodes do not have optical memories to deal with contention. To overcome the pernicious effects of these higher loss levels, special-purpose end-to-end congestion control algorithms run at the ends of the network, allowing conventional applications to run unmodified over OPS and OBS, with performance levels in terms of throughput and flow completion times that match and surpass those in equivalent EPS networks.

The datagram model makes sense for most applications developed in the context of the Internet. However, another possibility in all-optical networks is to use end-to-end resource reservation, where optical paths are reserved along the entire network before data transmission starts. This allows packet loss to be avoided, but at the price of denying service to connections, imposing startup delay when setting up circuits, and adding non-negligible complexity to the control plane. These circuit technologies are not well suited to provide service for most applications; instead, they are most useful to create clearly separated networks that share one physical infrastructure. An example of this is to divide a given all-optical network into several subnetworks with the same topology as the main one but lower capacities. Each subnetwork may then be leased to a different provider, and the traffic in each subnetwork will not affect the traffic in the others. This service is costly, as it does not allow the leverage of statistical multiplexing advantages that have been key to the development of the Internet. Anyway, for scenarios where this may be appropriate, all-optical technologies based on resource reservation have already been proposed and used, viz. optical label switching, optical flow switching, optical slot switching, wavelength division multiplexing (WDM) networks, and many other embodiments of the optical circuit switching paradigm. These circuit reservation networks can use optical packet switches to implement their nodes; they can also use slower optical circuit switches depending on the coarseness of their switching unit granularity.

20.2 Optical Burst Switching: Concept and Technology

20.2.1 Technology of Transmission

Energy dissipation in electronic devices increases with the bit rate. Conventional high-bit-rate packet switches experience extremely high power consumption, which leads to the need for powerful cooling systems and large device footprints. In practice, this directly constrains the deployment of electronic networks.

As the data transmission in high-bit-rate networks is predominantly optical, the goal of optical packet switching technology is to tackle the aforementioned problems by avoiding data conversion to the electronic domain to the greatest extent possible.

Packet switches, whether electronic or optical, need to be able to detect the presence of incoming packets and to know their destinations in order to forward them to their output ports if these are available, or to decide what to do if these are not (e.g., buffer them if buffers are available, drop them, route them to alternate available ports if any, etc.). Electronic switches have mature technology available to perform these operations. In contrast, optical packet switches try to do the same but relying as little as is practical on electronics but using less mature, optical technology. The main difficulties for optical packet switches arise from the lack of optical processor technologies and of random-access memories to implement the buffers used to deal with contention and packet losses.

Optical burst switching (OBS) [20.2–4] is a practical approach to the deployment of optical packet switching networks. As optical data processors are still too immature for deployment, OBS relies instead on electronic processing to manage the bits of information needed to perform its forwarding operations; in OBS, this information is sent separately from the main packet payloads, which remain in optical form throughout the entire OBS network. As packet headers are usually much smaller than packet payloads, OBS is thus able to save large amounts of electronic processing compared with electronic networks.

In addition, OBS takes advantage of the fact that high-bit-rate networks typically aggregate many flows from bursty applications – coming from users on metropolitan area networks (MANs)/wide area networks (WANs), or from workload-heavy distributed applications in datacenter local area networks (LANs) – leading to situations where the packets that arrive closely at a given ingress node often share a network egress node. As all of these packets share forwarding information while traveling through the network, it is possible to aggregate them into larger packets whose forwarding information is sent only once. Such a concatenation is customarily called a burst, while the forwarding information that needs electronic processing is called a control packet.

Savings in electronic processing-related activities arise then from two facts: First that, in a packet switching network, only packet headers need to be subject to processing inside the switches, i.e., payloads are irrelevant for the switching process, and second, that empirical traffic patterns tend to produce trains of packets that share a source and destination, thereby allowing them to share a single header in their transit through the network. From the point of view of network deployment, this allows direct benefits with noticeable gains in power efficiency and footprint compared with electronic networks that must electronically process both the packet headers and payloads. From the point of view of application performance, an OBS network offers flows an environment that has much lower end-to-end delay than electronic networks due to the lack of buffers in OBS switches, although this also makes the OBS network more prone to losses. These lower-delay and higher-loss characteristics tend to cancel each other out when using adequate congestion control algorithms across the OBS network, but usually with a noticeable advantage for OBS in many important scenarios. This, in turn, implies that OBS allows the achievement of higher application throughputs with zero losses, and consequently lower flow completion times, than equivalent electronic packet switching networks. OBS is thus an effective replacement for existing electronic networks.

An overview of OBS networks and their fundamentals, characteristics, performance, and applications is now provided. First, an overview of the main rationale for avoiding electronic packet switching in high-bit-rate scenarios is given. Then, the fundamentals behind OBS are presented. Afterwards, a detailed description of the multiple flavors of OBS is given, followed by the presentation of several applications that are especially well suited for OBS, including datacenters and MANs. Next, performance issues and implementation factors relevant to the design and deployment of OBS networks are discussed.

20.2.2 Why OBS?

Optical transmission has long been the preferred way of transmitting data across distances over a few hundred meters due to its high bit rates and low error probabilities [20.5]. The only practical exceptions to the supremacy of optical transmission are found in niche applications such as wireless transmission in environ-

ments where laying optical fiber is not an option (due, e.g., to cost or mobility) or due to the slightly longer latency of electromagnetic signals in glass with respect to air.

As optical technology is already the preferred means to manage transmission between network nodes, a logical step is to try to introduce it inside the network nodes themselves. The potential benefits of doing this [20.6] are similar to those found in the transmission field, namely the ease of achieving high bit rates, the lack of interference from the electromagnetic environment, and the low cooling needs derived from the use of light instead of electric currents that dissipate heat due to the resistive effects of semiconductor substrates.

Accordingly, intense research has been carried out over recent decades on the introduction of optical processing in the realm of optical switching [20.7]. Optical circuit switches have already been successfully developed and deployed in networks to create static or semistatic circuits to serve high-bit-rate streams, with common circuit lifetimes of days to years [20.8]. These devices have traditionally relied on manual operation for configuration or on automated control planes that take a substantial fraction of a second, or more, to change their internal configuration. These switches typically operate in a purely circuit-switched fashion, where outputs are not shared among several inputs, and as a result do not need any kind of data buffering. This is convenient since effective optical buffer devices do not exist yet, i.e., able to meaningfully compete in performance with electronic buffers. Data are thus allowed to bypass power-hungry electronic processing, never leaving optical form from their input to output.

Optical circuit switches are too slow to offer the sub-μs switching speeds needed to operate effectively as packet switches. These speeds would enable them to directly manage conventional packet data such as that found on the Internet and numerous private LANs, MANs, and WANs. Optical circuit switches are thus relegated to the creation of end-to-end circuits across existing topologies; these circuits will in turn act as a single hop for an overlayed packet switching network. This kind of arrangement, although common, introduces a middle layer that would be otherwise unnecessary, increasing network inefficiencies.

Thus, one of the goals of the optical switching research community is to create fast optical switches that can act as effective packet switches. Conceptually, the most straightforward option here is to replicate as far as possible in the optical domain the architecture of electronic packet switches. However, this has proven difficult: Most packet switch architectures [20.9] have essentially three main parts: A processor of packet headers, a fast switching device, and buffers. All three have mature solutions in the electronic realm: In the optical realm, however, only fast switching devices have been effectively manufactured [20.10]. Optical processors will probably need a considerable amount of time to mature before being available due to the difficulties involved in processing light. Optical buffers have also proved elusive to build: Optical random-access memory (RAM) seems still far from being available, and fiber delay lines are bulky and difficult to use compared with their electronic counterparts, and limited in performance. Thus, in practice, the lack of optical processors and the lack of optical buffers have traditionally been the main roadblocks that must be solved to enable optical packet switching technologies.

OBS is an optical packet switching network technology that solves the two aforementioned constraints in the following way: Regarding the first constraint, and as optical processors do not exist in practical form yet, OBS uses electronic processing for the packet headers; in this way, it can readily configure the fast optical switch that will in turn manage the passing packet payloads. As for the second constraint, viz. the lack of optical buffers, recent research [20.11–14] has shown that even a bufferless OBS network can match and surpass the throughput and flow-completion-time performance of an equivalent electronic packet switching network with plentiful buffers, through the use of suitable end-to-end congestion control algorithms across the OBS network. In practice, this means that the OBS network does not need buffers in its switches to outperform electronic networks in terms of throughput at zero losses and in terms of end-to-end delay, and they do so without paying any price in implementation complexity inside the switches nor in the network control plane. In addition, OBS can naturally increase efficiency by combining several packet payloads that share forwarding information into a single data burst. This is a common occurrence in current networks, where sources are bursty and packets to the same destination often appear close to each other. This allows the OBS network to use a single header to route several payloads at the same time throughout the network core, drastically decreasing the amount of data that needs electronic processing inside the network.

20.2.3 What Is OBS?

From an architectural point of view, OBS networks are composed of nodes that perform one or more of the following functions [20.2, 3]:

1. *Ingress nodes:* These nodes receive packets from non-OBS networks and start the process of forwarding them through the OBS core. They may

aggregate several packets that share an OBS egress node, then create and send their common forwarding information, or control packet, to the next hop, followed a short time later by a data burst containing the payloads.
2. *Egress nodes:* These nodes receive, from the interfaces to other OBS nodes, control packets and their subsequent data bursts destined to non-OBS networks reachable through them. They transform the data burst to the electronic domain, extracting in the process all the packets that may be inside.
3. *Core nodes:* These nodes send and receive control packets for traffic that traverses them from an OBS interface to another OBS interface. They act thus purely as a switch for the OBS network, processing the control packet and configuring accordingly the optical switching element in order to relay the data burst towards its next hop.

These three roles are not exclusive. Practical OBS nodes can combine more than one, and indeed all three. The above-described processes are in essence equivalent to others found in conventional electronic networks, save for the aggregation process that takes several data packets and assembles them into a single burst. As in electronic networks, there exist multiple possibilities to manage these processes in OBS networks. Let us take a closer look at these [20.15].

Ingress and Egress Nodes

Ingress nodes create the control packets and data bursts that will be managed by the rest of the OBS nodes. Burst creation implies the possibility of aggregating all the packets aimed at a given egress OBS node. Many possibilities exist here for the algorithm that performs this function, and the literature has described several [20.16]: for example, waiting a certain amount of time to aggregate all packets received in that window, or waiting until a given number of packets have been received before assembling them into a burst, or a mix of both approaches, where bursts are sent at the earliest occurrence of a given timeout or the availability for transmission of a given number of packets. These possibilities cover the spectrum of efficiency in the usage of the electronic control plane with respect to the amount of data sent in the data plane, and the spectrum of latency in the OBS ingress node while waiting for transmission. Different burst assembly algorithms will achieve different trade-offs in those two spectra; these in general will lead to different reactions with the sources creating the packets, usually the transmission control protocol (TCP) [20.17, 18]. Burst assembly algorithms thus act as aggregators of traffic, and in contrast to the effects of conventional buffers in electronic networks, they lead to a decrease in traffic burstiness [20.19] rather than an increase, easing traffic management concerns.

Once the control packet has been created, there are several possibilities for how to proceed [20.2, 3]. First, there can be variations about whether out-of-band signaling or in-band signaling will be used: In out-of-band signaling, the control packet uses a link that is not the same as the one the burst will follow; in in-band signaling, both are the same. Another degree of flexibility relates to how long to wait before sending the burst once the control packet has been sent [20.17]. One possibility here is to send it right away, thus on arrival to each switch the burst will need to be delayed a time roughly equal to the electronic buffering and processing time of the control packet, in order to ensure that the switch has enough time to process the control packet and change its internal configuration accordingly (this is called TAG, or tell-and-go); fiber delay lines can be used for this purpose. Another possibility is to wait a given time, or offset time, in the OBS ingress node before sending the burst after the control packet has been sent, and to not use any delaying procedure inside the OBS network switches; this gives the control packet a time advantage that will progressively decrease as the control packet is delayed slightly in each switch due to its electronic processing while the data burst does not suffer any comparable delay. The key here is to make the offset time long enough for the last OBS core node to have sufficient time to forward the burst successfully: The last OBS core node is, among all the OBS nodes crossed by the control packet and burst, the one where the time separation between both will be shortest.

Egress nodes are in charge of receiving control packets and data bursts and of converting the burst payloads into groups of conventional electronic packets that will be forwarded in a conventional way to their next hop towards their final destination. Their implementation is straightforward compared with ingress nodes.

Core Nodes

Core nodes forward OBS control packets and bursts. They run a scheduling process in the same way as in electronic networks. However, a fundamental difference in OBS networks with respect to electronic networks is that no RAM buffers are available in OBS, while alternatives such as fiber delay lines are inefficient and bulky; although, if they are used, several families of algorithms may be used to manage them, trying to optimize the application or switch performance. For a given incoming burst, an OBS scheduling algorithm will typically need to decide whether to forward it (if

its output port is available), drop it (if the output port is not available), reroute it, or allow preemption of some kind. The possibilities and trade-offs here are similar to those in electronic networks and circuit switching networks (e.g., dynamic alternate routing). Depending on the capabilities of the hardware, several scheduling policies are possible [20.20]; for example, the literature has described algorithms that delay bandwidth reservation, or just-enough-time (JET) [20.2], efficiently using voids between control packets and their bursts to schedule other bursts that are able to pass through in that interval. Other alternatives use immediate reservation, or just-in-time (JIT) [20.21], reserving bandwidth for a burst once its control packet arrives. These two can be thought of as two extremes of a spectrum that trades off implementation complexity against bandwidth utilization efficiency and illustrate how an OBS network may be built from a variety of heterogeneous optoelectronic technologies in different states of maturity. OBS schedulers can thus be effectively implemented in practice; prototypes have already been created in field-programmable gate arrays (FPGAs) [20.22].

If the scheduling process does not find an available output for a given burst, the simplest option is to drop it, either fully or partially [20.23]. Other options are to use alternative routes [20.24, 25], or to convert the burst wavelength to another one that is free if the network features that capability [20.26, 27]. If fiber delay lines are available, they can be used as well to reduce the likelihood of burst drops [20.2, 28]. In general, dimensioning tools can be readily used to configure the network to attain a given target for the loss probability [20.29]. If an alternate route can be identified and is available, multipathing may be used. As always with multipathing, reordering may occur, but higher-layer protocols such as TCP can take care of this: The specific combination of the OBS network and application will thus dictate whether multipathing is a reasonable option.

The behavior when dropping bursts because outputs are busy is very different from what happens in electronic networks, where buffers are used to store that data. As an OBS network does not have buffers, the associated scheduling and queuing issues are moot. Losses are unavoidably higher; to compensate them and offer a lossless service to applications, congestion control algorithms running at the OBS edges can be used [20.11–13]: These algorithms can match and outperform the throughput and flow completion time of applications running on top of equivalent packet switching networks with the same topology and link bit rates. This is a result of the low-delay characteristic of bufferless networks, which allows compensation of their a priori higher losses.

Quality of Service

Quality-of-service (QoS) issues can also play a role in the scheduling of OBS bursts [20.2, 30]. As in electronic networks, there are many possibilities, e.g., high-priority traffic preempting lower-priority bursts, high-priority traffic being reserved a given number of channels between nodes, etc. For some QoS algorithms, QoS information may need to be available inside the control packet to allow the control plane to explicitly know the QoS level associated with each burst and thus determine whether preemption or any other possibility is allowed. Some QoS schemes [20.31] modify the offset time between a control packet and its burst, in such a way that longer offset times lead to higher probabilities of the burst's not being blocked inside the OBS nodes. Other QoS schemes, such as segmentation, combine packets with different QoS levels inside a single burst [20.32–34], using burst preemption in ways that lead to a higher probability of losing packets towards the end of a given burst; in these schemes, lower-priority packets are thus placed towards the end of bursts. In general, multiple options [20.35] are available if QoS requisites need to be explicitly managed inside the OBS network. Thus, it is possible to fine-tune [20.36–39] the loss probabilities, and thereby the end-to-end delay at zero loss, of the different classes of traffic that pass through the OBS network. Options such as differential pricing are also available to control performance in the framework of utility maximization [20.40].

As a result of the above, OBS can directly support useful QoS architectures, e.g., the Internet Engineering Task Force (IETF) differentiated services (DiffServ) architecture [20.41], whose techniques are broadly available and counted on by many network providers. Protocols such as the session initiation protocol (SIP) that are used to create session-oriented applications can also be used effectively across OBS [20.42]. The practical advantages of OBS in terms of avoidance of optoelectronic conversions jointly with the possibility of fine-tuning its parameters also make it attractive for use in space-constrained deployments, e.g., onboard networks in satellites [20.43].

20.2.4 Energy Savings Compared with Electronic Networks

The basic idea of OBS is to aggregate as much data as possible at the edge of the network before sending it out in a single burst towards the destination. This burst is preceded by a given amount of time by a special control packet that contains routing information for that burst; this control packet will undergo electronic processing in

the intermediate switches in order to guide their control planes in the management of the optical switch itself. Thus, we have:

1. Total avoidance of electronic processing for the burst, which crosses each OBS switch in entirely optical form.
2. Usage of electronic processing for the control packet, in order to avoid the hurdles associated with all-optical processing of information.

For a given amount of data A (bit) in the burst and a given amount of data B (bit) in the control packet needed to direct the forwarding process, OBS processes electronically an amount $B/(A+B)$, allowing a fraction $A/(A+B)$ of all the passing information to cut through the switch in optical form. Obviously, in order to maximize $A/(A+B)$, one must minimize B.

The control packet, with a size of B bit, contains the routing information for the burst. Assuming an OBS network with N egress nodes and a fixed-length field to encode that destination, we can estimate $B \approx \log_2(N)$; for example, in a 1024-node OBS network, one needs $C = 10$ bit in the control packet dedicated to routing information. This would be a bare minimum, and in practice it would probably be advantageous to have a field to indicate the protocol being carried by the control packet itself, for example, to manage the address resolution protocol (ARP). A second field for the source address is common as well in many network technologies, to facilitate operations. Common network technologies illustrate these possibilities: For example, in the LAN arena, the ubiquitous 802.3 specification [20.44] uses a 48-bit field for the destination address and another for the source address, with 16 bit for the protocol carried in the payload. These addresses by convention encode the network interface card (NIC) manufacturer in the first three bytes, with the remaining three reserved for the address itself. Assuming that the OBS network uses a similar scheme, this would mean that the network can connect more than 16 million nodes. In practice, LAN networks are much smaller, often only connecting a few tens or hundreds of nodes at most. The same holds for MANs or WANs belonging to the same administrative entity, e.g., an autonomous system (AS). In the WAN arena for networks encompassing different ASs, the Internet at large as supported by IPv4 can directly address up to 4 billion nodes, with 32-bit address fields.

These examples clearly indicate that control packets can easily be limited to a few tens of bytes without any particular hindrance to their expected functionality. Assuming that we use 8 byte for the address of the destination, another 8 byte for the address of the source, 2 byte for the protocol, and 2 byte for reserved purposes, this would lead to extremely versatile 20-byte control packets.

Let us focus now on the size of the optical payload, A: If the burst consists of a single 1500-byte Ethernet maximum transmission unit (MTU)-sized frame (i.e., with no aggregation of several frames in the burst), the fraction of electronically processed data among all the information sent via the OBS network would be $20/(1500+20) = 1.3\%$. If the payload were a single TCP acknowledgment (ACK) packet with a size of 40 byte, that fraction would be $20/(40+20) = 33\%$. If the network workload consisted exclusively of these two kinds of packets appearing in the same proportion as each other (which is a reasonable approximation in practice [20.45]) and no aggregation happens in the OBS network edge, the fraction of electronically processed data would be 40/1540, i.e., close to 2.6%. Thus, even without any packet aggregation in the OBS ingress nodes, OBS can directly lead to a decrease of 97% in energy costs associated with electronic conversion.

As explained above, the proportion of data being processed electronically in OBS switches can be decreased even more from the previous 2.6% figure by taking advantage of the bursty nature of many applications, and of the effect of network buffers on nonbursty sources of data: Whenever a smooth, nonbursty source of data encounters a busy link in a switch/router and is thus enqueued in a buffer feeding that link, the data will join the rest of the data buffered there, plus the data being sent, inside a bigger burst; this bigger burst will be made even bigger by the data that will arrive while it is being sent. In other words, buffers make network traffic burstier at their outputs than at their inputs: If the OBS network is fed by conventional networks, the traffic arriving at the OBS ingress nodes can be expected to be a burstier version of the traffic that was generated by the applications due to the mere presence of buffers upstream from the OBS network, and this happens independently of whether the end applications generating the traffic are themselves bursty [20.12, 13].

In addition, end applications are often bursty. For example, web servers often act in a reactive manner, responding to requests with replies that often contain large files, but being silent the rest of the time due to browsing habits. In another example, video codecs are often designed in a way that concentrates data in definite moments in time, e.g., with Moving Picture Experts Group (MPEG) compression.

Moreover, protocols such as TCP, used pervasively in the Internet, often impose bursty characteristics on the data being sent by their applications due to the effects of congestion control algorithms implemented inside those transport protocols. For example, TCP

Reno and its many variants maintain a window that indicates the maximum amount of unacknowledged data that they are allowed to have in flight; if that amount is reached, they cannot send more, even if the application is sending data to TCP. Then, when acknowledgments (ACKs) arrive, a corresponding amount of buffered data is sent in a burst, in addition to any excess data allowed to probe for additional bandwidth.

In practice, all of these factors mean that applications often have bursty patterns, which are often amplified by the congestion control algorithms inside TCP, which are then also amplified by the existence of buffers inside electronic network switches. Such traffic is then fed to the OBS network. Many packets thus enter the OBS ingress nodes close to each other, and many of then will need to be sent to the same OBS egress node, either because they share the same final destination outside the OBS network or because that OBS egress node leads towards their respective final destinations. As the OBS network only needs a single control packet to route an entire burst of packets that share the same OBS egress node, the proportion of data that needs to be processed electronically in the intermediate OBS switches decreases correspondingly: Let us assume that bursts mix in the same proportions 1500-byte and 40-byte packets, and thus the average packet size is 770 byte. Then, if an OBS burst carries ten packets on average, the electronic processing in OBS can be estimated as $20/7700 = 0.26\%$ of the amount processed by an equivalent electronic network. Figure 20.1 plots this ratio as a function of the average number of constant-sized packets. In practice, once the burst carries more than a handful of packets, the total amount of electronic processing becomes negligible compared with the total amount of information flowing through the entire network.

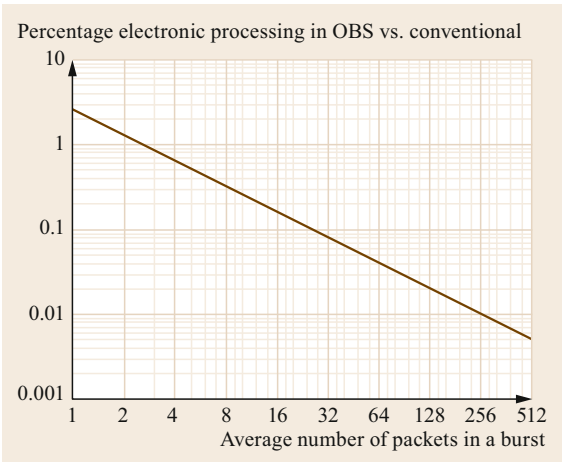


Fig. 20.1 Percentage of data processed electronically in an OBS network compared with the amount processed in an equivalent electronic network. The workload is assumed to be 50% 1500-byte packets and 50% 40-byte packets

In summary, OBS networks managing conventional traffic can be expected to reduce the energy spent in traffic processing to less than 3% of the original value. As most realistic traffic conditions are bursty due to application behavior, congestion control algorithms, and buffers in electronic networks, this 3% figure can reasonably be expected to become even smaller.

20.2.5 OBS Applications

OBS, as a network technology that takes advantage of statistical multiplexing, is a natural fit for any kind of environment where sources are bursty and where it is not economical to use dedicated end-to-end circuits for all end user/application pairs, as testbeds have shown [20.46–49]. This section discusses some relevant scenarios where OBS is an obvious fit to improve performance.

Local Area Networks (LANs)

LANs in need of large bandwidths are an obvious field of application of OBS due to the high bit rates allowed by optical transmission and their usual energy consumption constraints [20.11]. This is often the case for massive datacenter LANs, especially those where there is a strong need for delay-bounded services, such as those hosting the back-ends for interactive services.

An illustrative scenario of this kind would be a datacenter that manages the back-end of search engines that rely on massively parallel and distributed applications. For example, MapReduce applications and their derivatives often exhibit such behavior: When a datacenter gateway for a search engine receives a petition for a given search, a multitude of queries are transferred over the datacenter network towards internal nodes that are likely to host relevant information about the query. Then, after a given time, all those nodes will relay whatever relevant information they could find in their local resources towards another node or nodes that will combine all the information before creating the final response towards the external user. All these processes lead to bursty traffic patterns per query. The datacenter workload consists basically of the superposition of many of these groups of flows. As explained above, OBS is especially effective at handling bursty traffic due to its leverage of statistical multiplexing, and as it allows most electronic processing to be bypassed, it benefits the performance, footprint, and energy consumption of the entire network.

Metropolitan Area Network (MANs)

MANs also have an ever-growing need for high data rates at low delays in order to manage data traffic of most kinds, thus being a suitable scenario for OBS [20.12, 13]. A relevant example here are the applications proposed in the context of 5G networks. Many such applications follow a pattern dictated by the geographical coverage of the MAN; for example, consider a metropolitan area that uses a 5G network to obtain sensor data from a multitude of points scattered throughout the area (for example, real-time traffic and weather conditions at all the intersections of a large city) to then send them to a central controller in charge of processing that data and sending back a series of commands to multiple actuators across the city (for example, traffic light pattern changes or summoning of city services). These applications are currently the focus of a great deal of attention due to the penetration of cloud computing services in city management. The traffic is expected to be bursty due to the nature of sensors, which often concentrate their activity in a few intervals, being silent the rest of the time. Technologies such as circuit switching have been used in the past for similar deployments, but circuit switching does not scale easily, in terms of either cost or deployment constraints. OBS, on the other hand, is an ideal match for this kind of scenarios due to its high bit rates and effective use of statistical multiplexing. Another important practical advantage of OBS here is that OBS switches have a priori lower power needs than their electronic counterparts, since they do not need buffers to manage optical data while providing high throughput and low delay to applications; thus, OBS switches can achieve smaller footprints, less power consumption, and lower cooling needs than electronic nodes. This allows OBS nodes to fit more easily than electronic ones into environments where space is critical (e.g., cabinets in difficult-to-reach places across a city or building).

Wide Area Networks (WANs)

WANs are often created by point-to-point links between routers. These links are also usually created by circuits over an underlying circuit-switched optical network rented from a transport network operator. For the transport network operator, OBS introduces the possibility of doing away entirely with circuits, thus shifting its portfolio towards a packet-switching approach that is uses bandwidth more efficiently than a circuit-switched one. These OBS WAN providers would simply need to implement suitable congestion control algorithms at the ingress/egress points of the OBS network: The net result would be zero-loss service at higher throughputs and lower delays than electronic counterparts.

20.2.6 Performance of OBS

OBS networks are targeted at offering high bit rates, low latencies, and low losses to applications. The state of the art is presently dominated by electronic packet switching (EPS) networks, which, like OBS, use statistical multiplexing to share network resources efficiently among many bursty applications. From a high-level architectural point of view, an OBS network resembles an EPS network, save from the fact that all-optical transmission is used for data bursts and thus electronic processing is greatly minimized in the network core, and from the fact that OBS switches will not necessarily feature any kind of buffering for bursts, be it optical RAM, fiber delay lines, or optoelectronic conversion to buffers in the electronic domain. To counter the contention losses that happen due to this bufferless behavior, congestion control and retransmission should be used at the edges of the OBS network, leading to zero-loss service across the OBS core. In these conditions, and without the need to have any special kind of routing/switching algorithm specific to OBS inside the OBS switches, OBS can match and outperform the application throughput and end-to-end latency offered by EPS in important scenarios. As an illustration, Figs. 20.3 and 20.5 plot the performance that OBS achieves compared with EPS in two scenarios that have already been studied in related work [20.11–13]:

1. A datacenter LAN (Fig. 20.2) with 128 nodes interconnected by a fat tree with three levels of switches. This datacenter runs a MapReduce workload that models the behavior of applications where many entities are constantly being used to collaborate in the search for information (e.g., search engines). In this case, Fig. 20.3 plots the throughput – defined as the size of flows divided by the time it takes for all of them to be transferred with zero losses across the network – of a case with 1000 flows in the MapReduce application; details about the workload itself can be found in [20.11]. Three cases are considered: OBS where an optimized congestion control algorithm (TCP SAW (stop-and-wait) [20.8]) is used, OBS where a conventional congestion control algorithm (AIMD, additive-increase-multiplicative-decrease) is used, and a conventional electronic network with conventional congestion control (AIMD). It is clear that the electronic network is clearly much better than the OBS network with conventional congestion control; however, OBS with optimized congestion control clearly outperforms the electronic network.
2. A MAN (Fig. 20.4) in the context of a 5G application where a multitude of nodes scattered across a city

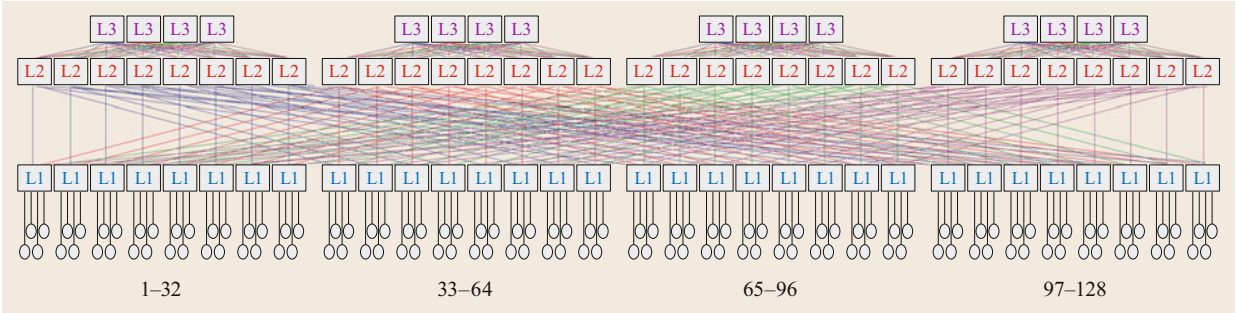


Fig. 20.2 Datacenter LAN: topology (© 1990 IEEE. Reprinted, with permission, from [20.50])

aggregate data from many sources and then send these data continuously to a central node that acts as a service planner (e.g., for traffic optimization across a geographical region). Figure 20.5 plots the average throughput per flow, defined as the size of the flow divided by the time it takes to finish transmission at zero losses, for the cases of OBS with TCP SAW, and electronic networks (EPS) with two choices of congestion control: SAW and SACK (an optimized version of AIMD) and with two different values of the buffers available inside the electronic switches; details about this scenario can be found in [20.12]. Note that the electronic network offers better throughput than OBS when there are few flows and if the electronic switches have large buffers; however, the relative performance of OBS improves as the number of flows increases, and after a given point OBS outperforms the electronic network. This behavior is also found in per-packet delay metrics (refer to [20.12] and references therein for more details).

These results assume that OBS bursts consist of single packets, so they do not take into account the optimizations that result when combining several packets in each burst. Scheduling is as simple as possible: Route to the desired port if available and drop otherwise. Even with this remarkably simple implementation, it is clear that OBS easily matches and often outperforms EPS.

It is interesting to point out that another technology that has also been proposed to leverage all-optical technology to replace EPS in environments such as datacenters is optical circuit switching (OCS), due to the relatively mature and low-cost nature of OCS switches. OCS, as with all circuit-switching techniques, is based on end-to-end resource reservation, preventing any given flow from accessing resources that have already been allocated for other flows (the term *OCS*, as used here, encompasses a broad array of techniques whose common denominator is that they are based on resource reservation, including ones such as optical flow switching (OFS)). The literature has described multiple OCS

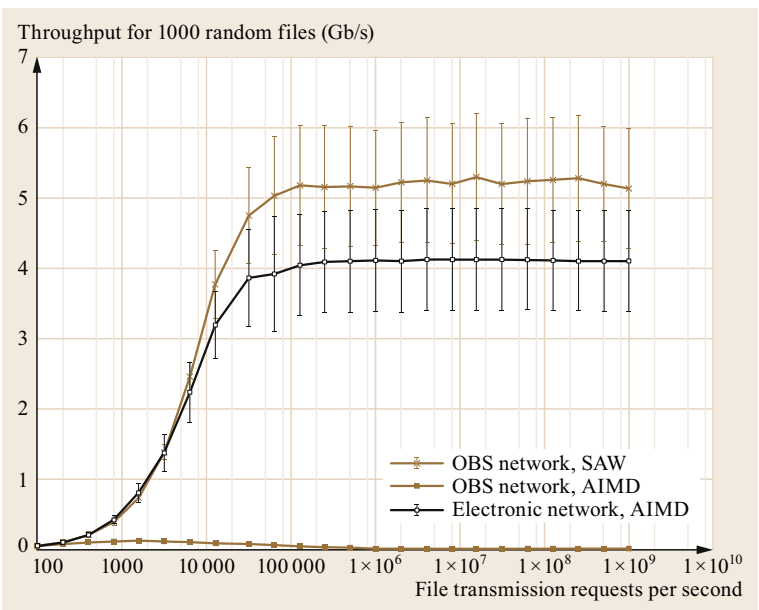


Fig. 20.3 Datacenter LAN: throughput performance of OBS in three cases: Sources that use an optimized congestion control algorithm (SAW), sources that use conventional congestion control (AIMD), and sources using conventional congestion control in a conventional equivalent electronic packet switching network (Adapted with permission from [20.11], The Optical Society (OSA))

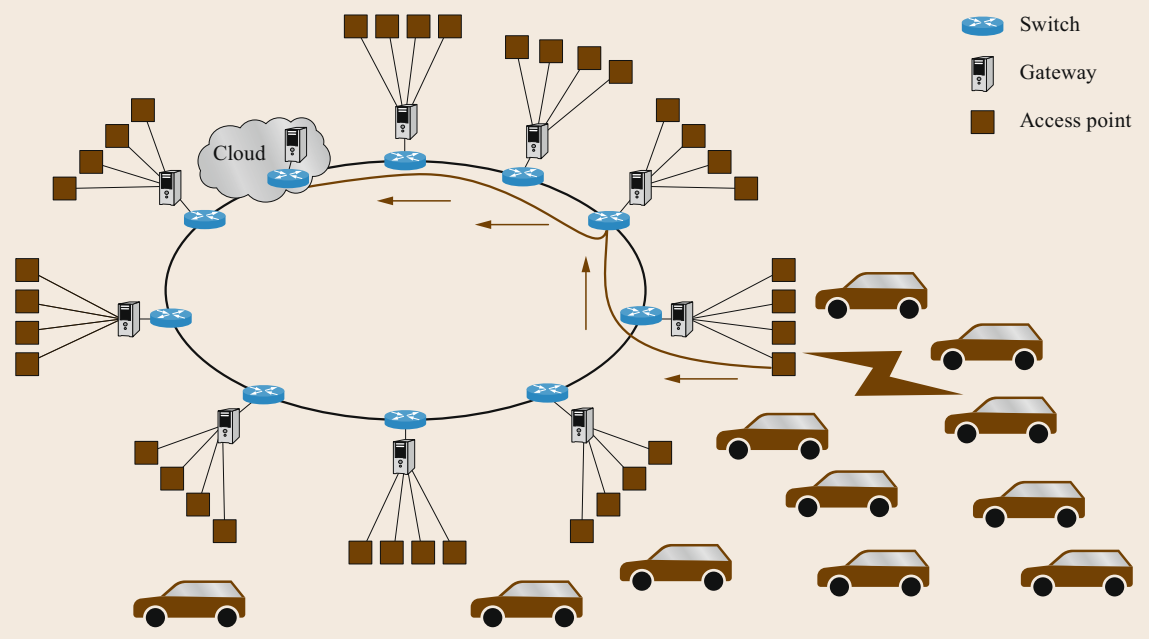


Fig. 20.4 5G MAN: topology (© 2009 IEEE. Reprinted, with permission, from [20.10])

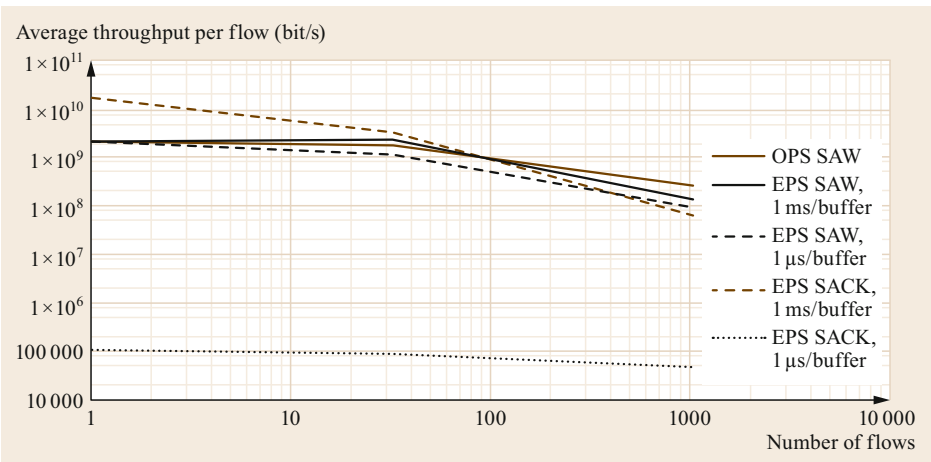


Fig. 20.5 5G MAN: throughput performance of OBS compared with an equivalent electronic packet switching network; MTU = 9 kB (1 MB, 10 MB) files, 10 km circumference (Adapted with permission from [20.12], The Optical Society (OSA))

varieties, but they all share this end-to-end reservation philosophy. As a consequence of it, flows that arrive at the OCS network edge may be momentarily blocked if resources are not available, being either delayed while waiting for them, or dropped. This behavior contrasts with the statistical multiplexing behavior found in OBS, which takes advantage of the bursty nature of many data applications, and leverages the fact that many data applications and transport protocols have not been created with a circuit-switched behavior in mind (i.e., detection of blocking events followed by reattempts, or equivalent alternatives) but rather with a best-effort behavior (i.e., some packets will make it across the network, some will not, while some transport protocol, usually TCP, will likely be used to recover the missed ones). OCS can in theory be a reasonable alternative to EPS when applications can be modified to adapt to it or it is transparently applied to flows, and the amount of data to be sent is much larger than the time that the OCS network takes to set up a circuit. However, the large bit rates of optical networks play against OCS, since most flows in current networks, even most large flows, usually take very little time to be transmitted at the Gb/s or higher rates found in optical networks, thus OCS circuit management al-

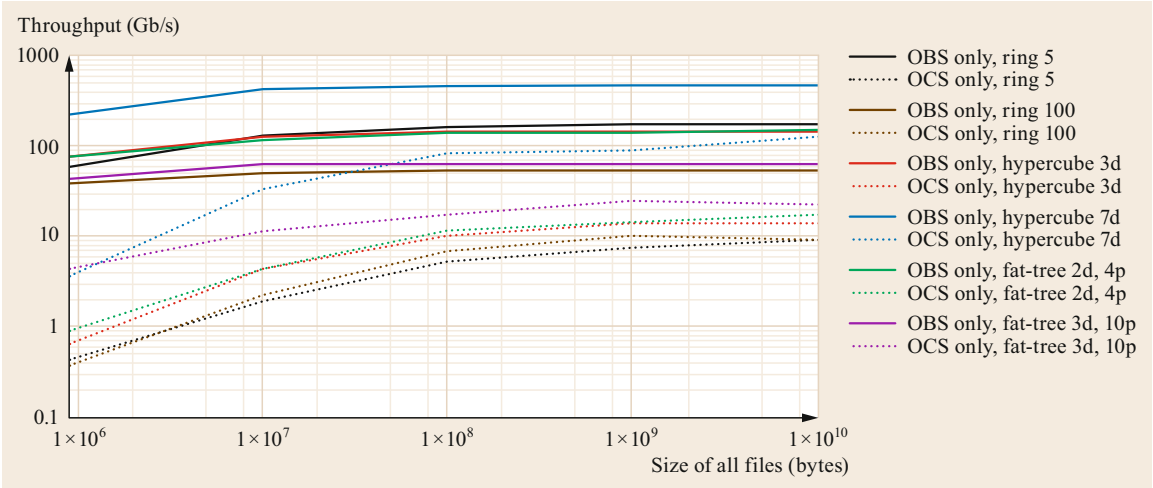


Fig. 20.6 Datacenter LAN: performance of OBS compared with an equivalent optical circuit switching network; 100 files of constant size transmitted in each experiment (Adapted with permission from [20.14], The Optical Society (OSA))

gorithms need to be especially fast to be much faster than those times: Unfortunately, no practical algorithms have been demonstrated that allow anything like that in complex networks with more than a few nodes.

As an illustration of these matters, Fig. 20.6 plots the performance of OBS versus OCS in a variety of networks (small and large rings, small and large hypercubes, and small and large fat trees) representative of datacenter networks, managing a MapReduce workload; details can be found in [20.14]. Figure 20.6 plots the throughput of the distributed application as a function of the number of flows in each MapReduce cycle. As seen here, OBS is much better than OCS at handling the application flows. This holds across the entire range of file sizes, even for very large flows, i.e., those that are often described in literature as well suited to be managed by OCS rather than by packet-switching alternatives such as OBS.

20.2.7 Implementation Issues

OBS networks, as with other all-optical network technologies based on packet switching, need fast optical switches to be able to work effectively. These switches are in general composed of an electronic control plane processing the control packets and of an all-optical switch that will manage the data bursts.

In contrast to electronic packet switches, there is no practical optical RAM available, so OBS switches may be equipped with banks of fiber delay lines, devices that offer fixed delays to bursts in order to control collision probabilities, or with opto-electro-optical (O/E/O) converters to transfer temporarily to the electronic domain the bursts that would otherwise be lost due to contention. In the absence of these devices, or in addition to them, sophisticated schedulers can be implemented in the electronic control plane, taking into account multiple routing possibilities, if available, towards a given destination, similarly to what happens in electronic networks. In general, all these technologies have the potential to increase the flexibility of the OBS network when dealing with congestion. However, as we saw above, buffering is not needed at all to adequately control losses, as totally bufferless all-optical switches have much lower end-to-end delays than equivalent electronic switches precisely due to the absence of buffers: Congestion control algorithms running at the network edge are able to effectively leverage this trade-off, allowing the compensation of the higher packet losses inside the network with faster reaction times. This is enough to overcome the effect of the packet losses, allowing the OBS network to transfer in an end-to-end fashion flows at zero losses, at lower delays, and at comparable or higher throughputs than competing electronic packet switching networks. Moreover, there is no need for complex scheduling schemes either inside the OBS nodes: The results shown above use the simplest strategy possible: Forward to the desired port if it is available, and drop otherwise.

As there is no need for buffering or complex scheduling algorithms, the most critical part of an OBS switch is its switching element. Currently, several technologies are available to manufacture OBS switches, and they feature switching times in the range from ps to ns [20.10]. For comparison, a 100-byte burst at 100 Gb/s lasts 8 ns, a 1-kB burst 80 ns, and a 10-kB one 800 ns. The number of ports in OBS switches is also important in practice: Prototypes with sub-μs la-

tencies have been developed with a handful of ports, usually from two to eight. These sizes are still small compared with equivalent electronic switches, but they are already suitable to implement important and robust network topologies with low node degrees. For example, existing switch prototypes may already be used to implement useful MANs, where ring topologies are prevalent due to cost advantages when deploying links across a city: Scenarios such as 5G applications can thus directly benefit from them. Datacenter networks making use of hypercubes or other similar topologies can also be effectively implemented with already existing switch prototypes.

The OBS network also needs ingress nodes in charge of assembling bursts from packets that share an OBS egress node in the OBS network, and to run the edge-to-edge congestion control algorithms that allow burst losses to be overcome. These activities take place entirely in the electronic domain, so current technology may be used here without the need for any significant technological development.

Finally, OBS transmission can reuse the physical layer (PHY) characteristics that already exist in electronic networks with optical transmission, or in any one of the proposals created for any optical packet switching technology [20.13]. Thus, no significant technological changes are needed to implement the PHY part of OBS edge nodes.

20.2.8 Conclusion for OBS Technology

OBS is an effective replacement for current conventional electronic networks with high-bit-rate transmission, especially in those environments where energy efficiency, application performance, or equipment footprint are critical. The key advantage of OBS is the simplicity and energy efficiency of all-optical switches due to their lack of need for buffers. OBS networks can easily match and outperform equivalent conventional electronic networks in terms of important application performance metrics, viz. throughput (and conversely flow completion time) at zero losses, and end-to-end delay.

OBS, as with all packet switching technologies, leverages statistical multiplexing and thus avoids the resource inefficiencies of circuit switching technologies. It offers good implementation feasibility, since all network elements needed to set up a practical network have already been successfully demonstrated; this is in contrast to the equipment needed by other all-optical options that try to exactly replicate the architecture of electronic packet switching networks in the all-optical domain.

OBS switches are in addition very simple from an architectural point of view: As they do not need buffers for the data that cut through them, they do not need any complex scheduling/queuing scheme either. The only electronics they need are those in charge of processing the control packets, and this is a negligible part of the corresponding load in equivalent conventional electronic networks. This naturally makes OBS a robust network technology.

An OBS network in the path of a given flow presents itself to that flow as a virtual single hop in that path, with zero loss by virtue of the congestion control algorithms controlling the end-to-end transmission across the OBS network. This virtual hop features better throughput and delay characteristics than an equivalent hop across a conventional electronic packet switching network, and with considerable energy and footprint advantages. This makes OBS an extremely well-suited candidate to effectively introduce very high bit rates into environments where these are difficult to introduce or manage due to environmental concerns (e.g., 5G deployments in space-constrained situations) or where energy consumption is important (e.g., massive datacenters).

20.3 Optical Packet Switching: Concept and Technology

The evolution of optical technologies, including research on all-optical variants such as OBS and OPS, is better understood in the broad context of the growth that information and communication technologies (ICT) have experienced over recent decades. ICTs were given a strong boost during the Cold War as a strategic asset for the defense of countries. As the Cold War ceased, commercial interests took the helm in the development of ICTs, creating the modern Internet. The following text recalls several key points that provide understanding of how optical technologies fit into the global ICT context.

20.3.1 Network Evolution Since 1990

Until the 1990s, optical synchronous transmission systems such as synchronous optical networking (SONET)/synchronous digital hierarchy (SDH) systems dominated telco networks. These systems were used to create end-to-end circuits between pairs of nodes. These circuits mainly supported telephony, the main cash cow of the times. Then, telco operators started preparing for the foreseeable emergence of enhanced services such as video telephony and data. These new services had bandwidth needs that were variable, and often higher than

voice. Telco networks were expected to manage both voice and enhanced services. To address this heterogeneity, the telco industry decided to implement a packet switching layer, either on top of existing synchronous lines or on raw ones. This packet switching layer was called asynchronous transfer mode (ATM). Negotiations in the telco sector led to the usage of relatively small ATM packets called cells, with a 53 byte size, viz. 5 byte for the header and 48 byte for the payload, lasting 42.4 ns at 10 Gb/s. This small size was targeted at limiting potential jitter encountered by high-priority traffic sharing non-preemptive queues with low-priority traffic. This small packet size also led to stringent requirements on processing of headers. At 10 Gb/s, for a 32×32 ATM switch, header processing must take less than 1325 ns if all lines are busy. At 100 Gb/s, the processing of each header must be done in less than 132 ps. Although promising, ATM technology was abandoned due to its being too late to be useful: By the time it matured, packet switching implemented directly over SONET/SDH was already a better alternative, and the option of removing SONET/SDH entirely and using hop-by-hop routing entirely under the packet switching layer was also a realistic possibility due to the advancement of Internet protocol (IP) routing technology.

Mid-1990s: The Internet Enters the Public Domain

In the mid-1990s, IP networks experienced tremendous growth due to investments from service providers. Many of these providers would later go bankrupt during the dot-com bubble burst in 2000, but the infrastructure they deployed was extremely useful for the development of the Internet. Services such as voice that were traditionally operated over SONET/SDH started to appear over IP. IP providers had experience with nonsynchronous LAN networks that were very efficient, such as Ethernet. IP routers in the MAN and WAN arenas often had to rely on synchronous telco systems, but improvements in transmission technologies led to the progressive usage of technologies that were previously limited to LANs, in the MAN arena, and later in the WAN arena. Economies of scale with these transmission technologies, and the Internet bubble, led to the progressive decline of SONET/SDH in favor of packet over fiber. This change ended up being so drastic that the roles of packet switching and circuit switching ended up being reversed: While initially packet switching networks were intended to use the services of the underlying circuit switching networks (e.g., by means of ATM over SONET/SDH), in the 2000s it was the circuit switching networks that ended up using the packet switching networks that formed the main infrastructure of network providers, e.g., by means of multiprotocol label switching (MPLS) over IP. Packet switching networks established themselves as a fundamental technology across the entire range of communication networks: From the access segment (e.g., asymmetric digital subscriber line (ADSL)) to LANs (mainly with the dominance of Ethernet), to the MAN and WAN segments.

WDM: A Key Technique to Increase Capacity

Optical transmission technology was the only reasonable choice for high-bit-rate, long-distance transmission due to its advantages in terms of available bandwidth and noise resistance. WDM technology, which allowed the creation of multiple parallel links in a single optical fiber, soon allowed a further increase in the available bandwidth of each fiber. Research into WDM technology evolved fast, and by the 1990s, erbium-doped fiber amplifiers (EDFAs) became a key element of submarine transmission systems, since they could replace batteries of O/E/O regenerators needed for the successful operation of long-distance links. As a result, in the mid-1990s, WDM systems were being deployed for both long-haul terrestrial transmission systems and submarine transmission systems. These WDM networks allowed the easy creation of independent networks inside a global network of nodes, which were then leased to different network providers to implement their services in such a way that the traffic from one did not impact the traffic from the others.

Early 2000s: OADMs Introduce Optical Switching

Wherever packet data transmission was optical, it made sense to also use optical technology inside the nodes that forwarded the traffic. Available packet switching technology was however only electronic, not optical. Electronic switching needs complex systems to transform traffic from optical to electronic and vice versa, and are power hungry. Thus, research into the development of optical packet switches was intense; however, the emulation of electronic systems, such as memories and processors, in the optical domain turned out to be very difficult. In the early 2000s, the complexity of high-connectivity optical packet switches pushed the scientific community to explore simpler systems such as add/drop multiplexers (ADMs), which although not as versatile as optical packet switches, offered a good trade-off between performance and practical implementation. In this way, the 2000s were the era of ADMs, which operate at the wavelength granularity, and can provide connectivity directly in the optical domain and in a quasistatic way.

2010s: The Arrival of Cloud Computing

In the 2010s, the *cloud* concept appeared, leading to the concentration of huge quantities of information in a network of datacenters, with a corresponding pattern of communication between those datacenters and the end users of their data. These end users were often located in Internet service provider (ISP) access networks. The main objective of cloud computing was to limit dependencies on local storage and local processing, greatly facilitating the development and usage of applications by end users, who did not have to overly worry about their local computers when using high-performance and/or data-intensive applications. Although the cloud creates new problems, including important security-related ones, it has been a powerful driver of the development of many current popular services of the Internet (search, social networks, storage, and virtualization) and has led to a huge growth in the datacenter sector. These applications are data intensive and usually interact with humans, and thus are also latency sensitive. As a result, optical technology, which was previously geared towards MAN/WAN but not LAN scenarios, made significant inroads into the LAN arena. Economies of scale followed, and nowadays, optical LAN technology is a serious contender with electronic LANs in most deployments. Thus, the main use case for electronic transmission in wired networks is the last hop to user computers in LANs. Aggregation links in LANs, MANs, and WANs have thus embraced optical packet technology for transmission. Optical packet switching is the last issue remaining for an arguably all-optical end-to-end path in the global, nonwireless Internet.

20.3.2 Optical Packet Switching: Taxonomy and Research Directions

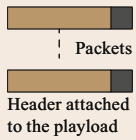
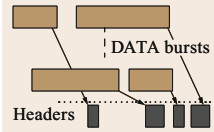
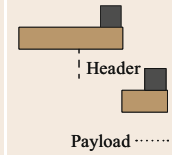
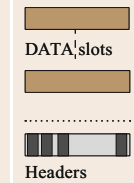
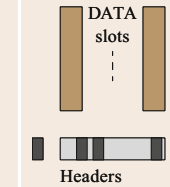
Optical Packet Switching: Taxonomy

Optical packet switching (OPS) describes a large domain from components to systems and networks. By optical technology in ICT, we understand all technology exploiting light for different applications: transmission, processing, and switching. The packet dimension can be assimilated to a virtual circuit or to a nonconnected technology such as the IP. However, in all cases, a packet means a small quantity of data limited in number of bits or in time, and interleaved with other data that could have different sources and destinations. Switching is the technology used to switch physically the information from one input port to one output port of a matrix of connections. So, we could define OPS technology as a technology able to switch a quantity of data limited in number of bits or in time, directly in the optical domain. In the terminology of OPS technologies, *packet* describes a technology that must switch in the range of ns or below, and not ms or higher.

Although there is no ambiguity regarding the terms *switching* and *optical,* one must define more precisely what a packet is. In classical packet technologies, there is a limited number of bits: up to 1500 byte for Ethernet, up to 64 kB for IP, and 53 byte for ATM cells. So, generally, a packet is defined by two main fields: the header and the payload. The header contains the information required to switch the packet at packet switches, while the payload includes the data to transport. So, one cannot dissociate the payload of the header, since it forms the packet structure. In optics, we may have different structures for a packet, since we want in some case to separate the header from the payload. This is particularly required when the payload has to cross a node transparently and we do not want to have to demodulate the whole packet structure to read only the header. The decorrelation of these two fields is particularly interesting when we want to decorrelate the bit rate of the payload from the bit rate of the header to offer transparency at the bit rate level or at the modulation format level without impacting the scheduling. The different structures of packets (Table 20.1) are now described:

- *Structure 1:* Payloads including native packets with in-band headers. However, the demodulation in the electronic domain of both the payload and header has reduced interest in this structure for optical switching. This pushed the community to study optical label switching (OLS) technology in the 2000s. If one does not want to demodulate the payload, the header must be processed directly in optics. This last technique, still under investigation, creates pressure in terms of optical processing and indirectly on the maturity of optical integration.
- *Structure 2:* Payloads including a burst of native packets and an out-of-band header sent in advance of the payload. This was the case for optical burst switching (OBS) technology [20.2]. The objective of this technology was to anticipate contention in the switches by sending the headers in advance to prereserve resources and solve the contention when the payloads arrived. This technique has been widely investigated in the scientific community, where the main objective was to eliminate optical buffers from optical switches while offering acceptable burst loss ratios.
- *Structure 3:* Payloads including a burst of native packets and out-of-band headers transported on a separate channel but synchronously to the payloads. The packet is then split into two physical channels: the channels transporting the payloads,

Table 20.1 Different structures for optical packets

	OPS optical packet switching	OBS optical burst switching	OLS optical label switching	OSS optical slot switching	WSS WDM slot switching
Concept					
Pros	Performance with fixed sizes	Separation of the data plane	Performance with fixed sizes	Performance/efficiency Real transparent concept	Performance/efficiency Any to any Simple system
Cons	Need for a lot of RX to detect the headers	Management of voids Synchro between DP and HP	Optical label processing limited	Need for tunable lasers and tunable filters	Need WDM TRX

and the channel transporting the headers. This technique offers the advantage, through full decorrelation of the headers and payloads, of making the system independent of the bit rate or the modulation format of the payload, while maintaining the bit rate of the header channel at a low value. The payload can then cross the node transparently while being controlled using a control channel that transports headers that are synchronously demodulated to route the payloads dynamically. In the last versions investigated, the optical slot switching (OSS) technique was proposed to minimize the insertion delay while guaranteeing fairness between nodes. A slot rhythm is mandatory to guarantee synchronization between the payloads and headers.

- *Structure 4:* This packet structure is very close to the previous one. The only difference is that the payload can be encoded over different wavelengths. We call it WDM slot. One advantage of introducing the WDM dimension into the packet structure is that the bit rate can be accelerated in the optical domain while keeping the frequency at low values in the electronic domain. A second advantage is that the system managing the pass-through traffic can be dramatically simplified. A third advantage is that broadcasting techniques can be implemented very easily, since a network of N optical packets (in the case of POADMs (packet optical add/drop multiplexer)) is transformed into a network managing only one packet but encoded over N wavelengths. Since this technique impacts the spectral efficiency, a WDM slot is more adapted to networks close to the access part, since the fiber is large enough to transport the required capacity. However, this technology could be also envisaged in the backbone, since its lower spectral efficiency optimization could be compensated by a gain in network efficiency thanks to the adoption of packet granularity.

Key Components Investigated for OPS Systems in the 1990s

As described above, in the early 1990s, OPS was foreseen as a promising technology for future switching systems. Beyond classical optical components such as optical couplers or optical connectors, six components (Fig. 20.7) were identified as key components of OPS systems: optical switches, fast tunable lasers, fast tunable filters, optical gates, fiber delay lines, and burst-mode receivers:

- The optical switch represents the central and key element of any packet switching system. Different technologies were investigated. One of them is based on the use of semiconductor optical amplifiers (SOAs) to switch in a very short time (typically a few 100 ps for the SOA) at the packet rhythm. A switch then includes optical couplers and SOAs integrated on one chip. Other types of switches were investigated, based on directive couplers. In all cases, a switch must be able to switch a packet in a short time (a few ns), with significant isolation with respect to the paths in off-state and with minimum loss (so as not to constrain the optical signal-to-noise ratio).
- Fast tunable lasers [20.51] were identified as potential components to switch data in space switches by exploiting the tunability of lasers and tunable filters. The objective is then to broadcast all the data to the output ports, but select only the correct output port for the data, a technique called *broadcast and select*. The selection can be done using an active switch

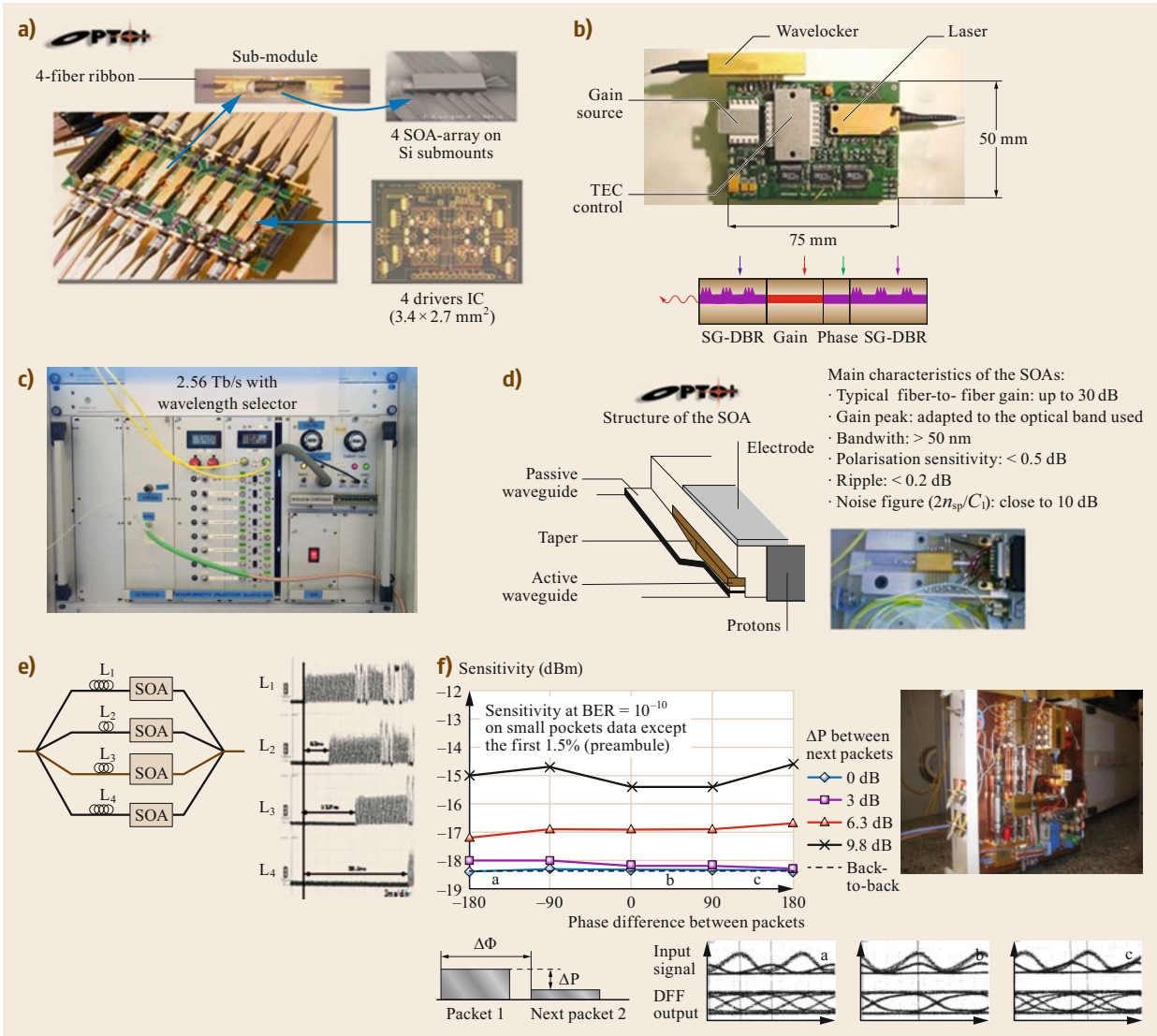


Fig. 20.7 **(a)** Optical switch (SOA array for the realization of a space switch) (N. Sahri et al., OFC'01,PDP OPTO+/Alcatel), **(b)** fast tunable laser, based on SGDBR structure, (Bhardwaj et al. IEEE PTL 2003) **(c)** rack-mounted fast wavelength selector embedding 16 SOAs, **(d)** optical gate for fast optical switching, **(e)** fiber delay lines selected by one SOA, **(f)** burst mode receiver based on a passive clock recovery device for fast CK recovery

including optical gates, with tunable filters at each output port.

- Fast tunable filters were required to support the broadcast-and-select mechanism. When broadcasting data through an optical switch, a fast tunable filter is mandatory to select one packet at a specific wavelength. One possible implementation of a fast tunable filter involves the integration of one optical demultiplexer, an SOA gate array, and one optical multiplexer. This device is called a wavelength selector [20.52].
- Optical gates [20.53], used as switches, can allow data to cross the gate or can block the data. This device, in a space switching architecture, can be the key element for a pure space switch. The key advantages of this component are its short switching time (a few 100 ps for the rise and fall times at 10–90%), high on/off ratio (generally > 40 dB), good stability, and simple realization, making it a serious candidate for use in many OPS systems.
- Fiber delay lines [20.54] can be used as buffers to solve the contention between two packets aimed at

the same output port at the same time. If deflected to the fiber delay line, packets are thus delayed to be serialized with respect to the other packet. The contention can then be solved. However, the buffering capacity of this technique is limited, since the depth of the buffer is directly linked to the number of fiber delay lines and their distribution in the structure of the buffer.

- Burst-mode receivers are required to demodulate optical packets/bursts at the reception side. Burst-mode receivers for amplified systems must be optimized to tolerate a smaller optical signal-to-noise ratio (OSNR). Therefore, the operating points of a burst-mode receiver for OPS systems are not the same as for passive optical network (PON) systems. The clock recovery must be fast enough so as not to constrain the preamble of optical packets too much. Finally, the burst-mode receiver must be tolerant to fast packet power variations, as in PON systems, but with a smaller amplitude to preserve the OSNR.

Towards an All-Optical Vision with New Functions

The all-optical vision was a target of the early 1990s in different research laboratories. To reach this goal, the OPS community explored different key subsystems including:

- OPS fabrics
- Optical buffering
- Optical reamplification–reshaping–retiming (3R) regeneration devices
- Optical synchronization subsystems.

OPS Fabrics. Different architectures for OPS systems were explored. Some of them were based on pure space switch architectures, while some were based on broadcast-and-select techniques exploiting tunable lasers and tunable filters. Some examples of realization are described later on.

Optical Buffering. Different structures for optical buffers were investigated. However, this technique was abandoned in the late 1990s, even if some actions to identify how to realize optical memories restarted in the late 2010s. A set of fiber delay lines can hardly compete with an electronic memory offering the four basic functions of write, store, cancel, and read. Even in the case of fixed-duration optical packets, where it was demonstrated that the packet loss rate was acceptable, this technology was not retained for the design of commercial switching systems.

Optical 3R Regeneration. Optical 3R regenerators were studied in the mid-1990s to overcome the power budget limits of high-capacity packet switches. Optical 3R regenerators based on active elements were investigated with O/E/O or O/O/O clock recovery systems. Different structures for all-optical 3R regenerators were investigated. One, based on cross-gain modulation semiconductor optical amplifiers and Mach–Zehnder interferometers with O/E clock recovery, gave excellent performance. The key point was the integration of the device, despite initiatives such as the LASOR project at UCSB. In 1997, a first demonstration was carried out, with a cascade of several 3R OOO regenerators at 10 Gb/s over 20 000 km up to 800 000 km. Later, the possibility of exploiting all-optical clock recovery using self-pulsating lasers was demonstrated. Although several demonstrations of the potential of optical 3R regenerators over quasi-unlimited distances were achieved, this device was abandoned in the mid-2000s, with the emergence of new modulation formats and forward error correction (FEC) systems.

Optical Synchronization. Optical synchronization appeared in the late 1990s as relevant to synchronous switching of optical packets, as performed in electronic switches to optimize the performance of the switch. To synchronize packet streams, two types of optical synchronizers were investigated: a coarse synchronizer to realign the packets in the guard band, and a finer synchronizer guaranteeing that the distance between consecutive packets was respected within a certain tolerance to avoid overlap of consecutive packets. Thus, the first synchronizer acts at the flow level, whereas the second synchronizer operates at the packet level. This structure based on a set of fiber delay lines was investigated in the IST DAVID project.

Revisiting the Concepts: Towards Hybrid Systems and Networks

In the early 2000s, the European IST DAVID project proposed to compare an OPS network with an optoelectronic packet switching network. The difference between the two was that, in the hybrid model, OPS technology was limited to the space switching and 3R regenerators. The switching capacity targeted was 2.56 Tb/s. It was demonstrated that the structure of the wavelength selector (the end device of the optical switch) could be different if one implemented an all-optical packet network or an optoelectronic packet switch (an optical packet switch surrounded by O/E/O transceivers). For the all-optical node, a polarization converter is needed in the wavelength selector to reaffect only one polarization state to all the consecutive

packets, before OOO regeneration. With an O/E 3R regenerator, conventional wavelength selectors can be adopted.

To address the problem of buffering, add/drop multiplexer architectures were investigated in parallel. These systems are close to an electronic technology, since with a contention order limited to 2 and an add port that has an electronic buffer, contention can be properly solved as in a fully electronic systems. Add/drop multiplexer architectures based on tunable lasers and wavelength selectors were first investigated. In the late 2010s, the ECOFRAME project focused only on packet optical add/drop multiplexers (POADMs) using optical slot switching (OSS) technology. Interesting results were obtained, demonstrating the potential of this new technology.

Towards Simplicity

The necessity for simple systems goes hand in hand with the identification of opportunities for the access and metro network segments, reinforced today by the support for cloud computing. In parallel, the lack of commercial fast tunable lasers and the difficulties in producing effective wavelength selectors have reoriented research towards simpler devices. The new goal then became how to make the system simpler, based on available commercial components to propose low-cost and high-energy-efficient nodes adapted to a 5G strategy and to support the emergence of new vertical markets (Industry 4.0, IoT (Internet of Things), Smart Cities, etc.). It was then mandatory to rethink the system design to tackle this new objective. The energy is also a key issue to solve, leading to the need to focus on eco-designed systems to target new markets in this perspective. Figure 20.8 summarizes the evolution of OPS technologies, from complexity to simplicity. Note that optical flow switching (OFS) was envisaged in the early 2000s as a potential alternative to OPS technology. OFS proposes the use of long packets (in the range of 100 ms) to support switching times for the nodes in the range of ms, thereby creating opportunities for cross-connects in switching fabrics. However, these systems showed some limitations in terms of network efficiency without any satisfactory ratio strategies, thus dramatically impacting the latency of each node.

20.3.3 Optical Switching Technologies by Network Segment

What Optical Technology in What Network Segment?

OPS has been identified as a potential technology to provide a solution for future packet systems and networks. However, the migration from circuit to packet is not obvious due to several physical barriers. Therefore, we need to come back to simple systems, in network segments where this technology could be competitive or provide new features. The following briefly describes the directions observed and the opportunities:

- The first network segment that paid attention to optical packets was optical access networks. Optical packets are generated at the optical network unit (ONU) level, then sent synchronously in a tree topology. Contention resolution is solved by adopt-

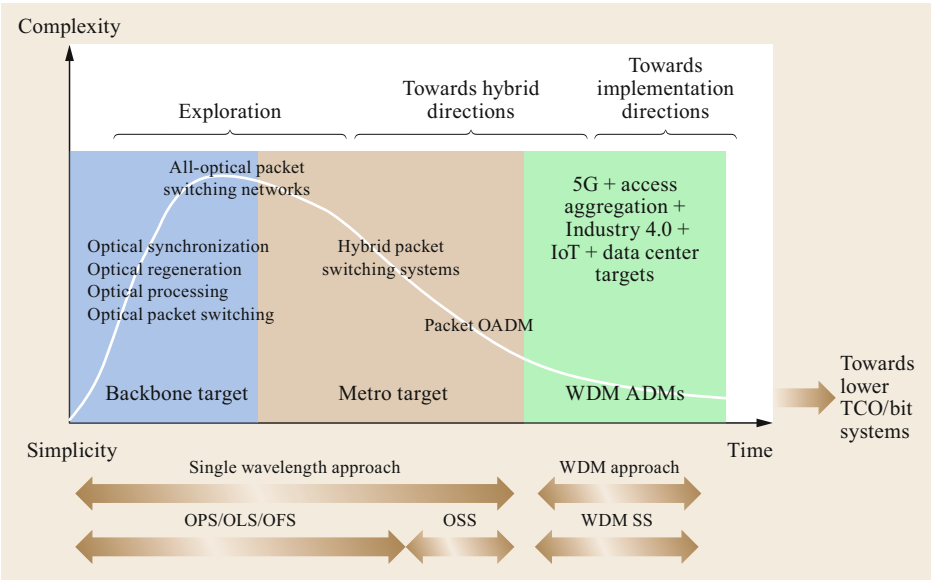


Fig. 20.8 Evolution path of OPS technology: from complexity to simplicity

ing a centralized synchronization in order to serialize the optical packets that will be transported on the same optical fiber interconnected to the optical line terminal (OLT). Japan was the first to propose an Ethernet passive optical network (EPON) system to their subscribers, in June 2001. The EPON proposed offered 1 Gb/s for upstream traffic and 1 Gb/s for downstream traffic. PONs were the first to propose optical packet technology in the infrastructure of the network with burst-mode transmitters at the ONU side, optical packets including a physical header to support burst reception, and optical burst receivers at the OLT side to be able to receive asynchronous packet streams potentially suffering from power discrepancies.

- In the metro arena, one can distinguish access aggregation and the metro core or regional metro. Access aggregation, whose role is to aggregate traffic from several PON systems, is identified as the first network segment where OPS technology could be introduced. The challenges for this network segment are cost, energy consumption, and performance. Innovative approaches have thus been investigated to create market opportunities through small incremental steps when compared with PON technology. This case will be particularly discussed in the next paragraphs. The metro core is probably more challenging, since there is a need for a mesh topology, imposing more complex nodes. However, if a generic box operating in the access aggregation part had the potential for easy scalability while supporting high bit rates and complex modulation formats, a new technology based on OPS technologies exploiting OADMs could be envisaged.
- In the 5G arena, there could be interest in the use of OPS technologies in fronthaul/backhaul/X-haul solutions. The need for ultralow latencies in these network segments positions OPS technology as a potential candidate through the exploitation of optical bypasses in a packet mode. Machine-to-machine (M2M) traffic, part of the Internet of Things (IoT) and requiring multiconnectivity networks enabled by a packet technology, could also be a catalyst for the adoption of OPS.
- Local area networks (LANs) could be a use case where OPS technologies could provide some advantages. Passive optical LANs (POLs) are already envisaged in LANs. With a small step in terms of technology, focusing on components coming from access networks, OPS technology could offer new solutions in line with the requirements of LANs (low cost, high network efficiency, and multiconnectivity).
- In core networks, opportunities for new OPS technology are becoming less and less convincing in areas where the stability of traffic allows use of circuit switching. However, OPS technology could offer new solutions for the interconnection of line cards of high-capacity routers, making they more compact and reliable, and with fewer optical fibers. Here, the potential of optical technology to switch in an ultrahigh-capacity mode by combining time, wavelength, phase, and amplitude modulation could be exploited in multimode and multicore fibers inside a small switch. One example of such a backplane is described in the next paragraphs.
- In data centers, OPS technology can find several applications. Firstly, to balance the load in edge routers of load balancers, an optical switch can be used with a simple scheduler to distribute the traffic in different line cards of the edge router. This is a good typical example application, since contention is automatically solved by the policy adopted for the scheduler. A second application case is probably the interconnection of disaggregated servers, using small optical switches to realize dynamic interconnection. Finally, OPS technology could be used for the interconnection of top-of-rack (ToR) switches.

Some Illustrations of OPS Testbeds

Optical ATM Switch of the RACE 2039 ATMOS Project. The European RACE 2039 asynchronous-transfer-mode optical switching (ATMOS) project (1991–1995) targeted the realization of an ATM optical switch to explore the potential of optical technologies. The optical packet was an optical cell (close to 170 ns at 2.5 Gb/s) including a guard band of a few nanoseconds to absorb the switching time of the active elements of the switch. In September 1993 at Montreux, Switzerland, the project demonstrated a 4×4, 2.5 Gb/s per input ATM switch, called a fiber delay line switch, including four tunable distributed Bragg reflector (DBR) lasers capable of switching four wavelengths spaced over 0.6 nm using 16 SOA gates, giving access to four fiber delay lines, with four fixed filters able to select channels with a 3-dB bandwidth of 0.3 nm and rejection above 30 dB at 0.6 nm. The switch was managed by an electronic scheduler to solve real-time contention. The demonstrator was shown during the ECOC 1993 exhibition (Fig. 20.9; [20.55]) over 4 days error free. This was the first public rack-mounted ATM optical packet switch demonstration including fast tunability, scheduling, optical buffering, and fixed optical filtering, paving the way for other investigations. At the end of the project, another important result was the first all-

Fig. 20.9 ATM optical switch presented at ECOC 1993, Montreux, Switzerland

optical wavelength conversion performed at 20 Gb/s by DTU-COM, a partner of the project. This key result opened the path to all-optical regeneration, which was then exploited in the KEOPS project.

The KEOPS Project. The European KEOPS project (1995–1998) aimed to go a step further by studying the feasibility of an optical packet network. A new packet format was proposed by CNET, different from the ATM cell of the ATMOS project since it was fixed in time and not in number of bits. The optical packet included a payload and a guard band to absorb the switching regime of the optical components (tunable lasers and optical gates). Three demonstrators were targeted: one by France Telecom focused on an optical packet switch, one by Alcatel-Alstom Research targeting an all-optical network, and one by CSELT targeting a ring network. The following key subsystems were realized:

- An optical 3R regenerator operating at 10G with on–off keying (OOK) modulation formats. It was demonstrated in the laboratory first over 20 000 km then up to 800 000 km, crossing more than 1000 3R 10G regenerators error free, with a limited sensitivity penalty (less than 2 dB). This was the first demonstration of the potential of active optical regeneration over long distances at 10 Gb/s. The 3R regenerator was a key element to test the cascade of optical packet switches in a network configuration without O/E/O conversion in the core of the network [20.56].
- An optical packet switch, offering 160 Gb/s of switching capacity with a 16×16 switch, operating at 10 Gb/s per input port. The architecture was quite different from the RACE 2039 ATMOS switch, being called a broadcast-and-select switch. The architecture included wavelength selectors (equivalent to fast tunable filters) at the end of the switch to select the packet coming from one of the inputs. The architecture was demonstrated to be feasible with small sensitivity penalties, but the power budget was completely exploited by the switch architecture. Thus, 3R regenerators were identified as mandatory elements to cascade several optical switches.

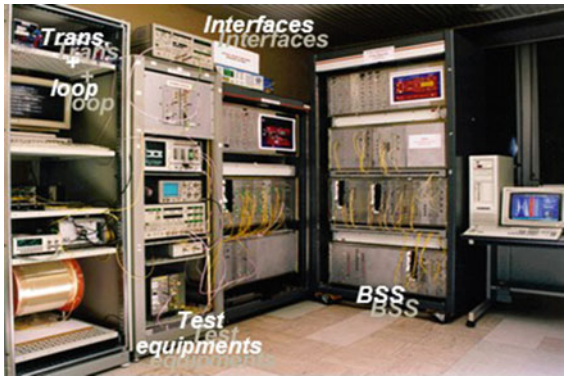


Fig. 20.10 Demonstrator of the European KEOPS project

Figure 20.10 [20.57] shows the final demonstrator of a network, to validate the cascade of 40 network sections, including an optical regenerator.

Towards Optical Slotted Systems

The KEOPS project demonstrated the need for new packet formats, larger than an ATM cell, to target high-capacity switching systems. Optical packets with a fixed duration were adopted to optimize the contention issues. It was effectively demonstrated that the adoption of variable optical packets could create some unfairness in the cascade of nodes: The first nodes crossed have access to an acceptable bandwidth, but the last nodes suffer from small holes, thus making the insertion of long packets impossible. The fixed-duration optical packet has the advantage of guaranteeing fairness in terms of bandwidth access to all the nodes of a network.

The IST DAVID Project. The IST DAVID project (2000–2003) explored different optical packet switching networks around five technologies:

- An optical packet switch in two versions: with O/E/O regenerators and with all-optical 3R regenerators
- A packet optical add/drop multiplexer (POADM) technology based on fast tunable lasers and wavelength selectors
- A packet optical add/drop multiplexer technology with a dual bus

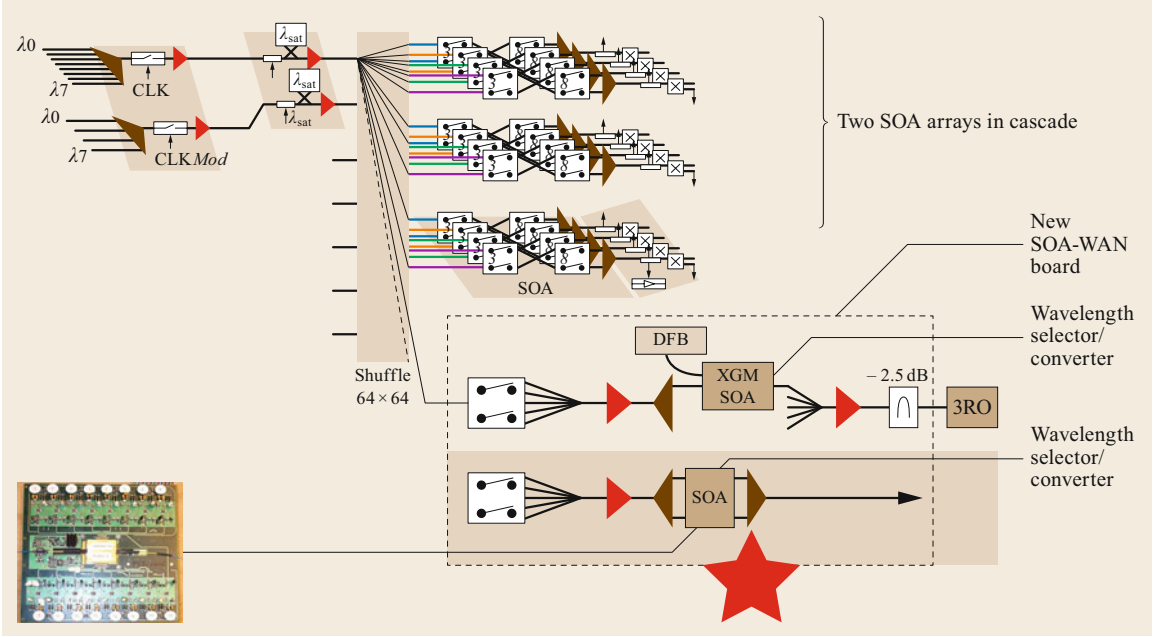


Fig. 20.11 Optical packet switch architecture, offering two output interfaces: one wavelength selector compatible with O/E/O regeneration, and a wavelength selector/converter to interconnect with an OOO 3R regenerator

- All-optical 3R regenerators with enhanced characteristics
- Optical synchronizers, to synchronize flows of packets and remove packet jitter accumulation.

The optical packet switch demonstrator (Fig. 20.11) in this project targeted a switching capacity of 2.56 Tb/s. The optical switching architecture adopted was the one investigated in the KEOPS project with no fiber delay lines in the core of the architecture. To solve the contention, a recirculating buffer was adopted, based on either fiber delay lines or an O/E/O buffer. In this switch, a saturated wavelength transporting the modulation of the WDM signal but with inverted polarity was used. This saturated wavelength was combined with the WDM signal to stabilize the gain of the SOAs used in the space switch. The gain was then dependent on a constant-amplitude signal (the sum of the data and the inverted data) to cancel the cross-gain modulation effect that can occur when injecting a signal closer to the saturation area of the SOA. The saturated signal was then rejected by the optical multiplexer at the output of the switch. Note that the switch includes two stages of SOAs. The first stage of SOAs selects the WDM input, whereas the second stage of SOAs selects the wavelength of the preselected WDM signal. We observe in Fig. 20.11 that the wavelength selector is also a polarization converter in order to guarantee a constant polarization state before the all-optical 3R regenerator.

In parallel, a packet optical add/drop multiplexer (POADM) was also investigated, based on fast tunable lasers and fast wavelength selectors. The project

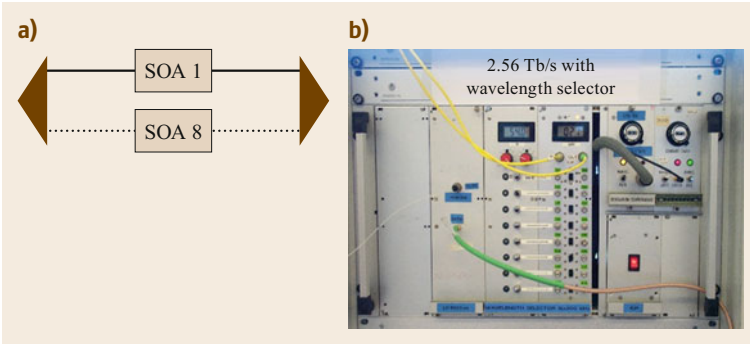


Fig. 20.12 **(a)** The 16-channel InP-based wavelength selector structure, **(b)** the rack-mounted wavelength selector including one optical demultiplexer, 16 SOAs, and one optical multiplexer

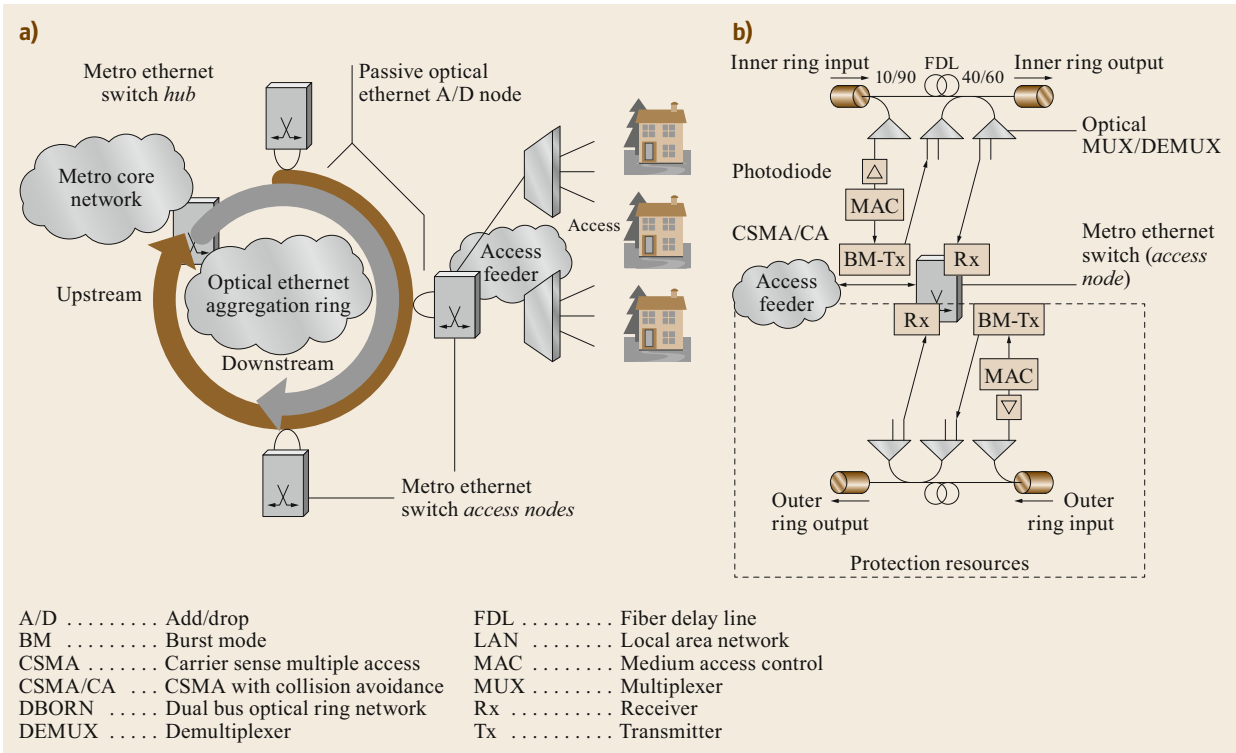


Fig. 20.13a,b Dual-bus optical ring network (DBORN). **(a)** DBORN ring topology deployed in metro access network, **(b)** passive optical Ethernet A/D node

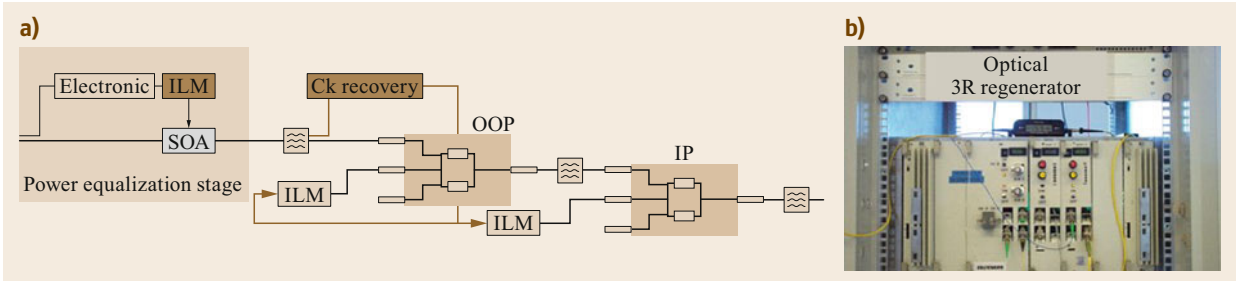


Fig. 20.14 **(a)** MZI-based all-optical 3R regenerator structure, **(b)** rack-mounted demonstrator

tested a wavelength selector (16 channels with spacing of 200 GHz) coming from the FLORA project (Fig. 20.12). The wavelength selector exhibited polarization sensitivity of less than 0.5 dB, a spectral flatness close to 1 dB, and a fiber-to-fiber gain close to a few dB.

A dual-bus optical ring network (Fig. 20.13; [20.58]) was also proposed as an alternative to the POADM technology, the objective being to simplify the technology of the node and to adopt a dual bus, as in PON systems, viz. one for the upstream traffic and one for the downstream traffic. The only difference with respect to PON systems was the adoption of a ring topology. Passive add/drop multiplexers with no embedded optical amplification were adopted. The upstream was characterized by burst transmission, whereas the downstream was fully synchronous, like in a PON system. A demonstrator was realized and shown at the end of the project.

To enable the cascade of switches, different optical 3R regenerator structures were investigated with Mach–Zehnder interferometer (MZI) devices. The structure shown in Fig. 20.14 includes: one power equalization stage and a polarization converter using a SOA in a cross-gain modulation scheme, one stage for the NRZ → RZ conversion sampled with a clock recovered using a MZI device and tuned to invert the polarity of

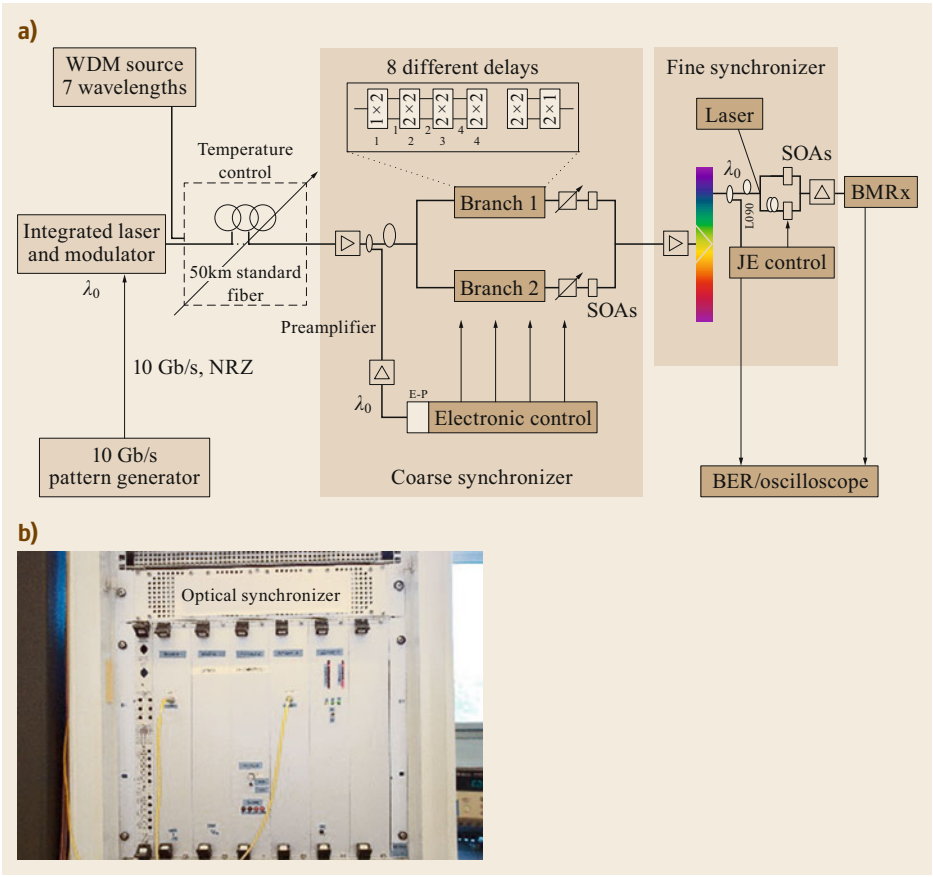


Fig. 20.15 **(a)** Optical synchronizer structure combining two stages: one coarse granularity synchronizer and one fine synchronizer, **(b)** photo of the optical synchronizer demonstrator

the data (OOP (out of phase)), and one final stage using a MZI maintaining the polarity (IP (in phase)). This 3R regenerator included all the functions necessary to regenerate the signal in the phase and amplitude domains. The principle of operation of Fig. 20.14 was as follows: The incoming signal was converted to a new wavelength through a cross-gain modulation scheme. The objective of this first stage was to stabilize the power and to fix the polarization that is imposed by the probe. This conversion also creates an inversion of the polarity. Before entering the MZI, a filter is used to reject the incoming wavelength transporting the data and a coupler extracts a part of the power of the modulated probe, to extract the clock. This recovered clock modulates the probe of the first MZI stage. After the first MZI stage, the data signal is sampled by the clock synchronously with the incoming data. The NRZ signal is then transformed to a RZ signal, but with inverted polarity. This first stage removes the time jitter created by the switching system. The output of the first MZI stage is then connected to a filter to reject the pump and to select only the probe modulated with the clock. The second stage of the MZI, operating in-phase to maintain the polarity of the signal, is used to regenerate the amplitude of the pulses. At the output of the all-optical regenerator, the signal has been retimed, and filtered in the time and amplitude domains.

Two optical synchronizers were realized (Fig. 20.15; [20.59]). The first one targeted the synchronization of the packet flow to align the packets at the input of the optical packet switch. The configuration of one branch is set up in order to have a fixed phase with respect to a reference signal. When a phase misalignment is detected due to a temperature evolution of the fiber, the second branch is then set up to a new delay value to readjust the phase of the data streams. In that case, the switching between branch 1 and branch 2 is made using a SOA. The second stage targets the reduction of the packet jitter to avoid all overlap of packets during the cascade of many nodes. Since the phase variations are quite slow, classical 2×2 optical cross-connects were used in the coarse synchronizer except for the optical switch selecting one of the two paths. For the fine synchronizer, SOAs were used to switch rapidly during the guard band and then be able to realign the phase of any packet in real time.

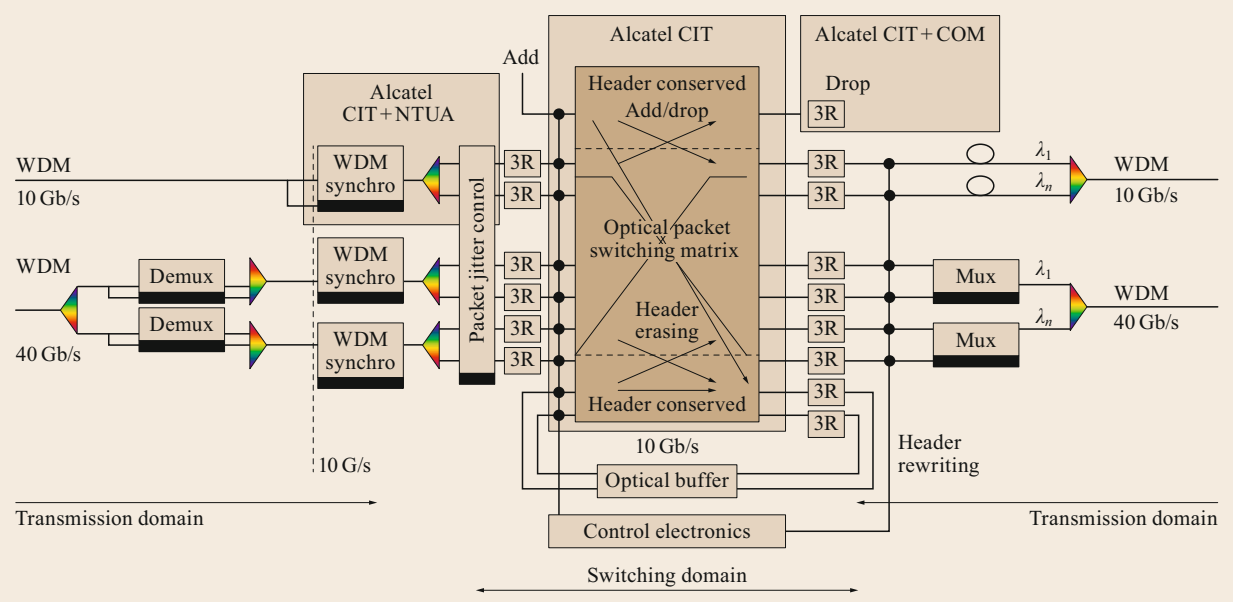


Fig. 20.16 Schematic of the all-optical packet switch, including a photonic switch, an optical synchronizer, and an optical regenerator, used to validate the feasibility of an all-optical packet network

All these elements were assembled, and the performance of a fully equipped packet switch was evaluated (Figs. 20.16 and 20.17; [20.60]). By cascading optical synchronization, optical regeneration, optical space switching, and optical regeneration, it is then possible to build an OPS network.

Although the DAVID project revealed some limits at the optical integration level, through the introduction of new technologies and subsystems, it highlighted a reorientation towards a hybrid approach combining the best of the optical and electronic technologies self-consistently.

The ECOFRAME Project. The ANR ECOFRAME project (2006–2009) [20.61] was launched to study the potential of POADMs. The technology (Fig. 20.18) was based on fast tunable lasers, optical gate arrays for management of pass-through (eliminating the wavelength selector in that case), and a photodiode array for the reception part. It was a trade-off in order to capitalize on the conclusions of the DAVID project. The ANR ECOFRAME project realized a demonstrator including three nodes. Excellent performance was obtained. The demonstrator was then exploited in two different new projects: One project with NTT whose objective was to interconnect two rings with an optical packet switch router, and a European project (ALPHA) whose objective was to build an integrated network including two apartments, two PON systems, and one metro ring network.

Fig. 20.17 Photo of the all-optical packet switch of the DAVID project

NTT Hybrid Packet Router. The hybrid packet router realized by NTT (Fig. 20.19; [20.62]) is a fundamental development, with extensive work focused on the exploitation of the best optical technologies to go beyond the potential of electronic technologies. The switch/router included an optical space switch, all-optical header detection, and a recirculation electronic buffer, having optical shift registers. The switch/router was designed to offer new advantages when compared with fully electronic solutions, in particular a strong reduction of the power consumption. However, more importantly, this team realized an unprecedented goal by proposing new devices and new components for many functions currently realized in electronics but

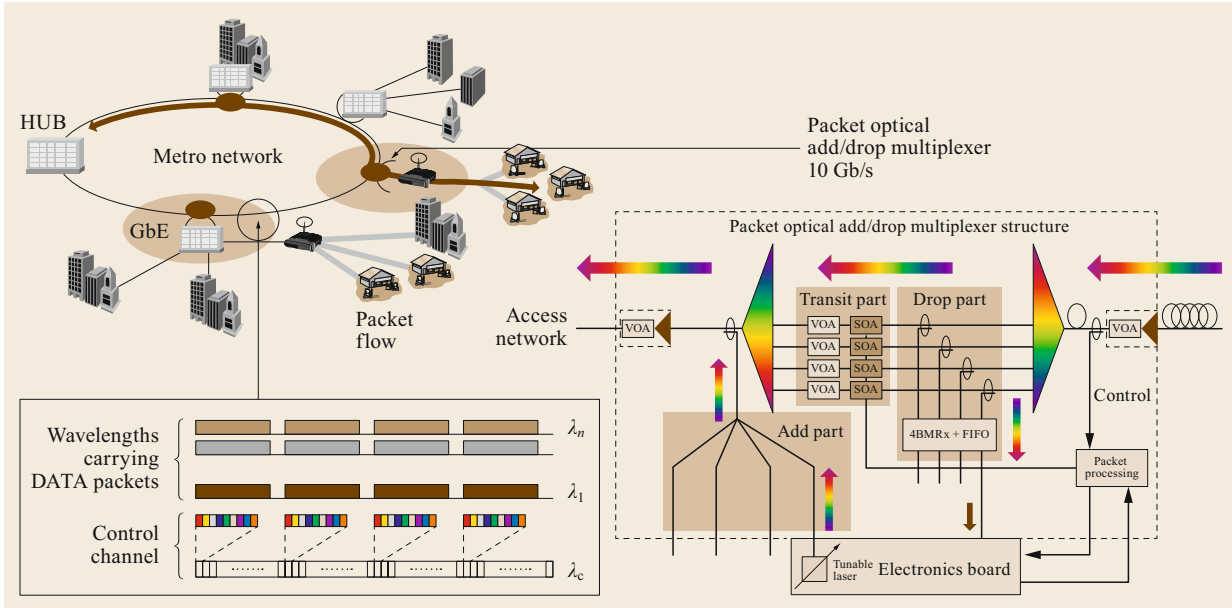


Fig. 20.18 Technology of the ECOFRAME project

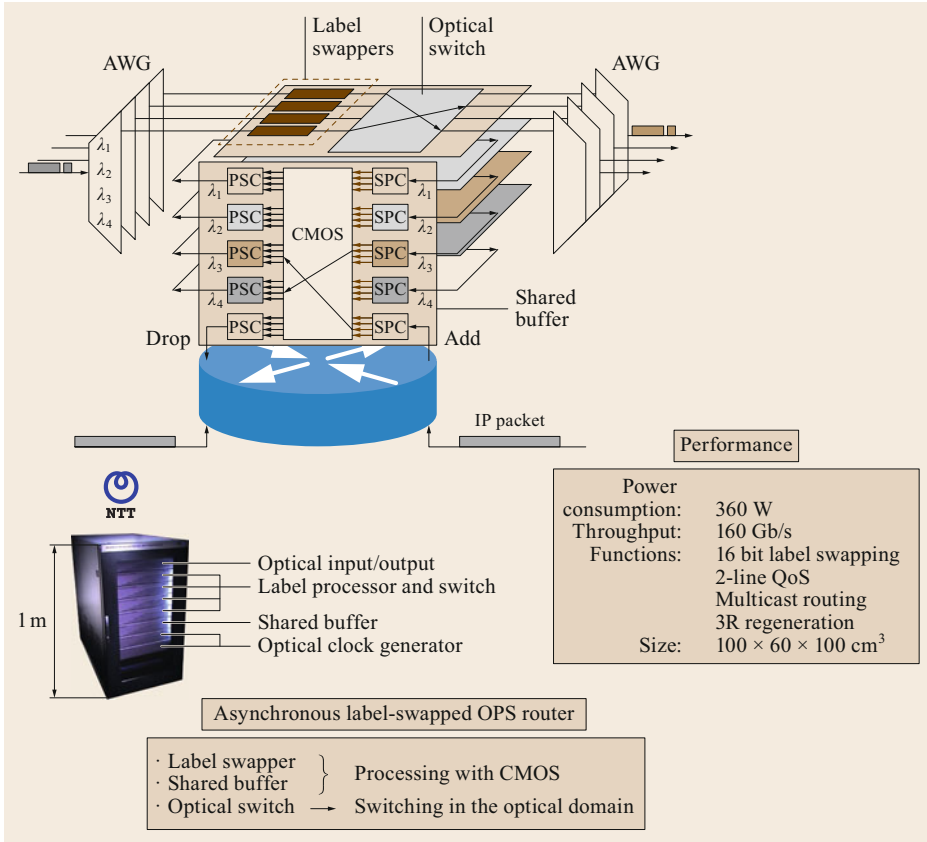


Fig. 20.19 Hybrid optical packet router presented at ECOC 2009 (courtesy of NTT)

Fig. 20.20 Demonstrator at ECOC 2010

identified as limited in capacity or frequency. NTT paved the way for a new generation of components and systems with an ecofriendly design approach.

A collaboration was set up between Bell Labs and NTT in early 2007 to realize the integration of two demonstrators: two optical rings based on the ECOFRAME POADM technology, interconnected with the hybrid packet router of NTT (Fig. 20.19). The integration was done in 2010, and the demonstrator was shown at the ECOC 2010 exhibition with error-free operation over the duration of the event (Fig. 20.20). For this experiment, an MPLS tag was placed in front of the optical slots of the packet rings to be able to process the destination of the optical packets when detected by the NTT switch. It was then possible to switch the optical slots from one ring to the second ring, through the switch/router.

Video streams where transported over slots, through the rings, and through the NTT switch/router to validate an end-to-end technology based on a new optical packet switching technology.

Furthermore, the double ring, rearranged in order to have a protected ring, was tested in the context of the ALPHA project (2007–2011) [20.63]. This test involved a complete network including two home networks, a WDM TDM PON system, a point-to-point PON system, and a double-ring POADM network (Fig. 20.21). An application server was implemented in the metro to provide services at homes. Once again, a multi-bit-rate, multimodulation format was used, on top of an access plus metro network, through real-time applications and the transport of a radio-over-fiber (RoF) technology on the WDM TDM PON system. This last experiment demonstrated the robustness of the technology.

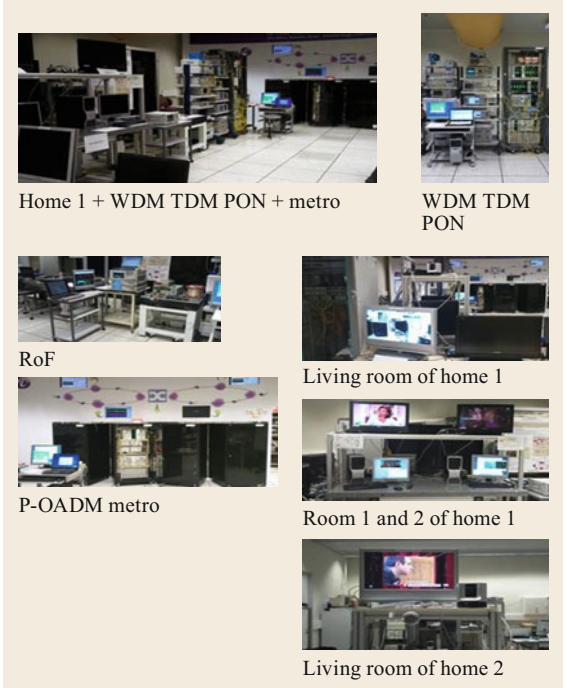


Fig. 20.21 Demonstrator of the European ALPHA project in 2011

20.3.4 WDM Switching Systems

Adopted in transmission systems, exploitation of the WDM dimension was rapidly successful, with EDFAs being able to amplify a WDM comb. The WDM dimension was not really identified as a key dimension to simplify the switching system design, even if some attempts were made in this respect in the mid-2000s using waveband OADMs. Professor Ken-Ichi Sato was the first to propose WDM OADM to simplify system design. Furthermore, multigranularity OADMs were also proposed, combining wavelength cross-connects and waveband cross-connects (Fig. 20.22; [20.64]).

The French RNRT ROM-EO project launched in 2003 with the objective of identifying the right directions to design a high-capacity metro network, reaching a similar conclusion. The WDM dimension has to be exploited to reduce the connectivity of optical switches and thereby relax the constraints on optical integration.

During the same period, NICT also proposed multicolor packets (Fig. 20.23; [20.65]) to increase the capacity of packet switching systems while preserving the propagation distance. This was a fundamental driver towards making these OPS systems more efficient.

The ANR N-GREEN project (2015–2018) then targeted a new approach for low-cost and low-power optical technologies. The ANR N-GREEN project proposed to explore an orthogonal direction when compared with the Flexgrid approach, viz. simplification of the transceivers and of the ADM structure to make the system cheap and energy efficient. The second idea was

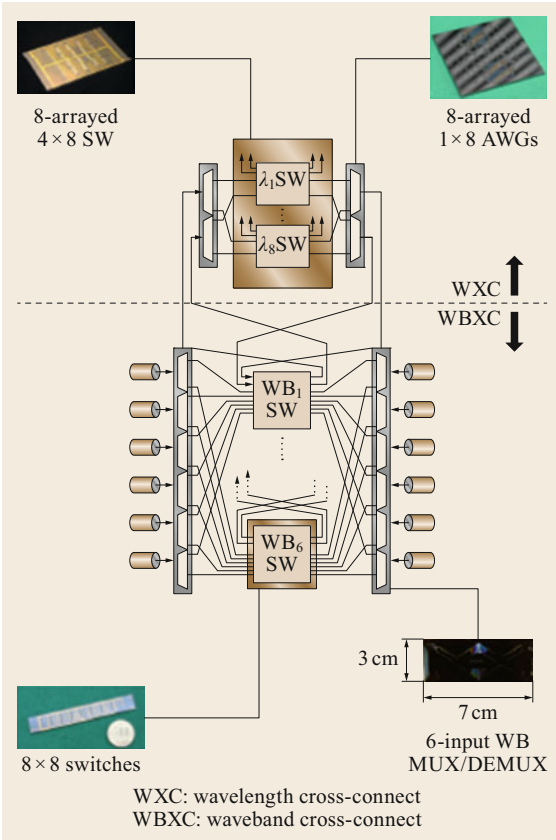


Fig. 20.22 Multigranular OADM exploiting waveband switching

to better exploit the WDM dimension (a natural parallel dimension that is well managed by optical technologies) and the optical integration capability (mainly in the transceivers). Thus, the N-GREEN project proposed two main ideas:

- Very simple OADMs through better exploitation of the WDM dimension. Eco-design rules were adopted to make the system very simple, highly scalable, and adapted to the access aggregation market. The objective was also to propose a cheap technology while offering good performance in terms of end-to-end delays. The proposed concept of WDM ADM was investigated for different 5G scenarios, together with access aggregation applications.
- Better exploitation of the WDM dimension through WDM packets in order to design a high-capacity backplane that was compact and simple enough to integrate self-protection capabilities. The switch size was limited to 16×16, in order to target an internal switching capacity close to 1 Pb/s.

20.3.5 Successes and Difficulties

Main Successes in Optical Switching

Reconfigurable optical add/drop multiplexers (R-OADMs) were identified as potential candidates for the optical transport layer in the early 2000s. The catalysts were probably wavelength-selective switches (WSS). An WSS is an integrated optical device including a microelectromechanical system (MEMS) matrix, and optical multiplexers and demultiplexers. Different structures of R-OADM were proposed for the marketplace. The R-OADM paved the way for optical networking based on circuit switching technology. R-OADM technology is implemented worldwide today in telco networks and is a full success. R-OADM technology allows the transport of high bit rates over long distances. R-OADMs represent the first generation of optical products offering both high capacity and rich connectivity. This technology is well adapted for the backbone, where the traffic matrix is quite stable, or for the exchange of data between distant data centers. However, the emergence of big data centers demanding regular reconfigurations of the network created the need for more flexible network concepts.

The FlexGrid concept is considered as the second generation of R-OADM-based networks. To overcome the rigidity of circuit switching technology, it was proposed to add some level of flexibility to the transceivers and to the R-OADMs. At the transponder level, the emphasis is on tunability for the bit rate and for modulation format adaptation, in order to select the correct bit rate at a specific wavelength. At the R-OADM level, the objective is to replace the existing R-OADMs with flexible R-OADMs able to adjust the channel spacing to the injected optical comb.

Main Successes in Optical Packets: Only in PON Systems

PON systems were investigated in the early 1990s. It was only in the late 1990s that PON systems were considered as potential products for the access network. The EPON system was first deployed in Japan, in June 2001. In the late 2000s, more than 10 million subscribers were connected to a fiber. In other countries, this dynamic of deployment was less evident, and we had to wait until the end of the 2000s to see support for fiber deployment everywhere for everybody. In Europe, the gigabyte passive optical network (GPON) system supporting Ethernet and ATM protocols was selected for massive deployment. This deployment allowed burst-mode technology to mature, creating opportunities for new network models. The burst-mode technology used in PON systems comprises

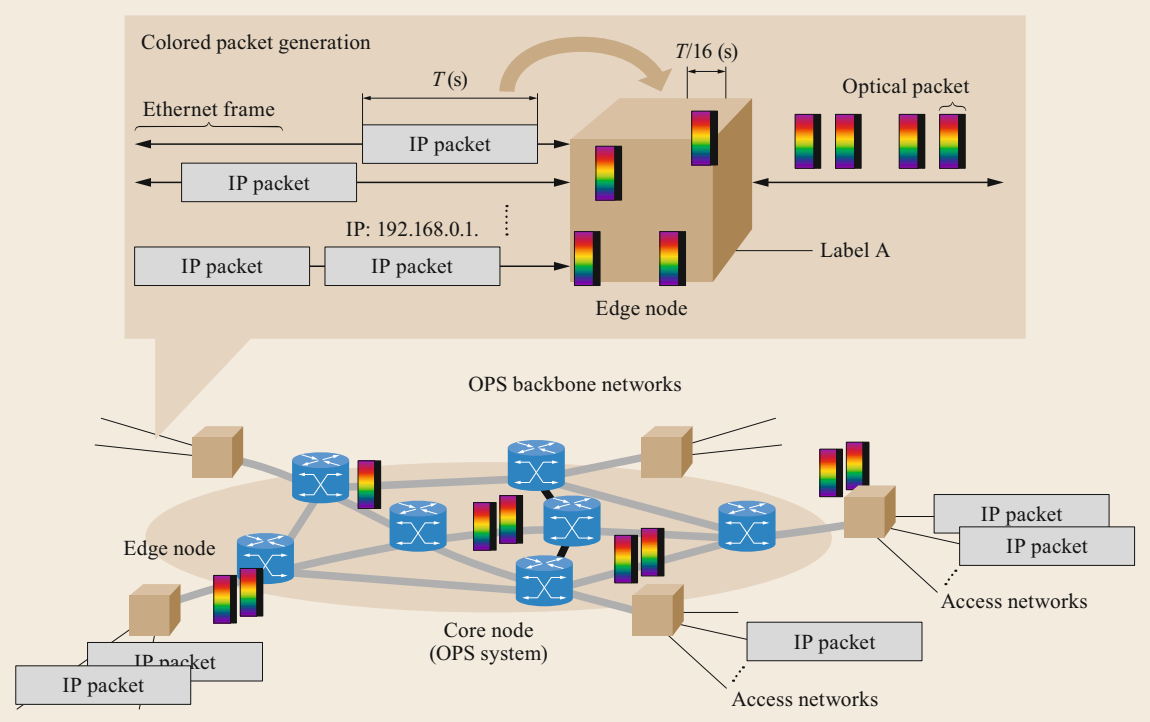


Fig. 20.23 Multicolor packets introduced by NICT (courtesy of NICT)

a burst-mode transmitter and burst-mode receiver. The burst-mode transmitter is designed to support long series of zeros to introduce a guard band in front of each packet. This guard band is mandatory to absorb phase variations between two consecutive packets to take into account static length dispersions between different ONUs connected to the same optical coupler, as well as dynamic length dispersions due to thermal effects impacting the index of the fiber (between 40 and 200 ps/(deg km)).

The WDM dimension, beyond point-to-point systems, has been identified as key to increase the capacity of optical PON systems. The first attempt used WDM PON without exploiting the TDM technique. These WDM PONs were deployed in the mid-2000s in some countries, particularly South Korea. However, the lack of efficiency imposed by the large granularity and without the possibility of multiplexing the traffic of different subscribers on the same transport fiber interconnecting the optical aggregation point (coupler or multiplexer) and the OLT rapidly pushed the scientific community to look at WDM TDM PON systems (Fig. 20.24). These last systems are simply an extension of classical EPON or GPON systems through the WDM dimension, allowing better exploitation of the capacity of the transport fiber. Because of the proximity to end users, cost issues dominated the choice of technology, with many studies demonstrating that WDM TDM PON systems are less expensive than pure WDM PON systems. However, this is still a stacking solution where the WDM dimension is simply used to increase the capacity of transport.

OPS Technologies Not Yet on the Market

One of the main barriers to optical packet technology is the optical component technology. To make this new technology feasible, new components are required, such as fast tunable lasers, optical gates, or more complex devices. In the following we explain the relevant difficulties.

Fast tunable lasers have been identified as key components for OPS systems and networks. However, they are extremely challenging to stabilize. A tunable laser needs to have several sections to guarantee output power, to enable tunability over a large optical bandwidth, to readjust its wavelength position to align it to a spectrum grid, and to gate the output power. This imposes a large number of electronic drivers to control the component. The control of the carrier density variation and the control of thermal effects when injecting currents to select the new wavelength are unresolved issues. Another aspect is the mass production of tunable lasers that emit exactly on the same grid. Finally, statis-

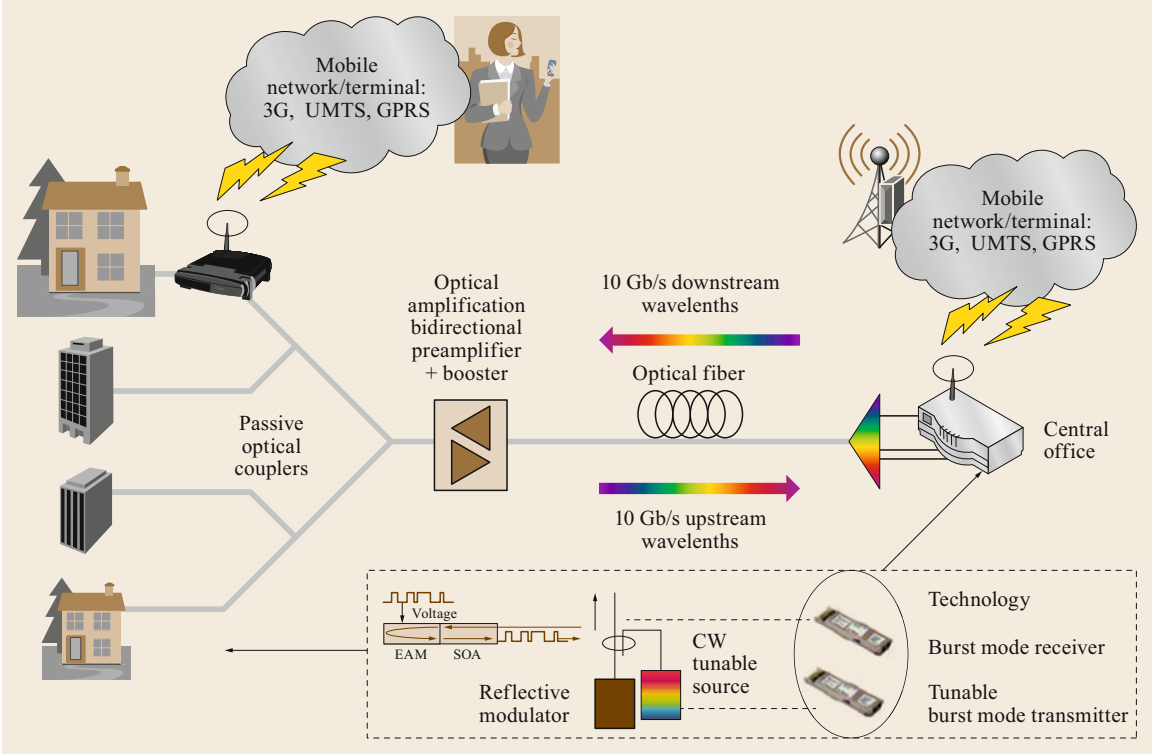


Fig. 20.24 One WDM TDM PON architecture explored in the ALPHA project

tical control, depending on the traffic, can impose low frequencies at the driver level, possibly modifying the generated currents and impacting the wavelength selection. This problem was studied in depth in the 1990s.

OPS technology also proposes an optical bypass in all-optical concepts; for example, to manage optical bypass at the packet granularity, wavelength selectors are required for POADM. Several attempts have been made in the past to realize fast wavelength selectors. One of the most relevant realizations was made by Alcatel Opto+ in the context of the FLORA project and then exploited in the DAVID project. A 16-wavelength selector was realized, including one optical demultiplexer, a 16-gate array, and an optical multiplexer, in InP. The component exhibited a $+2/+3$ dB gain, a flatness of the spectrum within 1 dB, and less than 1 dB of polarization sensibility. Several trials have been carried out more recently using Si photonics technology, but these devices require more functions to support cascades, in particular an automatic gain control per SOA, and variable optical attenuators (VOA) at the input of each SOA to control the injected power.

The packet power dynamic is a fundamental problem to solve. Without any control of the packet power in a cascade of nodes, the quality of signal (QoS) cannot be guaranteed at the end of a network due to the degradation of the optical signal-to-noise ratio (OSNR). Power management raises new constraints on the wavelength selectors, which must include an automatic gain control system, and monitoring systems to monitor the power at each node. Whereas the power dynamic range of PON systems can be handle by burst-mode receivers, amplified optical networks cannot be managed in the same way. Amplified optical networks require distributed optical power controllers and equalizers to guarantee the cascade of optical systems. OPS systems are very sensitive to optical packet power variations between consecutive packets, and require burst-mode receivers with small power dynamics and high OSNR degradation tolerance.

Finally, because of nonlinear effects when injecting a data stream into the saturated area of a SOA, it is important to control the nominal maximum input power. Therefore, a VOA has to be integrated in front of each SOA in the wavelength selection device to fix the maximum input power injected into each SOA at the calibration phase of the optical switching nodes.

20.3.6 Towards Simplicity

Because of the complexity of management of OPS technology, device simplicity is strongly needed in order to

ease its chances in practice. The questions then are how to build simple OPS technology, how to control optical signals from an end-to-end perspective, and how to minimize the optical integration constraints that limit the number of components to integrate. Since the mid-2000s, the scientific community has tried to simplify OPS systems in order to allow the creation of a new generation of systems.

How to Achieve Simplicity?

To make the systems simpler, we need to reexploit the WDM dimension. The WDM dimension can of course increase the transport capacity of a transmission system, but it can also introduce a parallelism in the optical domain that can be exploited to simplify the switching system architectures.

Effectively, if we switch subbands of wavelengths instead of wavelengths, we could reduce by one order of magnitude the complexity of the optical bypass system if the subband is formed from 10 wavelengths. So, by combining WDM switching and OADM structures, we can then propose simple nodes that could be envisaged for a new generation of OADMs.

At first sight, we could see a disadvantage in the introduction of WDM switching, since it requires WDM TRX. WDM TRX contributes to simplify the TRX structures. In WDM packet switching systems, we do not need fast tunable lasers or fast tunable filters. The ADM can be as simple as two optical couplers and one optical amplifier. The question is then: Is it not too expensive to install WDM TRX at the beginning when this capacity is not required? In fact, a total cost of ownership (TCO) analysis demonstrates that the opposite is true. By targeting end-of-life capacity we reduce operating expense (OPEX) costs, can benefit from the optical integration of current TRX to have cheaper TRX, and can better exploit network resources through a software-defined networking (SDN) orchestrator without the need to regularly reinstall new boards in order to enable capacity growth, in contrast to the single-channel approach.

Previous analyses have demonstrated that system simplicity is required for two main reasons:

- The cost of optical technology must be cheaper to be competitive with Ethernet.
- One must service 5G applications, where cost is a key issue.

Eco-design to Reduce Energy Consumption and Cost

Eco-design has been identified as fundamental in different industrial sectors to design new generations of systems and products. Eco-design is targeted at producing energy-efficiency designs to reduce the energy consumption of ICTs. The GWATT tool shows that two network segments are particularly important for eco-design, viz. access aggregation and data centers. Therefore, OPS systems should use eco-design techniques to reduce energy consumption.

The first attempt at eco-design was to add an eco-management layer, in order to put some equipment into sleep mode when no traffic was crossing nodes. However, this layer increased the cost of products, and operators were not ready to pay more, even if they would pay less for energy consumption.

Eco-design means to try to reuse existing technology in different configurations, to take into account the carbon footprint of all its enabling elements, to reduce as much as possible the operation frequency to control power consumption, and to identify business models where the product can be under control from the beginning of its production to the end of its life, including its recycling process. OPS technology has the potential to provide a new generation of eco-design systems if:

- We take into account the carbon footprint, from the design to the recycling.
- We design the system, when possible, with existing components.
- We adopt a long-lifetime subsystem, independent of the bit rate or modulation format, to increase the life time of the product to facilitate upgrades.
- We have modular parts, to pay as we grow on the electronics side.
- We have a programmable system, to manage its evolution easily, including easy upgrades.

Modularity, to pay as you grow, is important; but one must find a trade-off to limit upgrades on site that directly impact OPEX. Programmability is fundamental to control and adapt resources, reducing the need for physical maintenance. It is then important that OPS systems are fully programmable, to easily adapt their functionality to the needs of the network, and to simplify the provision of resources when required. The objective is then to design white/green boxes, at low cost and low energy consumption, and offering new perspectives in terms of performance when compared with electronic systems.

20.3.7 Opportunities

Aligning Optical Packet Technology with Network Evolution

Nowadays, packet technology is used in the following network segments:

- In the access segment with ADSL, EPON, and GPON systems
- In the metro segment with Ethernet and MPLS technologies
- In the core segment with IP/MPLS and Ethernet technologies.

Optical technology is used to enable high-capacity connections everywhere in the network. OPS technology could be more competitive than pure electronic solutions for the following main reason: Optical bypass overcomes the need for buffering and more generally electronic processing for passthrough traffic. Beyond the processing reduction, packet jitter can be kept at a minimum, at least for traffic crossing OPS nodes.

Network Segments Where OPS Could Bring New Advantages

Active optical LANs could be the natural evolution of passive optical LANs based on xPON systems. It has been demonstrated that PON systems can be deployed in LANs with potential benefits in power consumption and cost. However, the need for multiconnectivity networks, and not only hub-and-spoke architectures, could raise the need for a new technology. Active optical LANs could help in this regard. This possibility could maintain the double bus (one for upstream and one for downstream traffic) while adding some integrated amplification. The use of optical couplers to enable distributed broadcasting is a promising alternative to enable local point-to-point connections when using OPS technology.

Access aggregation and backhauling are interesting use cases to introduce OPS technology. By combining the WDM dimension and components from the access, a new low-cost technology such as that proposed in the ANR N-GREEN project can be considered.

For the metro core segment, more capacity is required than in the access segment. It also requires mesh topologies. Therefore, OPS technology could be exploited to replace the optical transport layer (based on R-OADMs) plus the L2 of the OTN technology to achieve an L2/L1 low-cost technology that would be compliant with IP/MPLS technology while offering a dynamic optical layer to facilitate the exchange of data between distant data centers.

Fronthaul is also an important use case, being very sensitive to the end-to-end latency and synchronization issues. One should be able to guarantee less than 250 µs in this part of the network for 5G technology. This small latency imposes short distances (maximum 20 km) to guarantee enough time for the processing carried out in the base band units. OPS technology in a WDM version to relax the time constraints through an acceleration of the bit rate and by adopting a TDMA mechanism for synchronization issues could be a good candidate for 5G fronthaul. In addition, a 4G technology could be put on top to enable a backhaul to offer a Xhaul technology.

Even if the need for ultrahigh-capacity backbones has slowed due to the cloudification of the metro area, OPS could be of interest in backbones to exchange data between distant data centers in a dynamic way. The objective is to reproduce the dynamic intraconnection inside the data center, but extended to the network. OPS technology could provide such a technology, and through sharing of the optical spectrum it is then possible to combine FlexGrid and OPS, like in the N-GREEN project. Another application case is in +100 Tb/s backplanes for interconnection of line cards of large switches [20.66]. OPS technology could be of interest to offer ultrahigh switching capacities at low cost and low power consumption.

Data centers are also an important segment where OPS technology can be introduced [20.67, 68]. We identify at least three use case for OPS technology:

- Load balancers to share the capacity between different line cards of the border router
- For disaggregated servers to interconnect dynamically memories and central processing units (CPUs)
- For interconnection of top-of-rack switches to offer minimum latency between the board router and the servers.

Market Perspectives

For a realistic introduction of OPS technology into modern networks, there is a need to capitalize on:

- Simple systems
- Optical components already proven to be robust, and developed for different use cases
- SDN to integrate programmability and reduce capital expenditure (CAPEX) and OPEX
- Low-cost designs
- Low power consumption and eco-design approaches
- The WDM dimension used as a parallel dimension to reduce the frequency at the edge of the network and in the line cards of the optical systems
- The identification of realistic use cases for operators
- Interoperability with Ethernet
- A multivendor approach and related standardization to ease market concerns.

20.3.8 Conclusions and Perspectives for OPS

Many achievements have been made in the overall OPS arena, from all-optical concepts to hybrid concepts, in-

cluding simple systems for rapid introduction into the market. However, it seems clear that the community needs to propose very simple solutions based on robust components, and to focus its attention on realistic use cases. Today, a need is clearly identified in access aggregation and 5G, but also anywhere where shorter end-to-end latencies are required.

Optical packet technologies have the potential to:

- Offer low-power-consumption solutions
- Offer low-latency systems thanks to the adoption of optical bypass
- Outperform Ethernet
- Better exploit the WDM dimension to pave the way to white/green boxes for fully programmable systems
- Increase the reliability of the network by minimizing the number of component crossed in the passthrough
- Change the upgrade model, by doing upgrades without any interruption of service
- Pave the way to new business models by exploiting subsystems offering longer lifetimes and other subsystems offering modularity.

The limitations of OPS technology are mainly due to the lack of optical memories. Even if some work has been carried out in the past to identify some approaches able to store some bits, the storage capacity is not enough to be envisaged in packet switches.

The second limitation is the duality of analog + digital dimensions. These constraints today limit OPS systems to optical add/drop systems that reduce connectivity to 2, to the requirement for at least one electronic buffer on the insertion port, and buffering performance in terms of contention resolution comparable to full electronic systems. This is the only way today to compete with electronic technology.

Finally, another limitation is related to the components. The community has focused its energy on complex components such as fast tunable lasers and wavelength selectors. However, these dedicated components are hardly compatible with low-cost approaches and are antagonist to eco-design approaches. It is mandatory to rethink the system design, which must be articulated around optical components already present in the marketplace, taking advantage of their large distribution volumes.

Perspectives

Over the next decade, the exploitation of OPS technologies seems interesting for the following network segments:

- Access aggregation networks, based on ring topologies
- FH/BH/XH to satisfy the key performance indicators (KPIs) of 5G, with low-cost and high-performance technology [20.69].
- Active optical LANs to create an evolution scenario to passive optical LANs
- Inside data centers:
  - For load balancers
  - For the interconnection between ToR switches and their servers to minimize latency and suppress packet jitter
- Optical backbones where latency is a key issue.

In the medium term, the metro core is probably a good target, since we could reuse existing IP/MPLS switches/routers with a new optical transport layer combining L2 and L1 through OPS technology such as WDM slotted add/drop multiplexer technology.

Requirements for OPS

There is a need for efficient processes for high-volume production of generic WDM transmitters (subband, selectable) to reduce their cost. New approaches are also needed for system and component design (in particular, tunable components). Si photonics could be a strategic direction to explore to make ultralow-cost components and devices.

The cloud environment is creating new opportunities in terms of end-to-end latency. OPS technology could create a breakthrough when repositioned in the 5G context. Cloudification puts new constraints on the network and leads to a new question: How to provide a packet-based solution offering zero jitter and the shortest end-to-end latency (excluding the propagation delay)? For example, in the 5G context, OPS technology could be the only known technology offering short end-to-end latencies, no packet jitter when adopting synchronous mechanisms, compatibility with M2M traffic through distributed connectivity, high energy efficiency, low cost, easy upgrades, and SDN compliance. Therefore, proposals are required to identify pragmatic directions to support cloud environments.

Energy efficiency will have a major impact on the industry in the next decade in terms of its capacity to reduce CO_2 emissions. Therefore, eco-design and eco-management are extremely important. OPS technology has the potential to offer high switching capacities with energy efficiencies that could be as low as 100 fJ/bit (10 Tb/s crossing an optical gate, assuming 1 W of power consumption). As an example, the ANR N-GREEN project targeted new WDM add/drop multiplexers, reducing power consumption by comb-

ing parallelism in the optical domain to slow down the frequency in the edge electronic interfaces, low-power-consuming components, and simple systems leading to small losses [20.70, 71]. The WDM approach is extremely powerful to make compact systems for OADMs, but also for systems used in the interconnection of high-bit-rate line cards of large switches/routers.

The L2 technique can be reconsidered and broadcast-and-select mechanisms envisaged to offer minimum latencies and prepare convergence towards Industry 4.0 to achieve coexistence of M2M traffic.

The route to market could be rethought by proposing an incremental scenario based on well-known components and systems, while integrating a long-term vision by proposing disruptive systems via an incremental evolution; for example, one could start with PON systems and introduce optical amplification and replace the tree topology by a bus topology, to introduce optical add/drop systems. A second example could be to start with a circuit technology, and introduce a low-TCO/bit OADM to create an incremental scenario, better adapted for the access part.

OPS will then define the new generation of sustainable systems and networks. Many opportunities exist in this regard, and the scientific community should look at promising directions in the short and medium term; system simplicity is a key enabler to open up markets, and incremental steps could be found as in the case of POL systems. OPS technology could provide a real solution for 5G problems, since it can address its KPIs. Finally, with the emergence of new vertical markets, low-cost OPS technology could be advantageous in terms of latency, cost, and energy consumption, with respect to other classical technologies.

20.4 General Conclusions

OPS and OBS technologies have been extensively studied with the perspective of finding an alternative to pure electronic switching systems. In the early 1990s, there was a strong emphasis on these new technologies, but the lack of optical memory and the poor maturity of highly integrated optical technology prevented its introduction into networks. It then became necessary to rethink the technical approach to find a realistic path to market. OBS has demonstrated that burstification of native packets is extremely efficient to transport data, and that the transport of headers on a separate resource could be more efficient than adopting headers attached to the payload. OPS technology also reached the same conclusions. The lack of optical memory pushed the community to propose OADMs to have electronic buffers at the add and drop ports, and the transport of the headers on a separate wavelength was also seen as an efficient technique to route the optical packets in a network.

Finally, OBS and OPS led to innovative concepts, adopting the burstification of native packets, the transport of headers on a separate wavelength, and electronic buffering to offer performance that could be competitive with fully electronic systems.

This technology, mainly based today on optical slot switching (on a single wavelength or in WDM), opens new perspectives in terms of short end-to-end latencies (thanks to the adoption of optical bypass and broadcast-and-select mechanisms), lower energy consumption than fully electronic systems, and easy upgrade when adopting transparent and colorless ODAM structures.

With the need for energy-efficient systems satisfying the KPIs of 5G, OBS/OPS technology could be a good candidate for the next generation of systems and networks if a strategy for their introduction into the network is proposed. An incremental evolution of the technology in a disruptive long-term vision is probably one path to market and will provide concrete end-to-end 5G solutions.

References

20.1 W.L. Ha, R.M. Fortenberry, R.S. Tucker: Demonstration of photonic fast packet switching at 700 Mbit/s data rate, Electron. Lett. **27**, 789–790 (1991)

20.2 C. Qiao, M. Yoo: Optical burst switching (OBS) – new paradigm for an optical Internet, J. High Speed Netw. **8**(1), 69–84 (1999)

20.3 C. Qiao: Labeled optical burst switching for IP-over-WDM integration, IEEE Commun. Mag. **38**(9), 104–114 (2000)

20.4 Y. Chen, C. Qiao, X. Yu: Optical burst switching: A new area in optical networking research, IEEE Network **18**(3), 16–23 (2004)

20.5 B. Mukherjee: *Optical WDM Networks* (Springer, New York 2006)

20.6 M.J. O'Mahony, D. Simeonidou, D.K. Hunter, A. Tzanakaki: The application of optical packet switching in future communication networks, IEEE Commun. Mag. **39**(3), 128–135 (2001)

20.7 G.I. Papadimitriou, C. Papazoglou, A.S. Pomportsis: Optical switching: Switch fabrics, techniques, and architecture, J. Lightwave Technol. **21**(2), 384 (2003)

20.8 C.A. Brackett: Dense wavelength division multiplexing networks: Principles and applications, IEEE J. Sel. Areas Commun. **8**(6), 948–964 (1990)

20.9 L. Xu, H.G. Perros, G. Rouskas: Techniques for optical packet switching and optical burst switching, IEEE Commun. Mag. **39**(1), 146–142 (2001)

20.10 R. Takahashi, R. Urata, H. Takenouchi, T. Nakahara: Hybrid optoelectronic router for asynchronous optical packets. In: *Proc. Int. Conf. Photonic Switch. (PS), Pisa* (2009), https://doi.org/10.1109/PS.2009.5307752

20.11 P.J. Argibay-Losada, G. Sahin, K. Nozhnina, C. Qiao: Transport-layer control to increase throughput in bufferless optical packet-switching networks, J. Opt. Commun. Netw. **8**(12), 947–961 (2016)

20.12 P.J. Argibay-Losada, Y. Yoshida, A. Maruta, K.-I. Kitayama: Optical versus electronic packet switching in delay-sensitive 5G networks: Myths versus advantages, J. Opt. Commun. Netw. **8**(11), B43–B54 (2016)

20.13 P.J. Argibay-Losada, Y. Yoshida, A. Maruta, M. Schlosser, K.-I. Kitayama: Performance of fixed-length, variable-capacity packets in optical packet-switching networks, J. Opt. Commun. Netw. **7**(7), 609–617 (2015)

20.14 P.J. Argibay-Losada, G. Sahin, K.-I. Kitayama, C. Qiao: On whether OCS maximizes application throughput in all-optical datacenter networks, J. Opt. Commun. Netw. **7**(12), 1135–1147 (2015)

20.15 S.J.B. Yoo: Optical packet and burst switching technologies for the future photonic Internet, J. Lightwave Technol. **24**(12), 4468–4492 (2006)

20.16 X. Yu, J. Li, X. Cao, Y. Chen, C. Qiao: Traffic statistics and performance evaluation in optical burst switched networks, J. Lightwave Technol. **22**(12), 2722–2738 (2004)

20.17 A. Detti, M. Listanti: Impact of segments aggregation on TCP Reno flows in optical burst switching networks. In: *Proc. Conf. Comput. Commun. (INFOCOM), New York* (2002), https://doi.org/10.1109/INFCOM.2002.1019434

20.18 S. Gowda, R.K. Shenai, K.M. Sivalingam, H.C. Cankaya: Performance evaluation of TCP over optical burst-switched (OBS) WDM networks. In: *Proc. IEEE Int. Conf. Commun. (ICC), Anchorage* (2003), https://doi.org/10.1109/ICC.2003.1204627

20.19 A. Ge, F. Callegati, L.S. Tamil: On optical burst switching and self-similar traffic, IEEE Commun. Lett. **4**(3), 98–100 (2000)

20.20 J. Li, C. Qiao, J. Xu, D. Xu: Maximizing throughput for optical burst switching networks, IEEE ACM Trans. Netw. **15**(5), 1163–1176 (2007)

20.21 J.Y. Wei, R.I. McFarland Jr.: Just-in-time signaling for WDM optical burst switching networks, J. Lightwave Technol. **18**(12), 2019–2037 (2000)

20.22 G. Wu, T. Zhang, J. Chen, X. Li, C. Qiao: An index-based parallel scheduler for optical burst switching networks, J. Lightwave Technol. **29**(18), 2766–2773 (2011)

20.23 A. Detti, V. Eramo, M. Listanti: Performance evaluation of a new technique for IP support in a WDM optical network: Optical composite burst switching (OCBS), J. Lightwave Technol. **20**(2), 154–165 (2002)

20.24 C.-F. Hsu, T.-L. Liu, N.-F. Huang: Performance analysis of detection routing in optical burst-switched networks. In: *Proc. Conf. Comput. Commun. (INFOCOM), New York* (2002), https://doi.org/10.1109/INFCOM.2002.1019247

20.25 S. Kim, N. Kim, M. Kang: Contention resolution for optical burst switching networks using alternative routing. In: *Proc. IEEE Int. Conf. Commun. (ICC), New York* (2002), https://doi.org/10.1109/ICC.2002.997329

20.26 J. Ramamirtham, J. Turner: Time sliced optical burst switching. In: *Proc. Conf. Comput. Commun. (INFOCOM), San Francisco* (2003), https://doi.org/10.1109/INFCOM.2003.1209224

20.27 J.S. Turner: Terabit burst switching, J. High Speed Netw. **8**(1), 3–16 (1999)

20.28 C.M. Gauger: Dimensioning of FDL buffers for optical burst switching nodes. In: *Proc. Int. Conf. Opt. Netw. Design Model. (ONDM), Torino* (2002), https://doi.org/10.1007/978-0-387-35670-9_8

20.29 Z. Rosberg, H.L. Vu, M. Zukerman, J. White: Performance analyses of optical burst-switching networks, IEEE J. Sel. Areas Commun. **21**(7), 1187–1197 (2003)

20.30 M. Yoo, C. Qiao, S. Dixit: Optical burst switching for service differentiation in the next-generation optical Internet, IEEE Commun. Mag. **39**(2), 98–104 (2001)

20.31 M. Yoo, C. Qiao, S. Dixit: QoS Performance of optical burst switching in IP-over-WDM networks, IEEE J. Sel. Areas Commun. **18**(10), 2062–2071 (2000)

20.32 V.M. Vokkarane, J.P. Jue: Prioritized burst segmentation and composite burst-assembly techniques for QoS support in optical burst-switched networks, IEEE J. Sel. Areas Commun. **21**(7), 1198–1209 (2003)

20.33 V.M. Vokkarane, K. Haridoss, J.P. Jue: Threshold-based burst assembly policies for QoS support in optical burst-switched networks, Proc. SPIE (2002), https://doi.org/10.1117/12.475291

20.34 V.M. Vokkarane, J.P. Jue, S. Sitamaran: Burst segmentation: An approach for reducing packet loss in optical burst switched networks. In: *Proc. IEEE Int. Conf. Commun. (ICC), New York* (2002), https://doi.org/10.1109/ICC.2002.997328

20.35 T. Battestilli, H. Perros: An introduction to optical burst switching, IEEE Commun. Mag. **41**(8), S10–S15 (2003)

20.36 K. Dolzer, C. Gauger, J. Spath, S. Bodamer: Evaluation of reservation mechanisms for optical burst switching, Int. J. Electron. Commun. **55**(1), 18–26 (2001)

20.37 K. Dolzer, C. Gauger: On burst assembly in optical burst switching networks a performance evaluation of just-enough-time. In: *Proc. Int. Teletraffic Congr., Salvador* (2001)

20.38 S. Verma, H. Chaskar, R. Ravikant: Optical burst switching: A viable solution for terabit IP backbone, IEEE Network **14**(6), 48–53 (2000)

20.39 H.L. Vu, M. Zukerman: Blocking probability for priority classes in optical burst switching networks, IEEE Commun. Lett. **6**(5), 214–216 (2002)

20.40 T. Zhang, W. Dai, G. Wu, X. Li, J. Chen, C. Qiao: A dual price-based congestion control mechanism for optical burst switching networks, J. Lightwave Technol. **32**(14), 2492–2501 (2014)

20.41 K. Long, R.S. Tucker, C. Wang: A new framework and burst assembly for IP DiffServ over optical burst switching networks. In: *Proc. GLOBECOM, San Francisco* (2003), https://doi.org/10.1109/GLOCOM.2003.1258818

20.42 G. Zervas, Y. Qin, R. Nejabati, D. Simeonidou, F. Callegati, A. Campi, W. Cerroni: SIP-enabled optical burst switching architectures and protocols for application-aware optical networks, Comput. Netw. **52**(10), 2065–2076 (2008)

20.43 S. Lizao, Q. Zhang, Q. Tian, Y. Tao, B. Liu, N. Liu, Y. Shen, D. Chen, F. Tian, H. Wang, X. Xin: Research of on-board optical burst switching weight-weighted round robin assembly algorithm based on the QoS guarantee. In: *Proc. Int. Conf. Opt. Commun. Netw. (ICOCN), Hangzhou* (2016), https://doi.org/10.1109/ICOCN.2016.7875732

20.44 IEEE 802.3 Ethernet Working Group: http://www.ieee802.org/3/

20.45 WIDE: Packet traces from WIDE backbone, http://mawi.wide.ad.jp/mawi (2020)

20.46 K.-I. Kitayama, M. Koga, H. Morikawa, S. Hara, M. Kawai: Optical burst switching network testbed in Japan. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2005), https://doi.org/10.1109/OFC.2005.192713

20.47 Y. Sun, T. Hashiguchi, V.Q. Minh, X. Wang, H. Morikawa, T. Aoyama: Design and implementation of an optical burst-switched network testbed, IEEE Opt. Commun. **43**(11), S48–S55 (2005)

20.48 H. Guo, J. Wu, Z. Lan, Z. Gao, X. Li, J. Lin, Y. Ji: A testbed for optical burst switching networks. In: *Proc. Opt. Fiber Commun. Conf. (OFC), Anaheim* (2005), https://doi.org/10.1109/OFC.2005.193069

20.49 I. Baldine, M. Cassada, A. Bragg, G. Kamous-Edwards, D. Stevenson: Just-in-time optical burst switching implementation in the ATDnet all-optical networking testbed. In: *Proc. GLOBECOM, San Francisco* (2003)

20.50 P.J. Argibay-Losada, K. Nozhnina, G. Sahin, C. Qiao: Using stop-and-wait to improve TCP throughput in fast optical switching (FOS) networks over short physical distances. In: *Proc. Conf. Comput. Commun. (INFOCOM), Toronto* (2014), https://doi.org/10.1109/INFOCOM.2014.6848064

20.51 D. De Bouard, G. Da Loura, C. Chauzat, J. Jacquet, J. Benoit, D. Leclerc, J.M. Gabriagues, J.B. Jacob, G. Le Roy: Fast optical triggering and wavelength switching using a DBR laser with a saturable absorber. In: *Proc. Int. Conf. Photonic Switch. (PS), Salt Lake City* (1991), Paper WA2-1

20.52 D. Chiaroni, M. Sotom, P. Doussiere: Wavelength channel selector with subnanometric resolution and subnanosecond switching time. In: *Proc. ECOC, Firenze* (1994)

20.53 T. Durhuss: Semiconductor optical amplifiers as gates and wavelength converters in delay line memory blocks. In: *Proc. COST 240, Athens* (1992)

20.54 P. Gavignet-Morin, F. Masetti, D. Chiaroni, G. Da Loura, J.M. Gabriagues: Multiwavelength optical buffer based on fibre delay lines for gigabit packet switching. In: *Proc. Conf. Opt. Fiber Commun. (OFC), San Jose* (1993), https://doi.org/10.1364/OFC.1993.WJ1

20.55 D. Chiaroni, P. Gavignet-Morin, P.A. Perrier, S. Ruggeri, S. Gauchard, D. Boudard, J.J. Jacquinot, C. Chauzat, J. Jacquet, P. Doussiere, M. Monot, E. Grard, D. Leclert, M. Sotom, J.M. Gabriagues, J. Benoit: Rack-mounted 2.5 Gbit/s ATM photonic switch demonstrator. In: *Proc. ECOC, Montreux* (1993)

20.56 B. Lavigne, P. Guerber, D. Chiaroni, C. Janz, A. Jourdan, B. Sartorius, C. Bornholdt, M. Morhle: Test at 10Gbit/s of an optical 3R regenerator using an integrated all-optical clock recovery. In: *Proc. ECOC, Nice* (1999)

20.57 C. Guillemot, M. Renaud, P. Gambini, C. Janz, I. Andonovic, R. Bauknecht, B. Bostica, M. Burzio, F. Callegati, M. Casoni, D. Chiaroni, F. Clerot, S.L. Danielsen, F. Dorgeuille, A. Dupas, A. Franzen, P.B. Hansen, D.K. Hunter, A. Kloch, R. Krahenbuhl, B. Lavigne, A. LeCorre, C. Raffaelli, M. Schilling, J.-C. Simon, L. Zucchelli: Transparent optical packet switching: The European ACTS KEOPS project approach, J. Lightwave Technol. **16**(12), 2117–2134 (1998)

20.58 F. Dorgeuille, N. Le Sauze, B. El Khattar, L. Ciavaglia, E. Dotaro, T. Zami: Dual optical bus for efficient aggregation and backhauling networks and potential extensions, Bell Labs Techn. J. **14**(4), 243–264 (2010)

20.59 A. Stavdas, A. Salis, A. Dupas, D. Chiaroni: All-optical packet synchronizer for slotted core/metropolitan networks, J. Opt. Netw. **7**, 88–93 (2008)

20.60 L. Dittmann, C. Develder, D. Chioaroni, F. Neri, F. Callegati, W. Koerber, A. Stavdas, M. Renaud, A. Rafel, J. Sole-Pareta, W. Cerroni, N. Leligou, L. Dembeck, B.B. Mortensen, M. Pickavet, N.L. Sauze, M. Mahony, B. Berde, G. Eilenberger: The European IST project DAVID: A viable approach toward optical packet switching, IEEE J. Sel. Areas Commun. **21**(7), 1026–1040 (2003)

20.61 D. Chiaroni: The French RNRT ECOFRAME project: Packet technology perspectives in metro networks. In: *Proc. APOC, Hangzhou* (2008)

20.62 R. Takahashi, R. Urata, H. Takenouchi, T. Nakahara: Hybrid optoelectronic router for asynchronous optical packets. In: *Proc. Int. Conf. Photonic Switch. (PS)* (2009), https://doi.org/10.1109/PS.2009.5307752

20.63 D. Chiaroni: Green metro network proposal for a convergent scenario: Results of the ALPHA project. In: *Proc. Opto-Electron. Commun. Conf. (OECC), Kaohsiung* (2011) pp. 764–767

20.64 K. Ishii, O. Moriwaki, H. Hasegawa, K. Sato, Y. Jinnouchi, M. Okuno, H. Takahashi: Efficient ROADM-ring connecting node switch architecture that utilizes waveband routing and its realization with PLC

technologies. In: *Proc. Eur. Conf. Opt. Commun. (ECOC), Vienna* (2009)

20.65 N. Wada: WDM-colored packet switching. In: *Proc. Conf. Opt. Fiber Commun./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC), San Diego* (2008), https://doi.org/10.1109/OFC.2008.4528728

20.66 J.M. Finochietto, R. Gaudino, G.A. Gavilanes Castillo, F. Neri: Simple optical fabrics for scalable terabit packet switches. In: *Proc. Int. Conf. Commun. (ICC), Beijing* (2008), https://doi.org/10.1109/ICC.2008.1000

20.67 S. Yoo, Y. Yin, K. Wen: Intra and inter datacenter networking: the role of optical packet switching and flexible bandwidth optical networking. In: *Proc. Int. Conf. Opt. Netw. Design Model. (ONDM), Colchester* (2012), https://doi.org/10.1109/ONDM.2012.6210261

20.68 K.-I. Kitayama, Y.C. Huang, Y. Yoshida, R. Takahashi, T. Segawa, S. Ibrahim, T. Nakahara, Y. Suzaki, M. Hayashitani, Y. Hasegawa, Y. Mizukoshi, A. Hiramatsu: Torus-topology data center network based on optical packet/agile circuit switching with intelligent flow management, J. Lightwave Technol. **33**(5), 1063–1071 (2015)

20.69 B. Uscumlic, D. Chiaroni, B. Leclerc, T. Zami, A. Gravey, P. Gravey, M. Morvan, D. Barth, D. Amar: Scalable deterministic scheduling for WDM slot switching Xhaul with zero-jitter. In: *Proc. Int. Conf. Opt. Netw. Design Model. (ONDM), Dublin* (2018), https://doi.org/10.23919/ONDM.2018.8396114

20.70 D. Chiaroni: Network energy: Problematic and solutions towards sustainable ICT. In: *Proc. Congr. Int. Comm. Opt. (ICO), Tokyo* (2017), Tu1E-07

20.71 M. Nakagawa, K. Matsumoto, H. Onda, K. Matsumura: Photonic sub-lambda transport: An energy-efficient and reliable solution for metro networks. In: *Proc. Int. Conf. Opt. Netw. Design Model. (ONDM), Dublin* (2018), https://doi.org/10.23919/ONDM.2018.8396125

Pablo J. Argibay-Losada

BITS
SBI Group
Tokyo, Japan
pargibay@ieee.org

Pablo received his PhD in telecommunications engineering from Vigo University (Spain) in 2009. He worked on optimizing TCP/IP stacks for optical networks at the LANDER lab at SUNY Buffalo (USA) and at the Photonics Network Lab at Osaka University (Japan). He currently develops high-performance distributed trading systems in the Japanese financial sector. Pablo is a senior IEEE member who also enjoys developing autonomous mobile robotic systems.

Dominique Chiaroni

Nokia Bell Labs France
Nozay, France
dominique.chiaroni@nokia-bell-labs.com

Dominique Chiaroni received his Bachelor's and Master's Degrees in thermal sciences and his BSc in optics and microwaves. He is currently working in the IP and Optical Networks Lab within Nokia Bell Labs France, with a focus on low-cost and low-power-consuming smart fabrics for 5G applications. He has authored and coauthored more than 200 publications and patents, including contributions in books.

Chunming Qiao

Dept. of Computer Science & Engineering
University of Buffalo
Buffalo, NY, USA
qiao@buffalo.edu

Chunming Qiao is a SUNY Distinguished Professor and currently chairs the CSE Department at University at Buffalo. He has led the Lab for Advanced Network Design, Evaluation and Research (LANDR) at UB since 1993. His current research interests cover not only the safety and reliability of various cyber physical systems, but also algorithms and protocols for the Internet of Things, including smartphone-based systems and applications.

Part C Datacenter

Part C Datacenter and Super-Computer Networking

Ed. by Ioannis Tomkos

The Internet (and consequently modern society and lifestyles) is based on different types of interconnected datacenters (DCs) that are owned by major industry giants such as datacenter operators (e.g., Google, Amazon, Facebook, Alibaba, and Microsoft) and telecom operators (e.g., Verizon, AT&T, Orange, DT, Vodafone, NTT, and China Telecom) alongside many smaller datacenter and telecom operators. Due to end-user needs, overall Internet traffic is increasing at a significant rate each year. All of these industry players are therefore forced to make significant investments in infrastructure (e.g., datacenters, central offices, and transmission equipment) to satisfy the ever-growing demands of their customers for more bandwidth-hungry services based on video as well as the emergence of new technologies (e.g., artificial intelligence) that require more and more data to be exchanged among people, machines, and things (as in the Internet-of-Everything). Notably, the vast majority of the total internet traffic generated stays within the datacenters, some of it is exchanged among datacenters, while only a small percentage represents end-user to datacenter traffic. In order to handle the continuous growth in traffic, more large datacenters comprising thousands of servers that are interconnected using ultrahigh-throughput switches are needed. Datacenters based on electrical switches and interconnects consume excessive power when they are employed to handle the increased communication bandwidths of available and emerging cloud applications. As a result, optical interconnects and even optical switching have gained significant attention over the past decade as promising futureproof solutions that offer higher interconnection speeds along with reduced energy consumption. Their applicability ranges from inter- and intra-datacenter interconnections down to the level of chip-to-chip and on-chip interconnections, supporting high-performance computing (HPC). The associated datacenter interconnection (DCI) and high-performance computer networks that are based on novel optoelectronics and photonics technologies, devices, subsystems, and systems represent an enormous market opportunity for the optical communications industry ecosystem, and are therefore the focus of Part C of this Handbook.

Chapter 21 defines the interconnect types used in the global network of datacenters, describes the inter- and intra-datacenter network architectures used by the major datacenter operators, and discusses their evolving high-level requirements in terms of interconnection bandwidth, switch-port speeds, energy consumption, and associated scalability trends.

Chapter 22 first defines the metrics that are used to evaluate and compare the performance levels of high-performance computing (HPC) systems. It then focuses on the state-of-the-art performance of these systems and predicted future requirements for the metrics. An analysis of the topology and characteristics of HPC interconnects is then provided, and the HPC interconnect building blocks for realizing adapters, routers, and high-speed connectivity links are discussed in detail. Finally, the main photonic technologies that are envisioned to provide the basis for optically connected router ASICs are explored.

Chapter 23 focuses on the interconnection topologies and technologies for implementing intra-datacenter interconnections. Currently implemented architectures are presented first, after which their possible evolution—which relies heavily on the use of photonics technologies—is analyzed. Interconnect network topologies are classified into categories, and the relevant literature is surveyed. The most important topologies are compared in terms of their scalability by considering the maximum number of servers that each topology can accommodate while retaining the same number of ports in each switch and the same number of ports in a server for a given topology. Other performance metrics (e.g., topology diameter, bandwidth, and bisection width) are also compared. The authors then discuss the pros and cons of today's electronic interfaces as compared to embedded optics approaches using standard currently available pluggable optical interfaces, as well as possible future implementations based on 2.5-D and 3-D integration using novel interposers and the potential monolithic cointegration of CMOS switch ASICs and silicon-photonics optical interfaces. Finally, the chapter summarizes the advantages and disadvantages of the main all-optical switching technologies and architectures that have been proposed for possible use in future optical interconnects in an effort to alleviate the limitations of electronic switches.

Chapter 24 discusses the implications of hyperscale datacenter architectures and interconnection requirements for optical transceivers. The authors summarize the implementation of available standardized interfaces, survey the current technologies and signal modulation/-multiplexing approaches that can be used to realize cost-effective and power-efficient optical transceivers, and discuss the ways in which the desired data-rate scaling may be achieved in the future. According to the ethernet roadmap, switch interface speeds of 800 Gb/s and 1.6 Tb/s will be required after 2020. Hence, the speeds of optical interconnect transceivers must soon be increased from the 200 and 400 Gb/s speeds that are currently commercially available. The need for a paradigm shift from the current era of pluggable transceiver module dominance towards new packaging approaches is discussed. Based on trends in the optical communications industry and the needs of hyperscale

datacenter operators and ASIC switch vendors, the on-board-optics approach (OBO) and co-packaged optics (CPO) solution are introduced. CPO, which refers to the intimate coupling of optics with an ASIC in one package, represents the next step in attempts to move the optics closer to the switch chip. Such tightly coupled optics promises to overcome the growing system challenges associated with linking the high-speed signals of an ASIC to pluggable optics modules residing on the platform's faceplate, thus allowing the faceplate bandwidth density to be increased as desired.

Chapter 25 provides a detailed discussion of the benefits that can result from the introduction of optical switching inside datacenters as a next step in the evolution of optical interconnects. The authors first discuss the slow switching-speed technologies used in the conventional optical circuit switching (OCS) approach and compare their performance levels based on a number of important characteristics. They then introduce the key technologies that can be used to implement the faster optical switches needed for optical burst and even optical packet switching (OBS and OPS). Technologies that could enable ultrafast switching are compared in terms of switching speed, scaling capability, and other important parameters. Finally, the technical challenges that must be addressed by the research community before fast optical switches for OBS/OPS can be made commercially feasible are analyzed.

21. Evolving Requirements and Trends in Datacenter Networks

Hong Liu, Ryohei Urata, Xiang Zhou, Amin Vahdat

In this manuscript, we present an overview of Google's datacenter network, which has led and defined the industry over the past few decades. Starting inside the datacenter, we cover all aspects, from networking/topology to key hardware components of interconnect and switching, traffic/throughput, and energy usage/efficiency for intra-datacenter networks. Likewise, we discuss topology and interconnect for inter-datacenter networks. With particular focus on optical interconnect, we also discuss future technology directions for scaling bandwidth through a combination of higher baud rates, wavelength-division multiplexing (WDM), coherent communication (polarization multiplexing, I/Q modulation), and space-division multiplexing (SDM), along with the corresponding trade-offs between these various dimensions and how these trade-offs are adjusted at different length scales. Although questions remain on the exact implementation to be adopted in the future, one thing is clear: the evolution of datacenter networks and the underlying technologies have been and will remain a critical driver for enabling new compute capabilities in the cloud.

Over the past decade, the datacenter has become the technology enabler for web-based applications, with prominent examples being web search, social media applications, and enterprise software (email/docs/storage). With the user interface often being a thin client device (stateless mobile/laptop device), the actual running of the application is performed in a remote datacenter. Starting from humble beginnings several decades ago [21.1], Google's compute infrastructure has been dramatically improved on all axes, from the cooling and power infrastructure and associated efficiency (power usage effectiveness, PUE), to the underlying hardware of servers, storage, and networking. In fact, the datacenter has quickly evolved from just racks of servers to a unified, global computer consisting of *copies* of massive datacenters interconnected throughout the world to deliver various services with low latency and high reliability, in a synchronized/consistent fashion. Google has thus built a global computing infrastructure and, in turn, developed and incorporated transformative datacenter networking technologies in order to keep pace with the increasing number and bandwidth demands of these applications.

The importance of the datacenter will be further enhanced as the cloud IaaS (infrastructure-as-a-service) growth migrates an increasing share of all compute to the cloud. Cloud-based platforms enjoy the same benefits as the aforementioned computing for web apps in terms of scalability, accessibility, and reliability. Without the concerns of running and maintaining an information technology (IT) infrastructure, users have immediate access to a single machine or thousands of machines to quickly and easily scale and handle increasing user workloads/services whenever the need arises. All user data reside in the cloud and are accessible anywhere in the world as long as there is a network connection. Lastly, the data are automatically backed up with multiple copies dispersed throughout the world,

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_21

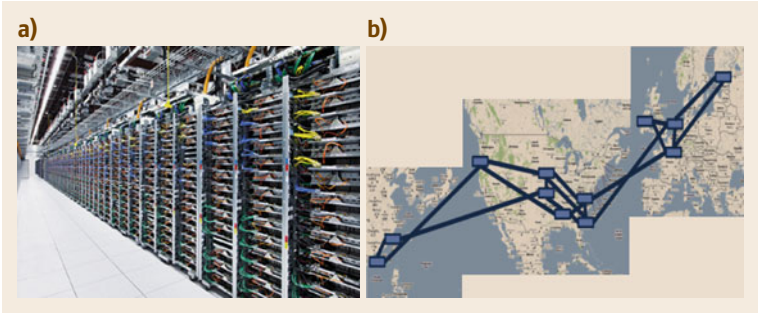


Fig. 21.1a,b Google's datacenter **(a)** showing rows of racks which house hundreds of thousands of servers inside the building (http://www.google.com/about/datacenters/) and **(b)** global B4 inter-datacenter wide area network (WAN) (after [21.2])

being fault-tolerant to local hardware or facility failure.

All of these benefits require a reliable and scalable network infrastructure. Datacenter networks are typically separated into two categories: (1) intra-datacenter networks used to connect machines/servers within the same building or same campus, and (2) inter-datacenter networks that interconnect multiple datacenters. As a result of their role/function, the intra- versus inter-type networks have different requirements regarding topology, reliability, and bandwidth capacity. Intra-datacenter networks employ a massively parallel fabric with rich path diversity for scalability and load balancing. Inter-datacenter networks are more point-to-point, with much higher capacity per link and interconnect having much longer reaches/distances. Figure 21.2 provides a high-level view of datacenter interconnects, divided into four segments based on reach and corresponding technology adopted:

1. The intra-datacenter interconnect for link distances up to 1 km within the same building (mix of copper and optics)
2. The intra-campus network, which interconnects clusters housed in different buildings within a 2 km campus neighborhood
3. The point-to-point metro edge access, which provides connections between datacenters and the global backbone network or POPs (point-of-presence), with a link distance typically less than 80 km; and
4. The long-haul and subsea backbone transport interconnecting datacenters throughout the world, with up to thousands of kilometers of reach.

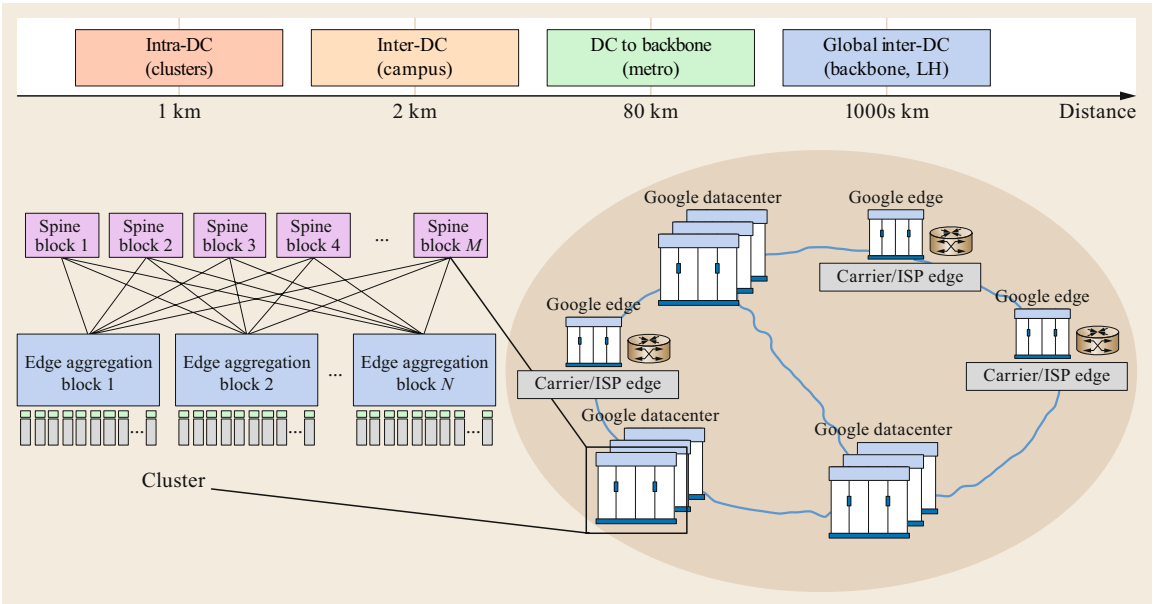


Fig. 21.2 A high-level view of interconnection types for the global datacenter network

21.1 Intra-Datacenter Network

The datacenter network at the building scale is typically a scale-out type of network that is wide and flat (such as a fat-tree network) through a richly interconnected and meshed network. This network interconnects thousands of commodity-class servers and storage devices beneath it. Leveraging a similar approach to scale-out computing with the implementation of arrays of commodity servers for increasing aggregate compute, scale-out datacenter networks are constructed using the following design principles: multiple stages of meshed switching blocks to maximize scale-out and nonblocking bandwidth provisioning with small failure domains; the use of low-cost, commodity merchant switch silicon for faster product cycles to take full advantage of Moore's law; and centralized control software with routing and management protocols tailored to intra-datacenter requirements [21.1]. A diagram of an intra-datacenter network example is shown in Fig. 21.3a,b [21.1, 3]. Machines are housed in racks, which connect up to top-of-rack (ToR) switches. There are multiple paths from the ToR to edge switches (often referred to as leaf switches) for path redundancy and fault tolerance. The edge switches may consist of multiple switches interconnected in a tightly coupled meshed fashion to constitute a nonblocking switching unit. These internal switches may consist of single switch silicon application-specific integrated circuits (ASICs) or an array of ASICs interconnected with a mesh topology. With its large radix, the edge switches can then be fanned out to all spine switches to create a network which maximizes bisection bandwidth (bisection bandwidth is the minimum possible bandwidth of the network when the fabric is bisected at all possible tiers).

An ideal datacenter network should be nonblocking to allow flexible placement of compute jobs among machines, flat with as few tiers as possible to reduce latency and cost, and wide enough to support all compute nodes with predictable latency from machine to machine. The two most important characteristics of intra-datacenter networks are scalability and cost/power efficiency, which are influenced by all aspects of the datacenter network design: fabric topology, physical layout, switch hardware, interconnect selection, network routing, and management control. With the anticipated growth of the cloud, the scaling and efficiency aspects will continue to influence future network designs. For example, a datacenter with $> 100\,000$ servers, each with 100 Gb/s of bandwidth allocated, would require an internal network with 10 Pb/s of aggregate bandwidth to support full-bandwidth communication among all servers. On the other hand, in order to maintain energy and space efficiency, as the bandwidth and number of servers scale up, the datacenter network needs to scale within the same footprint and cost.

There are many considerations in building a large-scale datacenter network, including but not limited to cost (capital expenditure and operating expenses; CapEx and OpEx) and performance. The key technologies are the fabric topology chosen, the switch silicon ASIC, and the interconnect to implement the fabric, each of which will be discussed below.

21.1.1 Fabric Topology

To connect tens of thousands of nodes, there are a variety of network topologies which can be chosen: flat-

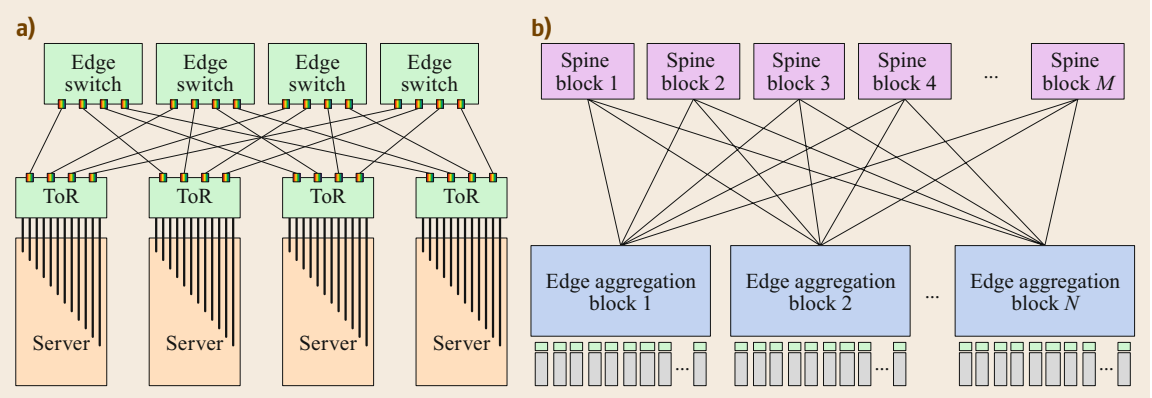


Fig. 21.3a,b Diagram of scale-out type of datacenter network. A tiered architecture is used with **(a)** fan-out from ToR to edge switches with multiple connections, **(b)** from edge aggregation block to spine blocks with all-to-all connections

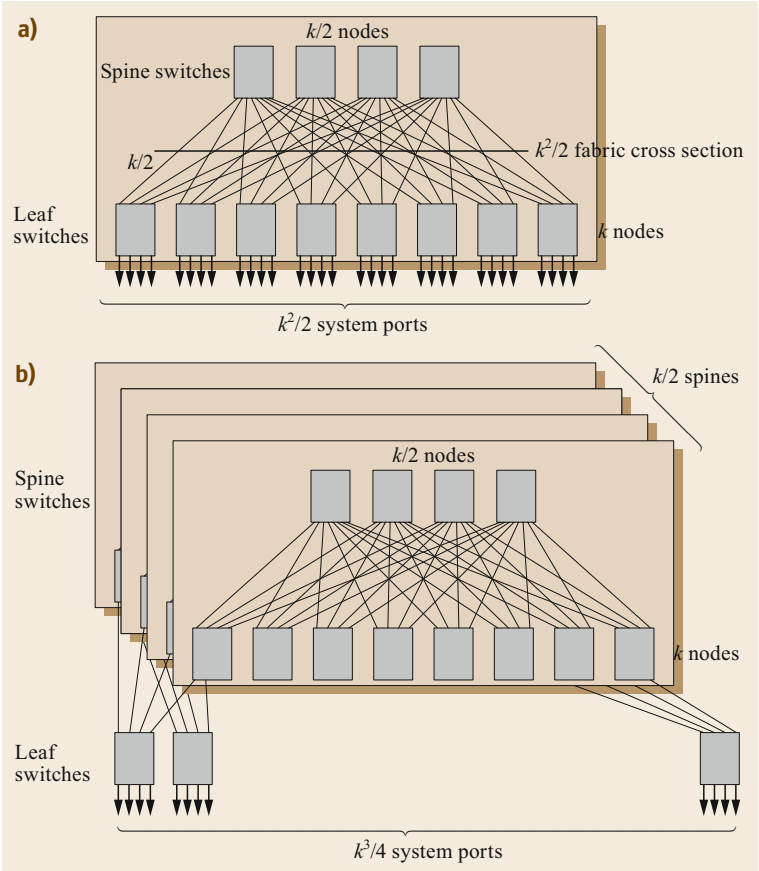


Fig. 21.4a,b Diagram of Clos fabric built with k-port switches showing **(a)** two-stage Clos, **(b)** three-stage Clos

tened butterfly [21.4], B-cube [21.5], Clos [21.6], and others [21.7]. The topology choice is often a trade-off between software routing complexity, physical link efficiency/number, and fabric scalability. For example, flattened butterfly trades off fewer physical links but requires adaptive routing to load-balance arbitrary traffic patterns, whereas a Clos/fat tree has multiple physical paths and simpler routing to handle arbitrary traffic patterns.

As described in Fig. 21.3, Clos topology is most commonly used for large-scale datacenters to support full-bisection bandwidth and graceful fault tolerance. This architecture enables the use of identical switching elements with smaller radix (i.e., the switch ASIC) to form a multistage Clos for a very large-scale network fabric that would be impossible to achieve with monolithic single-chip or single-chassis technologies [21.8]. Figure 21.4a shows a two-stage Clos constructed with identical k-port switches with all-to-all spine-to-leaf switch connections. The leaf switches have k nodes of k-port switches, with $k/2$ ports of each switch connecting to system ports, and $k/2$ ports fanning out to $k/2$ spine switch nodes. The k ports of each spine switch fan out to all leaf switches (k nodes)

$$\text{Total bisection bandwidth} = \text{number of spine links} \times \text{bandwidth per link} \times 2$$

$$\text{Total number of connected system ports} = (k^2)/2\,.$$

To further increase the bisection bandwidth and total number of system ports (or hosts) that can be interconnected, the two-stage fat tree can be extended to a three-stage fat tree (Fig. 21.4b) with $k/2$ port connections to each lower-stage switch. Layers can be continually added in this manner to increase the bandwidth of the network at the expense of latency and cost. Ideally, a fully meshed networking fabric that connects every system port in a datacenter provides full bisection bandwidth (same amount of bandwidth between any two nodes), which leads to easier programming (no consideration of the underlying network infrastructure and where bandwidth constraints lie) and better utilization of server compute capability. However, such a design would be prohibitively expensive, and modest oversubscription (i.e., bandwidth provisioned is less

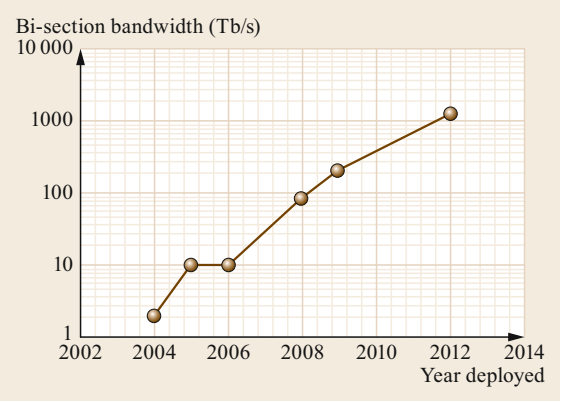


Fig. 21.5 Evolution of bisection bandwidth inside Google's datacenter

than maximum possible peak/demand bandwidth with fewer upward-facing than downward-facing links) is often applied at some layer/tier to reduce CapEx.

Figure 21.5 shows the evolution of Google's datacenter network fabric bandwidth over time. Leveraging the Clos topology and the increase of switch ASIC bandwidths, a three-orders-of-magnitude increase in bandwidth was achieved over roughly a decade, with the Jupiter network capable of 1.3 Pb/s of bisection bandwidth at maximum scale (after [21.1]).

21.1.2 Switch Silicon

Driving sustained increases in bandwidth, merchant switch silicon has become a key building block for datacenter fabrics, as it enables faster time to market and quicker product refresh cycles at much lower cost than with customized ASICs. With the continuous scaling of technologies for signaling speed, the increase in the number of signal pins, and power scaling from Moore's law, the capacity of high-speed switch chips has seen an impressive tenfold increase over the past 10 years. The evolution of the bandwidth available from a single switch chip is shown in Fig. 21.6. A single switch chip today can offer > 12 Tb/s of switching capacity, and > 25 Tb/s switching capacity in the near future using a 7 nm complementary metal–oxide semiconductor (CMOS) technology node. The radix (number of ports) of the switch chip has also increased from 16 to 256 in the past 10 years. The higher switch capacity and larger radix enable large-scale and more efficient datacenter fabrics with > Pb/s bisection bandwidth, using fewer stages and external interconnections.

However, switch silicon bandwidth may become constrained by the saturation of package pin count and slowing power consumption improvements of the I/O (SerDes interface of each port), due to Moore's law slowdown. While opportunities exist to improve the energy efficiency of each channel with advances in CMOS technology, the package pin count scales more slowly than I/O speed and power, per the International Technology Roadmap for Semiconductors (ITRS) projection, thus slowing the aggregate switch bandwidth growth.

21.1.3 Intra-Datacenter Interconnect

Inside the datacenter, one of the most notable characteristics of the network is the large amount of fan-out. A much larger number of transceivers and interconnects are thus required to implement the topology, which mo-

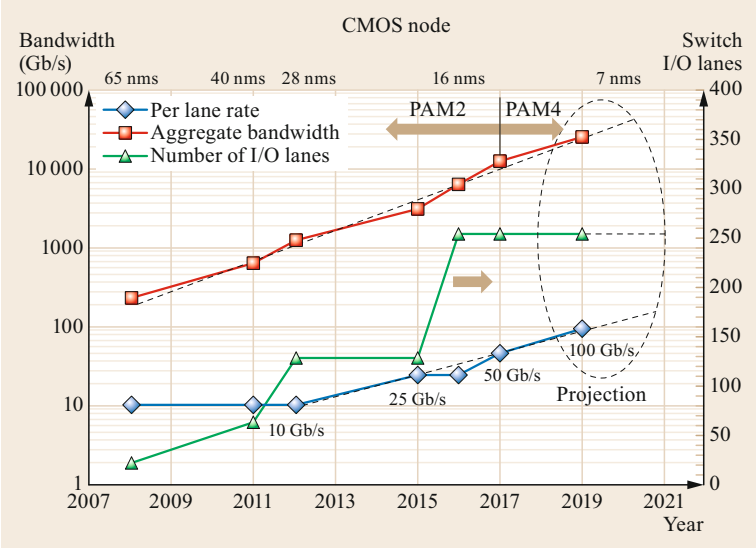


Fig. 21.6 Switch I/O bandwidth and technology scaling trend

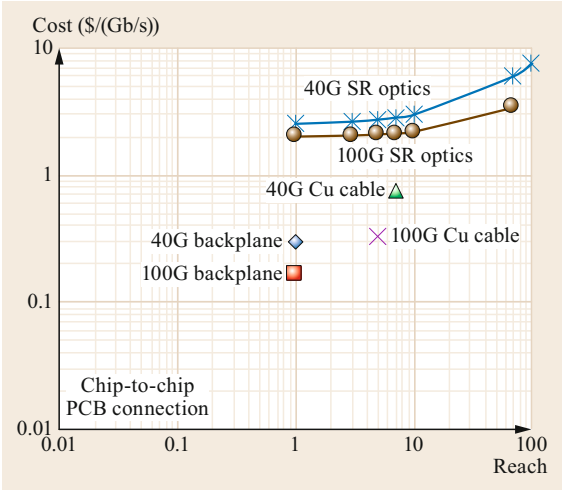


Fig. 21.7 Cost and reach comparison for electrical and optical interconnection for 40G and 100G

tivates a strong focus on improving cost, power, and density/size, at a variety of reaches.

Figure 21.7 shows the cost comparison for electrical and VCSEL-based short-reach optics at 40G (4×10 Gb/s) and 100G (4×25 Gb/s). Printed circuit board (PCB) traces are used for the switch I/O connections between chip-to-chip, chip-to-module, and backplane interconnects up to 1 m reach. Although there has been much anticipation and debate regarding optical intra-rack connections, for the data rates shown in Fig. 21.7, copper direct attach cable (DAC) still seems to be the most cost-effective and power-efficient interconnect for intra-rack connections up to 3 m at 25 Gb/s. Vertical-cavity surface-emitting laser (VCSEL)- and multimode fiber (MMF)-based short-reach technologies have shown the best overall link cost (transceiver cost plus fiber cost) for up to 25 Gb/s per lane speed and for reaches up to 100 m.

VCSEL- and MMF-based technologies have been widely deployed in Google datacenters since 2007 [21.1]. This has tipped the scales away from copper interconnects, which were bulky and very difficult to deploy at the number and lane speed scales required for inter-rack connection. Over the past 10 years, the performance of VCSEL arrays has improved to extend the reach to 100 m at 40 Gb/s (4×10 Gb/s) over OM3 fiber and 100 m at 100 Gb/s (4×25 Gb/s) over OM4 fiber. Beyond 50 Gb/s lane speed, due to the performance limitations on bandwidth and reach, it may be challenging for VCSELs to scale for 100 Gb/s PAM4 without advanced digital signal processing (DSP).

The interconnections between the edge aggregation switches and the spine switches form the core of the network fabric, which can span an entire mega-datacenter building or campus. Long-reach optical transceivers and single-mode fiber (SMF) transmission technologies must be used to achieve the required bandwidth reach and cabling efficiency. In 2012, Google achieved the first large-scale deployment of wavelength-division multiplexing (WDM)/single-mode fiber (SMF)-based interconnects inside the datacenters with 40 GbE quad (4-channel) small form-factor pluggable (QSFP) form-factor transceivers [21.1]. Using wavelength-division multiplexing and photonic integration, the solution can scale to higher data rates of 400 Gb/s, with four wavelengths each running at 100 Gb/s PAM4 modulation, and longer reach (> 2 km).

Next Generation: Intra-Rack Interconnect

For traditional on-off keying (OOK) modulation, beyond 50 Gb/s lane speed, the reach of signal transmission over copper is limited by the large loss at high frequency. Pulse-amplitude modulation (PAM) signaling overcomes the loss limitation at the expense of signal-to-noise ratio (SNR) penalty due to multilevel signaling. Lower-loss PCB materials and DSP-powered PAM4 (four-level PAM) SerDes may be needed to enable low-cost PCB interconnects. Because of the higher power consumption of SerDes I/O required to compensate for the physical impairments (loss, inter-symbol interference (ISI), reflection, etc.), achieving a PCB backplane trace of 1 m, which is typically needed for interconnecting line cards, is challenging. Although

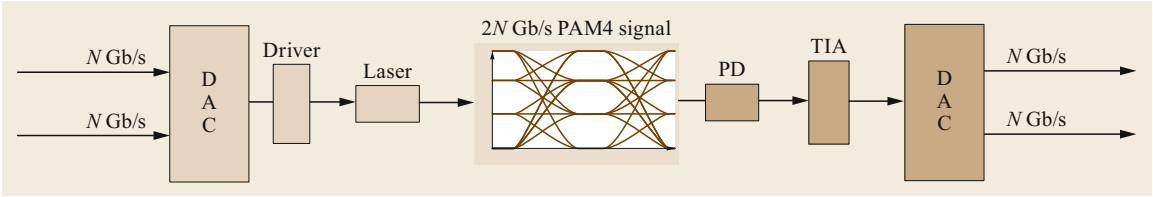


Fig. 21.8 Pulse-amplitude modulation-based link (specifically PAM4, four levels). Leveraging of technological developments for the line side (long-haul), in the form of high-speed ADC (analog-to-digital converter) and DAC (digital-to-analog converter) design, allows increased bandwidth with the same baud rate and number of optical components

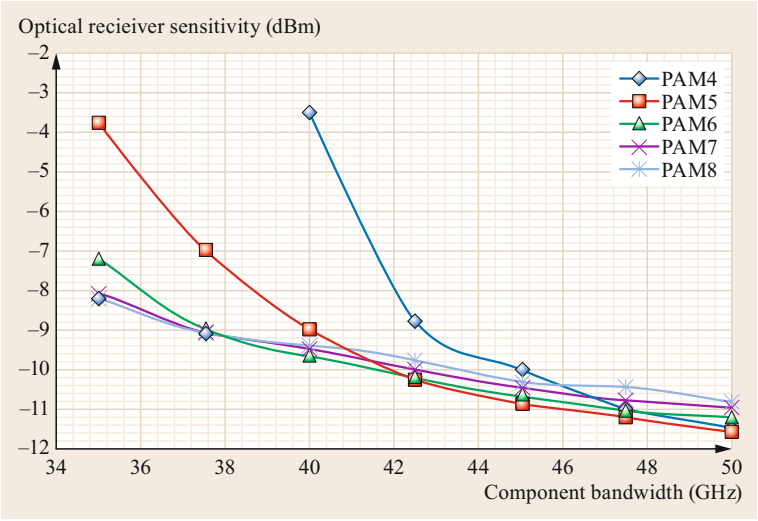


Fig. 21.9 Achievable optical receiver sensitivity using different PAMs with a net throughput of 200 Gb/s for a pragmatic short-reach optical system (thermal noise and bandwidth limited) (after [21.9])

high-speed I/O can scale beyond 100 Gb/s, the electrical reach may be too short even for chip-to-module connection with good energy efficiency. This may necessitate the closer integration of optics to electronics.

Next Generation: Inter-Rack Interconnect

The next-generation electrical I/O will be PAM4-based, with linear interfaces for the optical transmitter and receiver. Technologies developed for modern coherent systems used for dense wavelength-division multiplexing (DWDM) line-side optics (such as linear laser drivers (LD) and linear transimpedance amplifiers (TIAs) at the front end and high-speed analog-to-digital converter (ADC) and digital-to-analog converter (DAC)) form the necessary basis of next-gen high-speed datacenter transceivers. In addition, high-performance DSP capability has emerged for datacenter applications in order to relax the bandwidth and linearity requirements for optical components and to compensate for other link impairments [21.10]. Compared with long-haul transmission, the power consumption of the DSP functions could be significantly reduced for the much shorter-reach datacenter applications with the elimination of chromatic/polarization mode dispersion (CD/PMD) and polarization/phase recovery to meet the power and density requirements. Another potential area of development is low-latency and low-overhead forward error correction (FEC) specifically targeted for datacenter applications.

To scale the performance of direct-detection PAM signaling beyond 100G PAM4, PAM-N technology may be extended by a recently proposed FlexPAM concept [21.9], where irregular PAM (such as PAM5) with flexible and fine spectral efficiency (SE) granularity can be realized by using a single chip with a common DSP architecture. The need for FlexPAM with fine SE granularity can be clearly seen from Fig. 21.9, which shows the achievable receiver sensitivity of different PAMs versus component bandwidth (by using Shannon mutual information theory) to achieve 200 Gb/s throughput for a pragmatic externally modulated short-reach optical system. A dramatic improvement in performance can be achieved with a slight increase in the modulation bandwidth efficiency for a bandwidth-limited system. For components with achievable bandwidth of 40 GHz, the receiver sensitivity can be improved by 5.5 dB using irregular PAM5 compared with regular PAM4. Although using regular PAM8 can achieve similar receiver sensitivity, PAM8 has a 5–6 dB higher link penalty than PAM5, due to level-dependent impairments such as multiple-path interference (MPI) [21.9]. Such a technique may be useful for scaling of short-reach systems to beyond 1 Tb/s.

The use of coherent transmission to increase data rate per channel/wavelength has been previously discussed as another possibility for intra-datacenter interconnect beyond 100 Gb/s per channel [21.11]. The benefits of a coherent approach include the use of a single laser for multiple degrees of freedom (in-phase, quadrature, and polarization) and increased receiver sensitivity. The sensitivity advantage is shown in Fig. 21.10. Operating at an identical net symbol rate of 50 GBd, 480 Gb/s coherent PM-16QAM, with moderate 13 dBm local oscillator (LO) power, can achieve 15 dB better sensitivity than 112 Gb/s PAM4 with direct detection.

On the downside, the DSP functions required to extract the various degrees of freedom at the receiving end consume more power than with a direct-detection receiver. Therefore, it has traditionally been used only

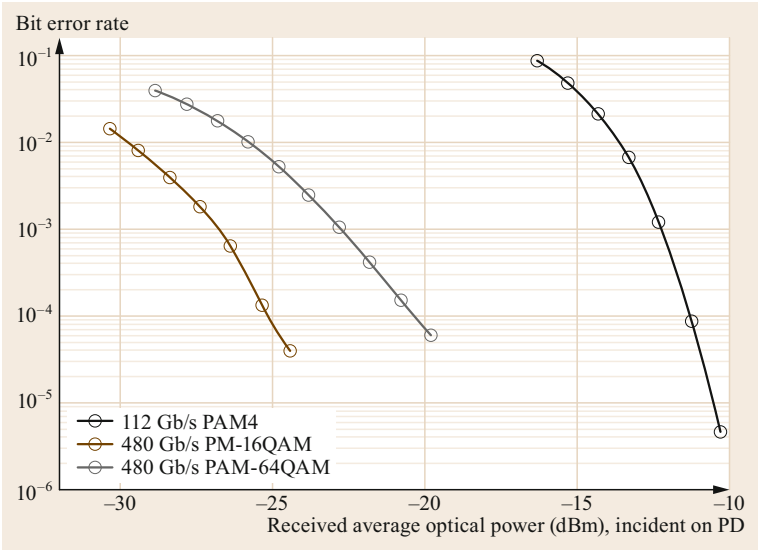


Fig. 21.10 Receiver power sensitivity comparison between coherent and direct detection

in long-haul and metro networks due to its implementation complexity and high power consumption. Other disadvantages are the increased requirements for the quality and control of components (modulator and laser), doubling the number of photodetectors, and test complexity. Finally, although not fundamental, the technology has not been developed for traditional datacenter wavelengths, affecting deployability. So far, disadvantages have outweighed advantages for intra-datacenter use of digital coherent technologies.

However, there have been enormous reductions in cost, power, and density with respect to coherent technologies for metro and long-haul transport systems over the last decade. Figure 21.11 shows power and linear density per gigabit per second as a function of time for datacenter transceivers and coherent transceivers (excludes external erbium-doped fiber amplifiers, EDFAs). Some estimates for target 200/400 Gb/s transceivers include 200 Gb/s (4×50 Gb/s PAM4) in QSFP for the datacenter, and 400 Gb/s in octal (8-channel) small form-factor pluggable (OSFP) for both datacenter and coherent (1×400 Gb/s) transceivers [21.12]. From these plots, it is clear that datacenter and coherent interconnect technologies are converging from a power and density perspective, with this trend expected to further accelerate with shorter coherent reaches (< 100 km, triangles in Fig. 21.11) shedding unnecessary DSP functions. The extension of coherent technologies to shorter reaches will likely enable aggregate interconnect cost reduction (inter- and intra-datacenter reaches) with shared volumes. In addition to relying on Moore's law for power reduction, low-voltage modulators and uncooled LOs need to be investigated to meet the power/density target for datacenter optics. For scaling beyond 1.6 Tb/s, coherent detection could be a more viable option than direct detection.

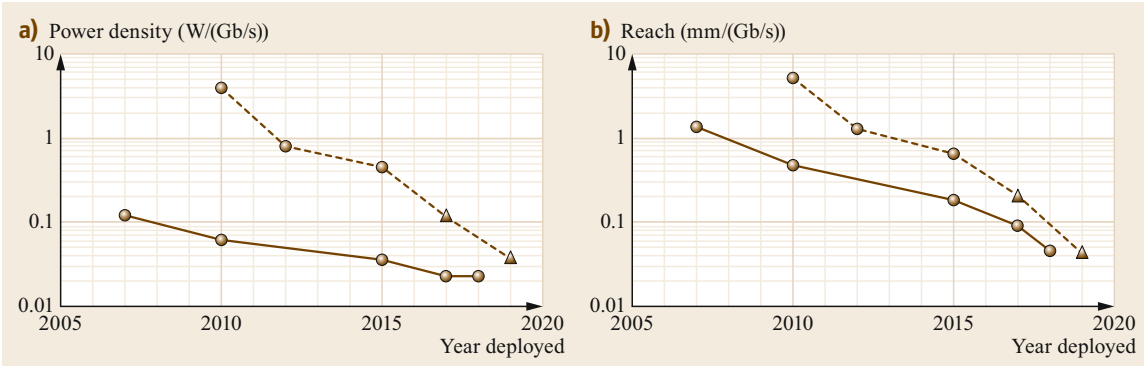


Fig. 21.11a,b Comparison of datacenter (*solid lines*) and coherent transceiver (*dashed lines*) power per Gb/s and linear density per Gb/s. *Triangles* indicate shorter reaches (< 100 km)

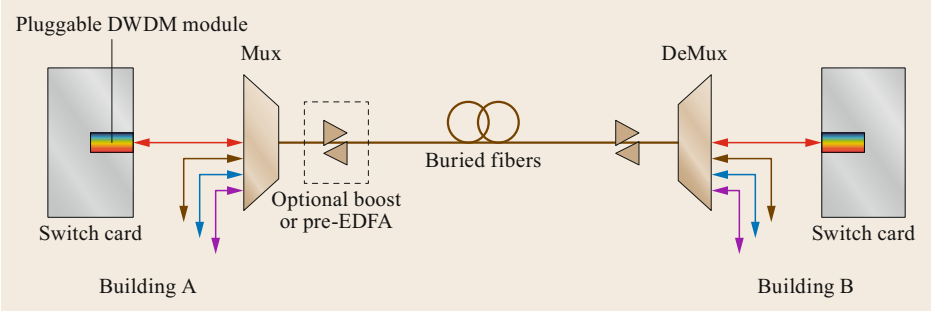


Fig. 21.12 Illustration of pluggable DWDM technique for intra-campus interconnection

Another option for increasing the aggregate data rate per interface is to increase the number of wavelength channels. With this approach, WDM integration will be critical, with three possible approaches: silicon photonics, monolithic III–V, or a hybrid solution. For silicon photonics, the main challenges are link budget, packaging, and optical and electrical integration. Grating and edge couplers with lower loss and improved alignment tolerance are essential to improving performance, yield, and cost. Electrical integration will minimize parasitics to reduce power and size. Silicon photonics may leverage the packaging technologies and techniques of traditional CMOS (f2 f, 2-D, 2.5-D, 3-D integration), giving it critical advantages. Monolithic III–V photonic integrated circuits need fundamental improvements in epitaxial growth and fabrication. Reduction and/or identification of defects early in the III–V process may be aided by data mining and better analytics.

Intra-Campus Interconnect

There are always trade-offs among spectral efficiency, power consumption, path diversity, and cabling complexity. For the intra-building network, a connection-rich mesh topology is desirable; hence, coarse wavelength-division multiplexing (CWDM) with lower spectral efficiency has been the technology of choice for achieving lower power, cheaper transceiver cost, and a richer network fabric. On the other hand, at higher aggregation layers such as the inter-building network, bandwidth is more concentrated over point-to-point links, and dark fiber is more expensive to deploy; hence, DWDM with higher spectral efficiency is preferred. As datacenter traffic continues to grow, the quantity of fibers required from building to building becomes difficult to scale and manage [21.13]. More spectrally efficient and pluggable DWDM technology is needed to scale the intra-campus bandwidth. The use of pluggable DWDM technology not only greatly reduces the number of fibers needed (e.g., a 100 GHz-spaced 48-channel DWDM system reduces the required number of fiber connections by a factor of 48), but also allows simpler fiber management. The higher optical link budget supported by the use of optical amplifiers may also enable new network-level functionalities such as larger campuses or intermediate switching nodes.

Density and cost are two key factors that determine the applicability of DWDM technologies for campus applications. Additionally, to facilitate deployment and inventory management, full-band wavelength tunability is highly desirable. DWDM technology requires stable wavelength grids, and this often necessitates temperature-stabilized (thermoelectric (TE)-cooled) packaging. Innovative laser chip and package designs are needed to bring down cost and power consumption, especially for full-band tunable lasers.

21.1.4 Throughput Requirements and Traffic Characteristics

Figure 21.13 shows the traffic generated by servers inside Google's datacenter, which has been doubling every 12–15 months [21.1]. This is much faster than the bandwidth growth of private backbone WANs, with 40–50% growth every 12 months [21.2].

Bandwidth provisioning to servers is a trade-off between performance (latency, energy) and cost efficiency. While various software techniques can be used to allocate resources more efficiently [21.14, 15] so as to extract the most value and efficiency of compute and storage resources, the networks that interconnect them must be correctly balanced. Otherwise, one or two resources would limit the value that could be extracted from other resources, or could cause some resources to be idle, and that would increase system costs. According to Amdahl's balanced system law, a system should have 1 Mb/s of I/O for every 1 MHz of computation in a parallel computing environment. Ideally, the network bandwidth would be slightly over-provisioned relative to both compute and storage to ensure there are no issues in efficiently connecting compute and storage.

Unlike compute and storage, networking bandwidth is a distributed resource. Despite the availability of software techniques to improve the dynamic allocation of networking bandwidth at the datacenter scale [21.16–

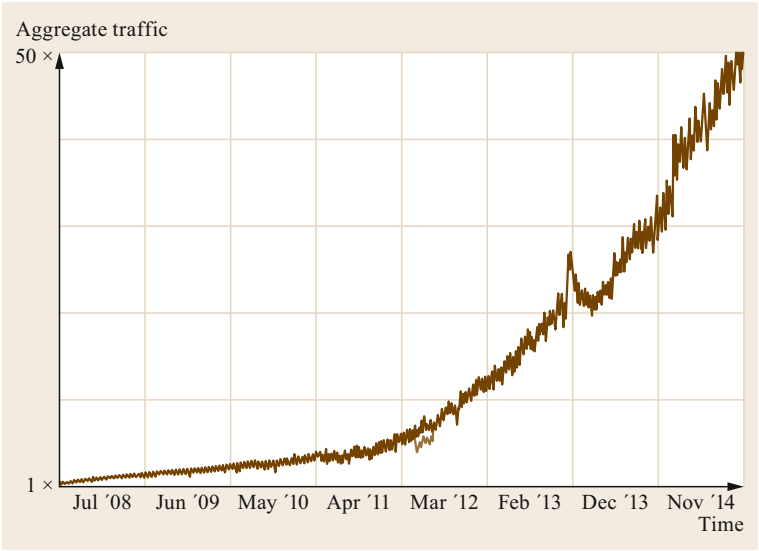


Fig. 21.13 Google's traffic growth generated by servers inside datacenters

18], reliable and scalable networking bandwidth sharing remains challenging even in network topologies with no oversubscription. Empirically, the bandwidth to the server can often be adjusted at the ToR by changing the oversubscription ratio and monitoring the utilization of the uplinks. For a rack of m servers, each capable of 100 Gb/s of burst bandwidth, if the ToR has $m \times 100$ Gb/s downlinks and $n \times 100$ Gb/s uplinks, the average bandwidth available to each server is $(n/m) \times 100$ Gb/s (with an oversubscription of $m : n$).

The total computing and communication capacity required for each target application also vary widely. Consider Google Search as an example: it often touches 50+ services and 1000+ servers with distributed data stored in memory or flash. Caching service (memkeyval) requires a large amount of networking bandwidth for copying of large files, while score and sort (websearch) consumes little network bandwidth [21.14]. With the adoption of flash today and other nonvolatile memory (NVM) technologies with large I/O capability in the future, designing a network to efficiently maximize the performance of compute and storage resources will remain a key challenge.

21.1.5 Energy Efficiency of Datacenter Network

With the unprecedented growth in users and traffic, datacenters continue to expand to house more machines and networking gear. The cost of power and its associ-

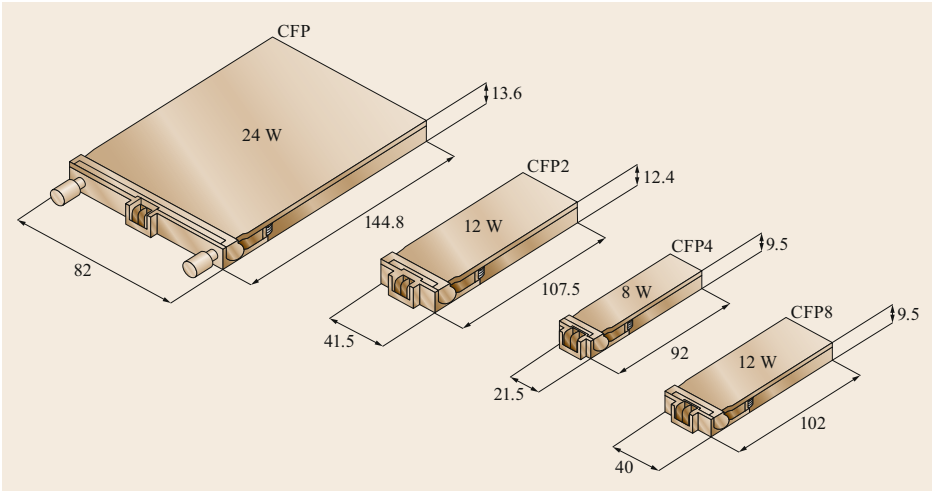


Fig. 21.14 Various form factors for 100 GbE pluggable transceivers and corresponding power envelope (www.cfp-msa.org)

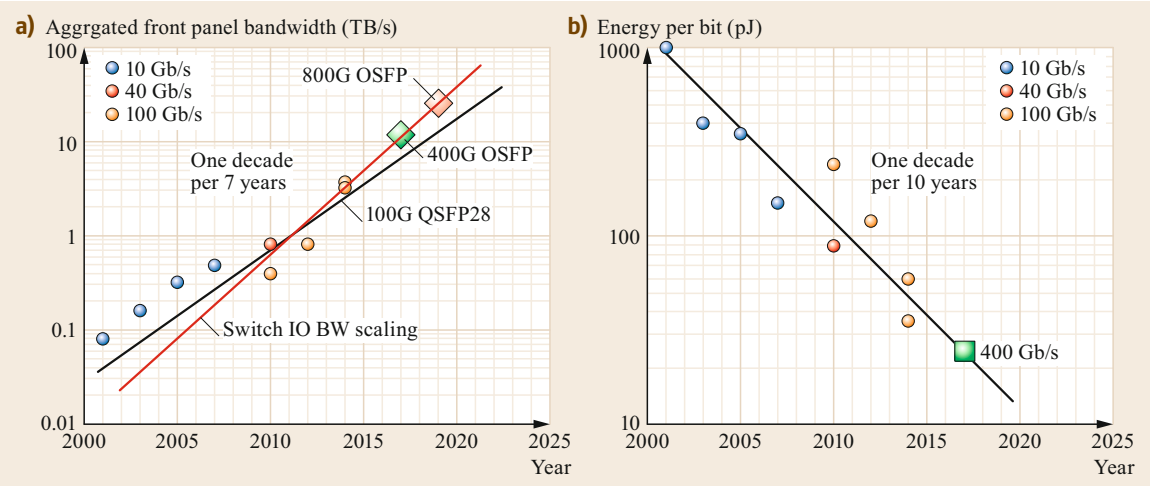


Fig. 21.15 **(a)** Front panel bandwidth of pluggable datacenter optical transceiver; **(b)** energy to transport a bit over 2 km intra-datacenter

ated delivery and cooling are becoming significant factors in the total expenditures of large-scale datacenters. Power consumption inside the datacenter is distributed in the following order [21.19]: servers consume around 33%, memory 30%, power delivery and building cooling 20%, disks 10%. Over the past 10 years, significant effort has been spent in improving energy efficiency, from the individual components to building-level design. Besides the improvement in cooling and power delivery systems, low-power hardware components and energy-proportional designs enable energy savings for processors [21.20], memory [21.21], storage devices, and networking systems [21.22].

The networking portion constitutes only 7% of the entire power envelope of the datacenter [21.19]. The portion of networking attributable to optics is a smaller subset of that power (half or less). Thus, a reduction in networking power consumption will not yield a dramatic improvement in overall datacenter power efficiency and corresponding reduction in CapEx and OpEx. However, the reduction of power in optical modules is still important for increasing front panel linear density and reducing the overall datacenter fabric footprint and cost. In practice, the switch line card/system's form factor is limited by the transceiver's power consumption and the electrical feeds to the transceiver. The size constraint on optics comes from the front panel linear density ((Gb/s)/mm) of the line card, where all the bandwidth of the switch ASIC(s) housed on the line card must be extracted via the front panel. Figure 21.14 shows some pluggable optical transceiver standard form factors for 100 GbE and the associated power consumption of each. Obviously, with the smaller form-factor solutions with small connector pin pitch, more bandwidth can be brought out per linear dimension of the line card. However, the small-form-factor optical transceiver often has a smaller power envelope because of a more limited surface area for heat extraction.

Figure 21.15a shows the switch I/O bandwidth trend line and the front panel bandwidth achieved for 10, 40, and 100 Gb/s. Pluggable optics technology saw a tenfold improvement over 7 years, with an increase in electrical I/O speed (from 10, 25, to 50 Gb/s) and a decrease in connector pin pitch from 0.8 mm for (40 Gb/s)/(100 Gb/s) QSFP to 0.6 mm for 400 Gb/s OSFP. Thus far, through the use of 400 Gb/s transceivers, there has been no scaling problem for datacenter pluggable optics to effectively match and extract the switch capacity.

Delving into what determines power at the component level can be quite complex, as the type of device and the design details give rise to a wide range of power numbers. Because of the analog nature of optical components, even with the advances in high-speed SerDes technologies, the power efficiency of datacenter optical transceivers improves only 10× every 10 years (Fig. 21.14b), which is slower than the growth of switch bandwidth capacity and front panel density. Improvements in thermal design for pluggable optics has helped to sustain bandwidth density improvements in the past. More aggressive thermal management and creative copackaging of electrical and optical devices will be needed to meet the more stringent bandwidth density requirements for port speed of 400 Gb/s and beyond.

21.2 Inter-Datacenter Network

While economies of scale suggest that datacenters should be as large as possible, typically restricted by the amount of power available for the site, datacenters should also be distributed across the planet for fault tolerance, latency locality, and better user experience. The rapid adoption of cloud services also drives a new set of requirements and challenges for wide area networking (WAN) in terms of capacity, accessibility, and flexibility. To meet the two sets of requirements, Google has built two separate WAN networks: public back-end backbone (B2) and private back-end backbone (B4). Figure 21.16 shows Google's wide area datacenter network topology [21.22]. Parallel to the private back-end network is a user-facing public backbone (B2), which end users interconnect to through peering, transit, or direct connection to gain access to various cloud services. In order to support better user experience with lower latency, the fiber topology of the public backbone is meshier, with more stringent availability performance. The private back-end backbone (B4) provides connectivity among datacenters only, to support large-scale data copies, remote storage access, and end user data replication for reliability. The private back-end backbone network is usually architecturally simple, with point-to-point links, but requiring much higher bandwidth to support large-volume internal application traffic [21.2].

The inter-datacenter private back-end network employs scarce and expensive long-distance fiber for transmission. The public-facing transport network, in addition, contains many high-capacity metro transport links to interconnect with other carriers' networks. Metro transport networks also serve to connect carrier networks to edge cache systems used by content providers in order to improve content distribution experiences with low latency without burdening the expensive backbone transport network. In recent years, fast-growing bandwidth-intensive services such as YouTube and Netflix and other cloud applications are accelerating the deployment of edge cache and metro optical transport systems [21.22].

21.2.1 Interconnect Network Architecture Evolution for WAN

With traffic engineering or buffering at the edge, the aggregated WAN traffic entering and leaving datacenters can be controlled based on capacity availability [21.2]. Unlike fibers inside datacenters, long-haul transmission fibers are expensive and time-consuming to build or acquire. Datacenters are often located in remote areas, which drives the requirement for a range of reaches all the way from a few hundred kilometers to 6000 km. Given the space and power-hungry nature of regeneration nodes, the focus of inter-datacenter links is on high-capacity, high-spectral efficiency, ultra-long-haul transmission systems. Considerable hardware and soft-

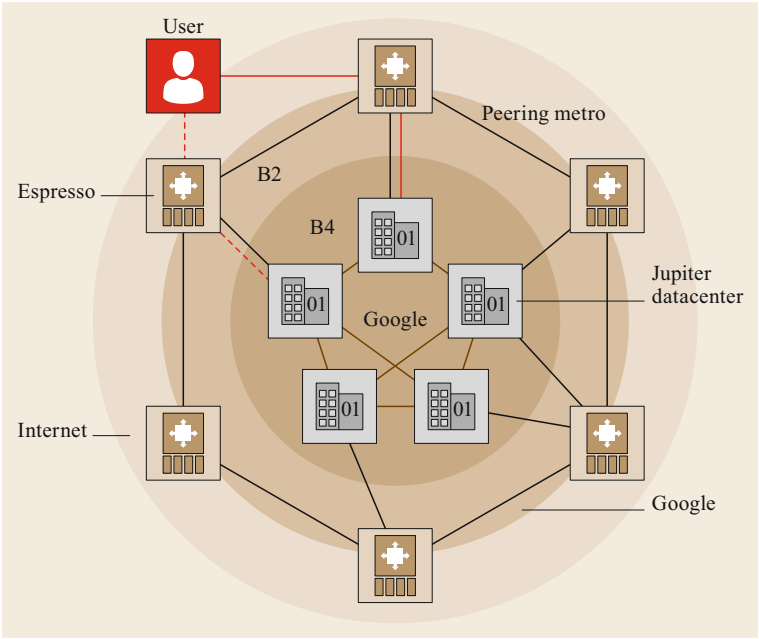


Fig. 21.16 Google's wide area network (WAN) architecture (after [21.23])

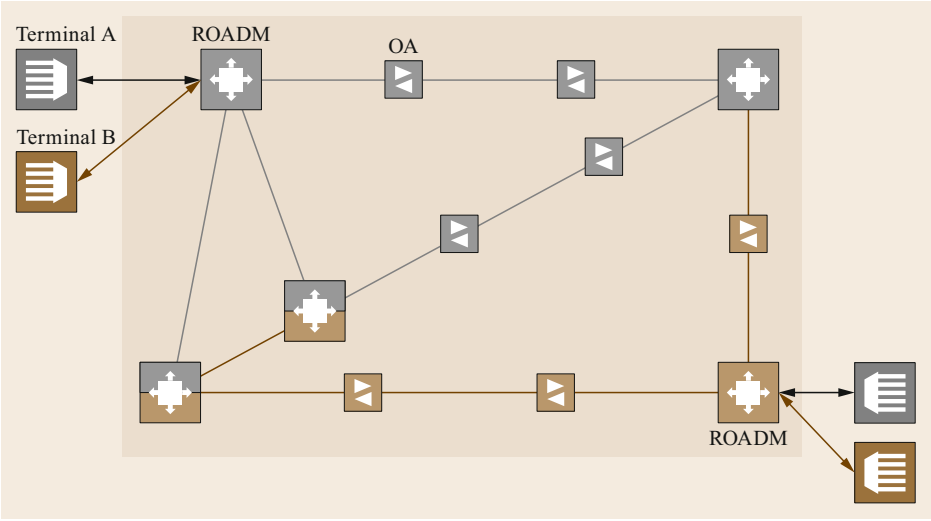


Fig. 21.17 Reconfigurable optical add-drop multiplexer (ROADM)-enabled mesh transport architecture (after [21.24]) for long-haul networks. (OA: optical amplifier)

ware [21.2] innovation and effort have been directed toward improving spectral efficiency and utilization of WAN links.

From a transport topology point of view, while metro networks (the network that connects our datacenters and peering POPs to our B2/B4 backbone networks) still use point-to-point link-based architectures, long-haul (LH) networks have evolved from traditional point-to-point link-based architectures to mesh architectures enabled by colorless, directionless, and contentionless (CDC) ROADM (reconfigurable optical add-drop multiplexer). With software control, CDC ROADMs can add-drop any wavelength from any direction to any transponder without contention in the add/drop path, thereby enabling an unconstrained dynamic network at the physical layer.

The CDC ROADM-based mesh network architecture has a number of advantages: (a) higher network utilization efficiency, (b) shorter reconfiguration time, and (c) operational simplicity:

(a) Network efficiency comes from a couple of areas. An arbitrary wavelength-routed mesh provides superior CapEx and OpEx performance relative to terminated point-to-point links. Secondly, CDC ROADM architecture enables bandwidth reconfiguration to meet the variation in network demand, unlike the traditional point-to-point network topology, where additional fibers and amplifiers must be deployed to transport those wavelengths or in the event of a bandwidth demand surge.
(b) CDC ROADM architecture enables optical layer reconfiguration around links under repair to reduce the mean time between failures, thus increasing the overall network availability. CDC-ROADM-enabled mesh networks could provide section-by-section (between ROADM nodes) $1:n$ share optical protection as well as dynamic mesh restoration (the capability to dynamically identify and activate new routes at the time of a failure). In particular, $1:n$ protection allows us to protect optical layer services without the need for reserving 50% of the network capacity, which is required with $1+1$ protection used in point-to-point-based networks.
(c) The use of CDC ROADM could enable end-to-end *touchless* provisioning and activation of network bandwidth without requiring technicians on-site to provision the connection of each circuit. Moreover, the sparing of transponders in the CDC add/drop node becomes a simple software command to activate a prepopulated spare instead of physically reseating a new card and reprovisioning it.

In addition to the introduction of CDC ROADM for a more flexible optical layer, disaggregation of key functional blocks is emerging as an important trend [21.24]. Traditionally, LH transport systems have been provided by system vendors as proprietary solutions, where the terminal optics (i.e., line card-based optical transponders) will work only with their proprietary line systems. The key functional blocks, such as ROADM and OA, can only be managed by their proprietary line management system, and have numerous control and equalization loops at the link level. Such an integrated solution has the advantage of well-defined transport performance, but it also has challenges in interoperability due to lack of common specifications, flexibility in vendor and technology selection, and

centralized network management and automation. The disaggregation of key transport functional blocks with a well-defined I/O interface addresses these problems. The first phase of disaggregation is to decouple terminal optics from the line system, which we have deployed for several years. In the second phase, ROADM disaggregation into key functional blocks provides a number of benefits, which include (a) modularity at the degree and link levels and the use of multivendor line systems, (b) faster insertion cycles for new subsystem building blocks, and (c) a greatly simplified control and management plane architecture by embodying modern software-defined networking (SDN) principles.

The ROADM technology has also been evolving from fixed grid (typically 50 GHz or 100G fixed DWDM grid) toward a flexible grid (with grid granularity reduced from 50 to 12.5 GHz). The introduction of flexible-grid ROADM could reduce the guard band between channels (e.g., 100 Gb/s quadrature phase-shift keying QPSK over 37.5 GHz channel spacing) and pack multiple channels together with minimal channel guard bands through the Nyquist pulse-shaping technique [21.25], improving spectral efficiency in long-haul transmission fiber.

21.2.2 LH Interconnect Bandwidth Scaling

Since long-haul fiber is a very scarce resource and time-consuming to deploy, in order to meet the increased bandwidth demands of private backbone WAN, the bandwidth capacity scaling of long-haul is achieved mainly by increasing the single-fiber capacity. Figure 21.18 shows an increase in single-fiber capacity from 3.2–40 Tb/s with the following techniques.

Dense Wavelength-Division Multiplexing (DWDM)

DWDM channels with 50 GHz spacing (with up to 96 channels over the extended C-band optical amplifier bandwidth) have been used for LH transport systems with 40–200 Gb/s per channel. Very tight laser wavelength control is required for DWDM systems with small wavelength spacing, because any wavelength drift will cause the system to experience higher loss and spectrum distortion from the filtering effects of Mux/Demux (multiplex/de-multiplex) (and cascaded ROADMs), as well as larger DWDM crosstalk from the neighboring channels. Wavelength control accuracy has improved from early DWDM systems with ±5 GHz accuracy, to current accuracy of ±2.5 GHz, in order to reduce the filtering and crosstalk penalties. Flexible-grid or gridless super-channels may be used to increase the optical spectrum utilization and thus the overall single-fiber capacity.

Four-Dimensional Modulation/Multiplexing

Each optical carrier has four degrees of freedom: X and Y-polarization, in-phase, and quadrature phase components. Polarization-multiplexed (PM) QPSK and PM M-QAM (quadrature amplitude modulation) such as PM-8QAM and PM-16QAM utilize all four dimensions of an optical carrier, and thus can increase the modulation spectral efficiency in terms of bits per symbol by a factor of 4 as compared with the single-polarization direct intensity modulation format with one dimension of an optical carrier only.

Digital Coherent Detection

Digital signal processing (DSP) has enabled coherent detection, where both polarization demultiplexing and

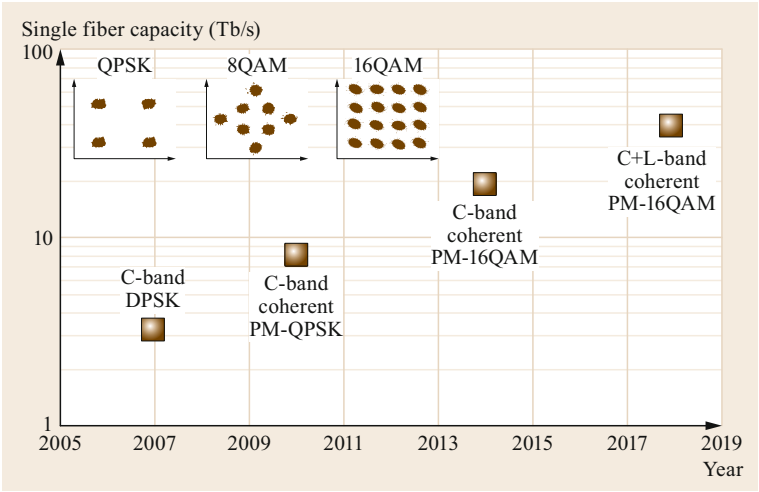


Fig. 21.18 LH network single-fiber capacity (with reach > 600 km) scalings. (PM: polarization-multiplexed, DPSK: differential phase-shift keying; QPSK: quadrature phase-shift keying; QAM: quadrature amplitude modulation)

carrier phase recovery are carried out in the digital domain by:

1. linearly mapping the incoming optical signal field into an electrical signal through a polarization- and phase-diverse coherent detection front end
2. digitizing the detected electrical signal; and
3. using a 2×2 multi-input and multi-output (MIMO) equalizer to recover the polarization and using a feed-forward-based phase estimation algorithm to track fast-changing optical phases.

Digital coherent detection not only encodes information into the full four dimensions of an optical carrier; it also enables electrical compensation of optical chromatic dispersion and polarization mode dispersion, which are the major impairments for high-speed optical transmission.

C+L Band

C-band (1528–1565 nm) and L-band (1570–1610 nm) are two communication bands that coincide with the third transmission window of silica fiber and the wavelength range of EDFAs. EDFAs can be made to operate at both C-band and L-band, and fiber loss in the L-band is still very low. When C-band capacity becomes insufficient, L-band is the next natural spectral region to double the single-fiber capacity.

Eventually, fiber nonlinear effects and OSNR (optical signal-to-noise ratio) requirements for higher-spectral-efficiency (i.e., higher bits per symbol) modulation formats [21.27] limit the single-fiber capacity. Figure 21.19 shows the estimated transmission reach limit for different higher-order modulation formats (with different spectral efficiency rates) at a fixed symbol rate of 40 GBd and fixed channel spacing of 50 GHz. The transmission reach rapidly decreases as the modulation spectral efficiency increases. Both hardware [21.26, 28, 29] and software techniques [21.2] designed to improve the resource utilization of network capacity have been investigated in recent years. Rate-adaptive optics [21.26], which optimizes the fiber capacity based on fiber route, can effectively increase network resource utilization: shorter routes using higher-SE (spectral efficiency) modulation formats (and thus higher capacity), and longer routes using lower-SE modulation formats (less capacity). Very fine-granularity rate-adaptive optical transceivers could be realized by using either the time-domain hybrid QAM technique [21.26] or the probabilistic constellation-shaping (PCS) technique [21.29]. The time-domain hybrid QAM has the advantage of simpler implementation, while the PCS offers better performance. As discussed earlier, the overall transport capacity can also be increased with Nyquist pulse-shaping-based super-channel techniques to increase spectrum utilization [21.25].

SDN has advanced greatly over the past 10 years. Not only does it enable the network-level programmability to make the above rate-adaptive optical transmission feasible, but the centralized optical control plane is also essential for allocating bandwidth among different services based on application priority. As a result, instead of over-provisioning and underutilization of bandwidth, the utilization of expensive long-haul links can be improved to near 100% [21.2].

Metro to LH: IP+DWDM to IPoDWDM

Traditional terminal optics-based transport technology has been used for metro and LH backbone networks. With this technique, the transport interface rate (i.e.,

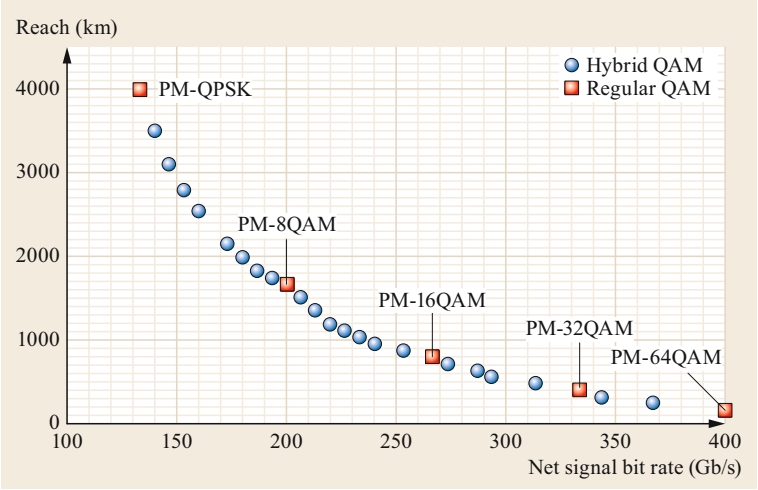


Fig. 21.19 Estimated transmission reach for different modulation formats all operating at 40 GBd symbol rate and 50 GHz channel spacing (after [21.26])

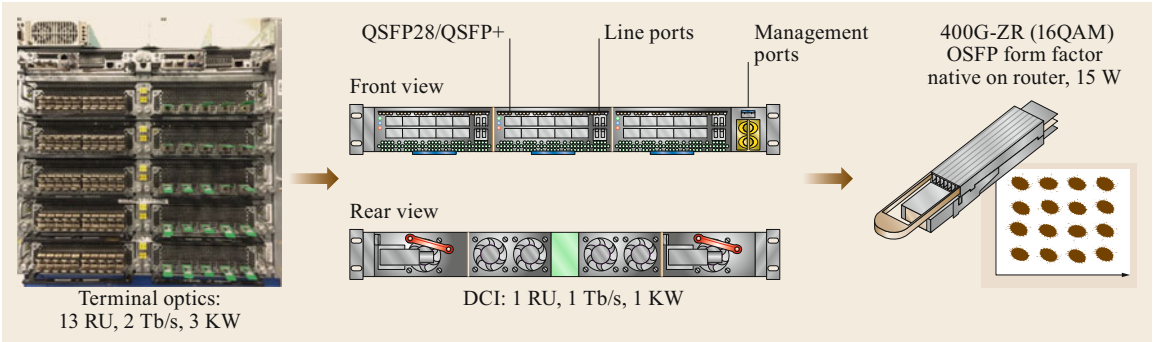


Fig. 21.20 Metro transport optics evolution (an example)

the line-side interface rate) is decoupled from the router interface rate (i.e., the client rate), with client optics interconnecting the router and the transport equipment. Such a technique was adopted mainly for two reasons: (1) the router interface rate is slower than the transport line-side interface rate, and (2) the router technology scales well with Moore's law, with much lower power consumption than transport optics. Thus the speed and density of router ports have evolved much faster than the transport optics. With the decoupling of the transport optics and router on separate systems, the router and transport optics can scale at their own pace to achieve the best overall economics. As the router interface rate increases to be comparable to DWDM line-side interface rates (100 Gb/s and beyond), not only is the cost of client optics greatly increased (relative to the line-side optics), but the role of traffic grooming using client optics is also reduced. On the other hand, with the advance of digital coherent detection, the transport technology evolution becomes largely dictated by the coherent DSP power, which has scaled well with Moore's law—the same as the routers. Therefore, starting from 100 Gb/s, a fundamental paradigm shift has taken place in the metro and LH networks: we are starting to move away from the traditional IP+DWDM architecture, toward a converged IP (internet protocol) over DWDM (IPoDWDM) architecture, by eliminating the client optics and putting the transport optics directly into the router interface ports. Such an IPoDWDM architecture not only saves power, cost, and space, but also simplifies network management and control.

Figure 21.20 shows the evolution of metro transport optics from traditional terminal optics to a lean datacenter connection interconnect (DCI) box with simplified management interface, and eventually to the pluggable 400G coherent module (i.e., the 400G-ZR, an Optical Internetworking Forum (OIF) standard which is under active development). As shown in the figure, traditional line card-based terminal optics occupies 13 rack units and consumes 3 kW of power for 2 Tb/s of transport capacity, whereas the 1 Tb/s DCI box occupies one rack unit and consumes 1 kW of power. When the 400G-ZR pluggable transceiver is available in 2019, its coherent optoelectronics is expected to consume < 15 W power and will be realized in an OSFP module [21.30], the same form factor that will be used for 400 Gb/s client optics. Such dramatic power and size reduction is made possible by the continuous power reduction with CMOS technology (40 to 28, to 14, to 7 nm) and the advancements in photonic integrated circuits (PICs), including both InP- and silicon photonics-based integration technologies.

21.3 Conclusion

The evolution of the datacenter network and its photonic technologies over the past decade has been driven by tremendous advances in hardware and software technologies, both within (intra-datacenter) and between (inter-datacenter) networks. The tenets of SDN will continue to have a major impact by enabling scalable, dynamic networks with better programmability and efficiency. The line-side technologies (wavelength-division multiplexing, digital signal processing) have migrated from traditional telecom applications to help drive the bandwidth growth of datacenter applications. Moving forward, the continuing growth of datacenters and cloud applications will become the key technology driver for more aggressive reductions in power, cost, and density through intimate integration of electronics and photonics.

References

21.1 A. Singh, J. Ong, A. Agarwal, G. Anderson, A. Armistead, R. Bannon, S. Boving, G. Desai, B. Felderman, P. Germano, A. Kanagala, H. Liu, J. Provost, J. Simmons, E. Tanda, J. Wanderer, U. Hölzle, S. Stuart, A. Vahdat: Jupiter rising: A decade of clos topologies and centralized control in Google's datacenter network, Commun. ACM **59**(9), 88–97 (2016)

21.2 S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Hölzle, S. Stuart, A. Vahdat: B4: Experience with a globally deployed software defined WAN. In: *Proc. ACM SIGCOMM* (2013), https://doi.org/10.1145/2486001.2486019

21.3 A. Vahdat, H. Liu, X. Zhao, C.J. Johnson: The emerging optical data center. In: *Proc. Opt. Fiber Conf.* (2011), https://doi.org/10.1364/OFC.2011.OTuH2

21.4 J. Kim, W.J. Dally, D. Abts: Flattened butterfly: A cost-efficient topology for high-radix networks. In: *Proc. 34th Ann. Int. Symp. Comput. Archit.* (2007), https://doi.org/10.1145/1250662.1250679

21.5 C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, C. Tian, Y. Zhang, S. Lu: BCube: A high performance, server-centric network architecture for modular data centers. In: *Proc. ACM SIGCOMM* (2009), https://doi.org/10.1145/1592568.1592577

21.6 C. Clos: A study of nonblocking switching networks, Bell Syst. Tech. J. **32**(2), 406–424 (1953)

21.7 W. Dally, B.P. Towles: *Principles and Practices of Interconnection Networks* (Morgan Kaufmann, New York 2004)

21.8 M. Al-Fares, A. Loukissas, A. Vahdat: A scalable, commodity, data center network architecture. In: *Proc. ACM SIGCOMM* (2008), https://doi.org/10.1145/1402946.1402967

21.9 X. Zhou, H. Liu: Constellation shaping: Can it be useful for datacenter reach communication. In: *Eur. Conf. Opt. Commun.* (2017), Paper WS4

21.10 R. Urata, H. Liu: Datacenter interconnect and networking: Present state to future challenges. In: *Opt. Interconn. Conf.* (2016)

21.11 X. Zhou, H. Liu: Pluggable DWDM: Considerations for campus and metro DCI applications. In: *Eur. Conf. Opt. Commun.* (2016), Paper WS3

21.12 R. Urata, H. Liu, X. Zhou, A. Vahdat: Datacenter interconnect and networking: From evolution to holistic revolution. In: *Opt. Fiber Conf.* (2017), Paper W3G.1

21.13 U. Hölzle: Ubiquitous cloud requires a revolution in optics. In: *Opt. Fiber Commun. Conf.* (2017), Plenary talk

21.14 D. Lo, L. Cheng, R. Govindaraju, P. Ranganathan, C. Kozyrakis: Heracles: Improving resource efficiency at scale. In: *Proc. 42nd Ann. Int. Symp. Comput. Archit.* (2015), https://doi.org/10.1145/2749469.2749475

21.15 V. Jeyakumar, A. Kabbani, J. Mogul, A.Vahdat: Flexible network bandwidth and latency provisioning in the datacenter, arXiv:1405.0631 [cs.NI] (2014)

21.16 A. Shieh, S. Kandula, A. Greenberg, C. Kim, B. Saha: Sharing the data center network. In: *Proc. 8th USENIX Conf. Netw. Syst. Des. Implement. (NSDI)* (2011) pp. 309–322

21.17 H. Rodrigues, J.R. Santos, Y. Turner, P. Soares, D. Guedes: Gatekeeper: Supporting bandwidth guarantees for multitenant datacenter networks. In: *Proc. USENIX WIOV* (2011) p. 6

21.18 T. Benson, A. Akella, D.A. Maltz: Network traffic characteristics of data centers in the wild. In: *Proc. 10th ACM SIGCOMM Conf. Internet Meas. (IMC'10)* (2010), https://doi.org/10.1145/1879141.1879175

21.19 L. Barroso, U. Hölzle: *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines* (Morgan Claypool, San Rafael 2009)

21.20 L.A. Barroso, U. Hölzle: The case for energy-proportional computing, Computer **40**(12), 33–37 (2007)

21.21 K.T. Mallad, F.A. Nothaft, K. Periyathambi, B.C. Lee, C. Kozyrakis, M. Horowitz: Towards energy-proportional datacenter memory with mobile DRAM. In: *39th Ann. Int. Symp. Comput. Archit. (ISCA)* (2012), https://doi.org/10.1109/ISCA.2012.6237004

21.22 D. Abts, M.R. Marty, P.M. Wells, P. Klausler, H. Liu: Energy proportional datacenter networks. In: *Proc. 37th Ann. Int. Symp. Comput. Archit. (ISCA)* (2010), https://doi.org/10.1145/1815961.1816004

21.23 K.K. Yap, M. Motiwala, J. Rahe, S. Padgett, M. Holliman, G. Baldus, M. Hines, T. Kim, A. Narayanan, A. Jain, V. Lin, C. Rice, B. Rogan, A. Singh, B. Tanaka, M. Verma, P. Sood, M. Tariq, M. Tierney, D. Trumic, V. Valancius, C. Ying, M. Kallahalla, B. Koley, A. Vahdat: Taking the edge off with espresso: scale, reliability and programmability for global Internet peering. In: *ACM SIGCOMM Conf.* (2017), https://doi.org/10.1145/3098822.3098854

21.24 V. Vusirikala: SDN enabled programmable, dynamic optical layer. In: *Eur. Conf. Opt. Commun.* (2017), Paper M.PL 1

21.25 G. Bosco, V. Curri, A. Carena, P. Poggiolini, F. Forghieri: On the performance of Nyquist-WDM terabit superchannels based on PM-BPSK, PMQPSK,PM-8QAM or PM-16QAM subcarriers, J. Lightwave Technol. **29**(1), 53–61 (2011)

21.26 X. Zhou, L.E. Nelson, P. Magill: Rate-adaptable optics for next generation long-haul transport networks, IEEE Commun. Mag. **51**(3), 41–49 (2013)

21.27 R.J. Essiambre, R.W. Tkach: Capacity trends and limits of optical communication networks, Proc. IEEE **100**(5), 1035 (2012)

21.28 P.J. Winzer: Spatial multiplexing in fiber optics: The 10x scaling of metro/core capacities, Bell Labs Tech. J. **19**, 22 (2014)

21.29 X. Zhou, C. Xie: *Enabling Technologies for High Spectral-efficiency Coherent Optical Communication Networks* (Wiley, New York 2016)

21.30 OSFP MSA Group: http://osfpmsa.org/

Hong Liu

Google Inc.
Mountain View, CA, USA
hongliu@google.com

Hong Liu is a Distinguished Engineer at Google Technical Infrastructure, where she is involved in the system architecture and interconnect for a large-scale computing platform. Her research interests include interconnection networks, high-speed signaling, optical access, and metro design. Prior to joining Google, Hong was a Member of Technical Staff at Juniper Networks. Hong received her PhD in Electrical Engineering from Stanford University.

Ryohei Urata

Google Inc.
Mountain View, CA, USA
ryohei@google.com

Ryohei Urata is currently a member of technical staff in the Platforms Optics Group, where he is responsible for datacenter optical technologies and corresponding roadmaps. Prior to joining Google, he was a research specialist at NTT Photonics Laboratories, Japan. He has over 125 patents, publications, and presentations in the areas of optical interconnects, switching, and networking. He received his MS and PhD degrees in Electrical Engineering from Stanford University.

Xiang Zhou

Google Inc.
Mountain View, CA, USA
xzhou@google.com

Xiang Zhou is currently with Google Datacenter optics group, leading 400Gb/s and beyond optical interconnect technologies and roadmap development. He received his PhD from BUPT (1999). He has extensive publications, including several record-setting results. He is the author of several book chapters and the holder of over 40 US patents. He has served on the Editorial Board of Optics Express and on the Program Committee of a variety of technical conferences.

Amin Vahdat

Google Inc.
Mountain View, CA, USA
vahdat@google.com

Amin Vahdat is a Google Fellow and Technical Lead for networking at Google. He has contributed to Google's data center, wide area, edge/CDN, and cloud networking infrastructure. Vahdat received his PhD from UC Berkeley in Computer Science, is an ACM Fellow and a past recipient of the NSF CAREER award, the Alfred P. Sloan Fellowship, and the Duke University David and Janet Vaughn Teaching Award.

22. Evolving Requirements and Trends of HPC

Sébastien Rumley, Keren Bergman, M. Ashkan Seyedi, Marco Fiorentino

High-performance computing (HPC) denotes the design, build or use of computing systems substantially larger than typical desktop or laptop computers, in order to solve problems that are unsolvable on these traditional machines. Today's largest high-performance computers, a.k.a. supercomputers, are all organized around several thousands of *compute nodes*, which are collectively leveraged to tackle heavy computational problems. This orchestrated operation is only possible if compute nodes are able to communicate among themselves with low latency and high bandwidth.

In 2004 the ASCI Purple supercomputer was the first to implement optical technologies in the interconnects that support these internode communications. However, research on optical interconnects for HPC applications dates back to the early 1990s. Historically, HPC has been a large driver for the development of short-distance optical links, such as the ones found in today's datacenters (as described elsewhere in this volume). As the number of research areas and industries that exploit HPC is growing, the need for improved HPC interconnection networks is expected to persist.

In this chapter we review the requirements of current HPC systems for optical communication networks and we forecast future requirements on the basis of discernible HPC trends.

22.1 Challenges of HPC

High-performance computing (HPC) denotes the design and use of computers specifically engineered to tackle large computational challenges. A computational challenge comprises the correct execution of the instructions declared in a program and, possibly, the appropriate transformation of input data into output results. According to this definition numerous devices such as personal computers, laptops, and even smartphones or watches tackle a myriad of computational challenges every day. Traditional or embedded computers, however, reach their limits when the challenge involves a very large number of operations, a massive amount of input, intermediate, or output data, or both. In those cases, traditional computers might at best deliver the correct output, but after an execution time so long that the results have very little or no value [22.1] (e.g., weather forecast delivered after the forecast epoch). In many cases, however, traditional computers will simply fail to execute the program altogether due to insufficient temporary storage space, to the presence of a random error in one of the

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_22

components, or to external causes such as a power failure. Some applications of computing thus require systems able to:

- Execute a large number of instructions per seconds to limit the total execution time
- Hold large enough intermediate states during the computation
- Be able to continue and/or resume correct operation for weeks or even months even in the face of a power or component failure.

Computers meeting these requirements are generally called high-performance computing systems.

Of course, requirements for a large number of operations per second, large memory and storage, and prolonged error-free operation are continuously growing as high-performance computing hardware evolves and progresses. Today's laptops and smartphones have capabilities that the most powerful supercomputers had about 20 years ago. High-performance computing can therefore be defined as the avant-garde of the computing industry. In particular, the field has seen the first use of optical fibers directly within a computing system, and might prefigure a generalized use of photonics in general purpose computing (e.g., in interfaces for ultrahigh definition displays [22.2]), making this focus on HPC relevant in the context of this book.

In this chapter, the main characteristics and rationales of HPC, and of the associated high-performance interconnects, will first be reviewed. If HPC is the entry door to computing systems at large for the photonics community, it is worthwhile analyzing past, current and future requirements and trends of HPC in the context of this book. If these HPC requirements are not properly understood, the opportunity to enter the computing system market might be lost. The next section (Sect. 22.2) thus provides a digest of what a high-performance system is, and how it differs from other systems, e.g., from datacenters (Sect. 22.2.1). In Sect. 22.3, a description of HPC interconnection systems is provided. This description comes with more details as the interconnection network is the part with the highest need for optical technologies. Finally, in Sect. 22.4, we review the photonic technologies that might be deployed in next-generation HPC systems.

22.2 Defining High-Performance Computing

In the definition provided in the introduction of this chapter, HPC was compared to traditional computers, but HPC systems are also distinct from datacenters.

22.2.1 HPC and Datacenters

HPCs and datacenter *machines* have many things in common: they comprise a large number of compute cores, memories, and storage, consume large amounts of power, require heat dissipation and cooling, are operated 24/7, and often span large areas. Both execute a massive number of instructions per second. Datacenters and HPC systems can be mainly distinguished from each other by looking at the size of the tasks they execute. In an HPC machine, a *single task* can easily use most or even all resources and do so for days or even weeks. In datacenters, in contrast, most tasks are completed within seconds, and mobilize only a small fraction of the datacenter resources. A typical HPC task is a weather forecast simulation for a whole country, whereas a typical datacenter task consists of servicing a user web request. Consequently, datacenters currently use components and networks that are performing worse than their HPC counterparts. Requirements in terms of error and failure recovery, in particular, are less stringent in datacenters.

This situation, however, will change as the complexity of datacenter tasks grows. Computationally intensive tasks that would have been considered as high-performance computing a few years ago are emerging now in some datacenters. In-depth processing of video streams, for indexing or illegal content detection, is a typical example. Another example is natural language processing (NLP) [22.3]. A certain level of convergence between HPCs and datacenters is expected to take place in the next few years [22.4]. For this reason, HPC can be considered as the avant-garde for improved datacenter technologies as well, in particular for technologies that directly apply to hyperscale datacenters, as cooling, automated server management, and, obviously, interconnects.

22.2.2 Quantitatively Measuring HPC

High-performance computers come in a large variety of architectures, designs and sizes [22.5] suited for executing specific applications or application classes. Yet since the early days HPC practitioners have attempted to build a universal metric to compare HPC systems. Floating point operations per second (FLOPS) has become the gold standard for measuring performance of HPC systems. A floating point operation (abbreviated

flop) can be any arithmetic or logic one, e.g., $+$, $-$, $*$, /, AND, OR, SQRT, and so on, applied to operands of a specific precision. When operands have 64 bits, which is the default case in HPC, the FLOPS are said to be *full-precision*, while *single-* and *half-* precision FLOPS correspond to 32 and 16 bit operands respectively. Note that flop is a unit, and should be distinguished from the FLOPS unit, which represents flop/s.

FLOPS can represent either the number of operations a computing element is theoretically capable of delivering per second, in which case the term *installed* FLOPS is used, or the number of operations a computing element has effectively delivered while performing a computation (*delivered* FLOPS).

Counting *Installed* FLOPS

Determining the FLOPS *installed* in a system begins by identifying the FLOPS capability of each processing unit (PU) present in the system. A processing unit can be a CPU (central processing unit), a GPU (graphics processing unit), or any other digital logic component destined to perform computational operations.

To identify the FLOPS capability, one first has to consider the largest set of computational units that can be used at the same time, and determines how many *floating point operations* these units can perform in parallel. This number indicates the *width* of the PU and its unit is abbreviated *flop*. For example, a PU with four computational units each able to perform a fused multiply-add $d = ax + b$ can perform $width = 8\,flop$ simultaneously.

As a second step, the *width* must be multiplied by the *rate* to obtain the $flop/cycle$ throughput of the PU. The *rate* indicates the number of cycles separating admissions of new batch of operations in the PU, and this regardless of the number of cycles required for this batch to be completed. In the ideal case, the *rate* is equal to 1, which means that a new batch consisting of *width* operations can be started at every cycle, and therefore that the PU micro-architecture is capable of carrying $width \times rate = width \times 1 = width\,flop/cycle$. Note that a *rate* close or equal to 1 is rather common in modern micro-architectures.

The $width \times rate$ figure (in $flop/cycle$) must finally be multiplied by the frequency of the PU clock, expressed in *cycles per second* (thus, in Hz), to obtain the *installed* FLOPS capability of the PU (in $flop/s$). For example, if the PU with $width = 8\,flop$ that we considered above has a $rate = 0.5$ and is clocked at 3 Ghz, its installed capability would be $8\,\text{flop} \times 0.5 \cdot 3 \times 10^9\,\text{s}^{-1} = 12 \times 10^9\,\text{flop/s} = 12$ gigaFLOPS (abbreviated GF).

The FLOPS installed in all the PUs in the system are finally summed up to yield the full-system installed capability. This generally works by multiplying the number of PUs present within a chip (the term socket is also used) by the number of chips installed in every *compute node*, and finally by the number of compute nodes present in the system. The term compute node, often simply abbreviated as *node*, is widely used in the HPC community and can be assimilated to a server, in the sense that a node generally runs its own instance of the operating system and can work in relative autonomy. In contrast to conventional computing, however, compute nodes might not be individually boxed. The BlueGene architecture from IBM, for instance, consists of motherboards accepting 16 daughter boards, with two *nodes* on each daughter board [22.6]. Thirty-two nodes thus share a *blade*, which also provides a common power supply and cooling system.

At the end of 2017, Sunway TaihuLight was the publicly disclosed computer with the most installed FLOPS. Each CPU can execute 8 flop/cycle. With a rate of 1 and clock frequency of 1.45 GHz, each CPU delivers 11.6 GF. With slightly more than a million of such CPUs, the full system has an *installed* computing power in excess of 120 PF (petaFLOPS) [22.7].

Counting *Delivered* FLOPS

Installed FLOPS must be considered as a higher bound of performance, as they only reflect the maximum number of operations per second the system can perform. A system, however, rarely operates at full capacity, essentially for two types of reasons. First, the mix of operations to be performed at a given point in time may not fully match the capabilities of the computational units. In the above fused multiply example, the maximum width is reached when four $d = ax + b$ operations can be realized in parallel. Should the executed program only require multiplications, half of the PU capability would remain unused. Second, the input data required to perform the programmed operations might not be available yet, either because it must be fetched from memory, is en route from another PU, or is currently being calculated within a computational unit. All these circumstances will cause the PU to stall, most often for a couple of cycles but sometimes for much longer, while waiting for data to be available in the registers and causing a net loss of FLOPS compared to the installed capability. These leaks of FLOPS are not only unavoidable; they also reflect the quality of system design. Hence, it is not hard to realize that even the most powerful chip multiprocessor (CMP) will underperform without an adequate supply of data from memory and other CMPs. To judge the practical capability of an HPC system, the notion of *delivered* FLOPS is introduced.

The delivered FLOPS metric is obtained by dividing the number of flop required to complete a specific

calculation (generally a benchmark) by the time taken by the system to perform the calculations. The former number can be obtained either by annotating the code with operation counters and executing it, or by means of a complexity analysis of the underlying algorithm. One of the most famous of the benchmarks used to evaluate the delivered FLOPS, LINPACK (linear equations software package), requires the solution of a system of linear equations, which is known to require $2/3n3+2n2+O(n)$ flop. The time taken to perform the calculations is generally experimentally measured, as predicting the number of PU cycles required, although theoretically possible, is generally intractable.

We note that the installed FLOPS metric is rather universal, possibly only subject to the interpretation of the width and rate parameters [22.8]. The delivered FLOPS metric of a given system, in contrast, depends on many aspects: the type of calculation performed by the benchmark application, the way it has been programmed, the compiler used and possibly even the input dataset. Different codes will challenge the components of an HPC system differently. The LINPACK benchmark, used in the *Top 500* ranking [22.9], is known to be extremely computationally intensive, stressing computational units primarily. Another benchmark called HPCG (high-performance conjugate gradient) [22.10], in contrast, stresses the communication infrastructure predominantly. Across multiple systems, the delivered FLOPS thus largely depends on the architecture and design choices. Two systems with equal installed FLOPS but different architectures generally deliver different delivered FLOPS when executing the same benchmark.

System Efficiency

The so-called system efficiency metric can be obtained by dividing the delivered FLOPS by the installed one. For the users of an HPC system, the efficiency denotes the intensity at which the tested benchmark leverages the computational resources. System designers, in contrast, will see the system efficiency as the result of efforts aimed at adequately connecting the computational units with the memory and among themselves to limit the number and duration of PU stalls. LINPACK calculations are relatively straightforward to conduct, even on highly parallel architectures (Sect. 22.2.4), and therefore system efficiencies for LINPACK are generally above 50% and sometimes above 90% [22.9, 11]. For HPCG, in contrast, efficiency rarely exceeds 10% and sometimes even falls below 1% [22.10].

Other Performance Evaluation Strategies

FLOPS-based performance evaluation methodologies are frequently criticized because they assign an oversized importance to computational units. As evidenced by the HPCG benchmark, computational units are not necessarily the performance bottleneck and, for some application, focus should be placed on other characteristics of the system. The Graph 500 is known as a benchmark that is not based on FLOPS [22.12, 13]. It comprises the pseudorandom construction of a massive graph with up to trillions of edges followed by a breadth-first search (BFS) over the graph. The performance is reported in terms of traversed edges per second (TEPS). This benchmark mimics calculations used in data analytics. The HPC system exhibiting the best performance on this benchmark, as of November 2017, is the six-year-old K computer, with 38 teraTEPS. This is more than the 24 teraTEPS demonstrated by Sunway TaihuLight (ranked second), even though the latter has more than 11 times more installed FLOPS.

Quantitatively Describing HPC Systems: Conclusions

Table 22.1 and Fig. 22.1 summarize the characteristics of some of the most powerful HPC systems active at the time of writing. FLOPS, installed or delivered, is a simple metric permitting description of the capabilities of an HPC system. As shown through the Graph 500 example, however, FLOPS counts are far from being a universal metric. Nevertheless, as we will see in the next section, FLOPS counts remain a powerful tool to scale various requirements, such as the memory bandwidth or communication bandwidth. It is therefore important to be aware of the FLOPS characteristics of the main systems.

In general, collocating many computational units within a single PU, and many PUs within a chip, requires engineering and technological exploits. Such exploits will boost a system's installed FLOPS, but offer no guarantee of increasing the delivered FLOPS and more generally improving the overall performance. In the design phase of a supercomputer architecture, HPC system designers face the choice of investing resources in installed FLOPS exclusively, which can cause a drop in the efficiency, or in other resources such as memory or the interconnection network that are likely to increase the overall efficiency. A system can be considered balanced for a given application when an additional dollar spent on installed FLOPS raises the delivered FLOPS by the same amount as a dollar invested on other components. For example, if adding more PUs in a system results in more delivered FLOPS, but boosting memory or interconnects of that system keeps performance unchanged, the computing part of that system is likely to be undersized compared to the noncomputing part.

Table 22.1 Main benchmarks for supercomputers

System name	Year of introduction	Top 500 rank	HPCG rank	Graph 500 rank	Delivered petaFLOPS (LINPACK/ Top 500)	Delivered petaFLOPS (HPCG)	TEPS (Graph 500)
Sunway TaihuLight	2016	1	5	2	93 015	0.481	23 756
Tianhe-2 (MilkyWay-2)	2013	2	2	7[a]	33 863	0.580	2061
Piz Daint	2016	3	4		19 590	0.486	
Gyoukou	2017	4			19 135		
Titan	2012	5	9	3	17 590	0.322	
Sequoia	2012	6	8		17 173	0.330	23 751
Trinity	2017	7	3		14 137	0.546	
Cori	2016	8	7	1	14 015	0.355	
Oakforest-PACS	2016	9	6	4	13 555	0.385	
K computer	2011	10	1		10 510	0.603	38 621
Mira	2012	11	12		8587	0.167	14 982
TSUBAME3.0	2017	13	10		8125	0.189	
JUQUEEN	2013	22	19	5	5009	0.096	5848

[a] Performance of a fractions of MIRA (ranked sixth, seventh and ninth in the graph500) excluded; if taken into account, Tianhe-2 ranks tenth

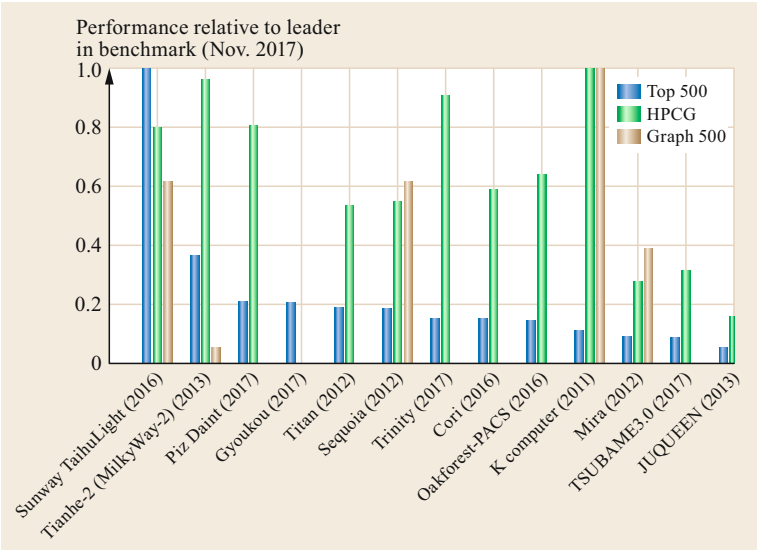


Fig. 22.1 Performance relative to leader of various supercomputers executing the Top 500 (LINPACK), HPCG (high-performance conjugate gradients) and graph500 (graph construction and search) benchmarks. Results from November 2017

22.2.3 Economics of HPC

Hyperion Research, a consultancy firm, estimated the HPC market to be worth $11.2B in 2016, and segmented the market into four categories: i) the largest systems, priced over $500k (and up to hundreds of $M), denoted as supercomputers, ii) divisional systems priced between $250k and $500k, iii) departmental systems ($100k–250k) and iv) workgroup systems (under $100k) [22.14]. These categories represent respectively 37, 20, 28 and 15% of the HPC market. For comparison, IDC, another consultancy firm, estimates the whole server market to be worth around $60B. HPC equipment thus represents an important but not dominant share of the total server computer market (which does not include desktop or laptop computers, but does include datacenters).

As of November 2017, supercomputers, the above listed category of most interest for optical networking, were defined as systems with 200 TF (teraFLOPS) or more of installed computing power, and 6700 TF were sufficient to be listed in the 500 most powerful supercomputers. The number of supercomputers installed in 2016 is of the order of one thousand. Due to the scale

of the systems sold, the HPC market, and more particularly the supercomputer market, represent significant revenues and margins, but not many units sold.

The cost per installed and fully connected TF, which has been constantly falling thanks to progresses in very large scale integration (VLSI) and semiconductor fabrication processes, in 2017 were roughly on the order of $ 2000. In 2017, a budget of ≈ $ 200M thus affords a 100 PF system. This amount of ≈$ 200M has been frequently invoked by the US Department of Energy, one of the main operators of supercomputers, as the upper limit for the largest scale machines. To reach an EF (exaFLOPS) installed capacity, the next grand challenge in supercomputing [22.11], while remaining within a $ 200M budget, the cost per TF should be reduced by an order of magnitude at least [22.15], and be brought below $ 200 per TF.

Energy consumption costs and other operational costs (OpEx), if generally inferior to acquisition costs (CapEx), represent a significant portion of the total cost of ownership (TCO). By dividing the delivered FLOPS metric by the average power consumption, the energy efficiency of a system (expressed in FLOPS per Watt, or flop per Joule) can be obtained. The current rule-of-thumb is 1 GF/W (gigaFLOPS per Watt); as of November 2017, out of the 306 systems disclosing energy efficiency in the Top 500 benchmark, 204 were showing an energy efficiency metric better than this mark. Seven systems were also capable of delivering more than 10 GF/W. An energy efficiency of 6 GF/W has been measured when executing LINPACK on the *Top 500* benchmark leader Sunway TaihuLight [22.9].

Although relatively energy efficient, the Sunway TaihuLight system consumed an average of 15 MW during the execution. Such high power consumption, comparable to that of a town, requires special access to the electric grid, as well as special-purpose power transform stations, which increases direct and indirect costs. To avoid excessive power-supply overhead costs, it is commonly agreed that the power consumption of a supercomputer should be kept below 100 MW. The US Department of Energy, both one of the largest supercomputer operators, and the operator of some of the largest systems, discourages power consumptions beyond 20 MW. Reaching an exaFLOPS capacity with a 20 MW power budget requires energy efficiency to increase to 50 GF/W at least.

Supercomputers usually have lifetimes of five to seven years. For instance, the K computer became operational in 2011 and is still active six years later. As performance growth tends to decelerate (as it will be addressed in Sect. 22.2.5), longer lifetimes might become the rule. Thus, it is worth remarking here that operating a 20 MW-system for ten years yields a lifetime consumption of 1752 MWh. At ≈ $ 75 per MWh, a fair estimate of the price paid by wholesale electricity consumers [22.16], the electricity cost amounts to $ 131M, a figure not very far from to the $ 200M acquisition cost mentioned above. This comparison shows that reducing power consumption can be relevant just from an operating cost perspective. Should the CapEx cost per TF be improved by a factor of ten in the next few years while the OpEx power per gigaFLOPS stays unchanged, an exaFLOPS-capable computer would consume 100 MW and thus cost $ 655M over ten years. This 3.2× more than its acquisition cost, not taking into account extra power-supply station costs. Progress on the energy efficiency front must thus be realized alongside improvements in the cost efficiency [22.15].

22.2.4 Parallel Nature of High-Performance Computing

HPC has traditionally relied on parallelism to increase the advantage over mainstream computers while relying on the same semiconductor process for chip manufacturing. In fact, historically, most of the advances in processor architecture have been introduced in HPC systems first. The CDC 6600 system, introduced in 1964 and considered to be the first supercomputer, with its central processor equipped with ten functional units, was capable of working on different instructions at the same time [22.17]. This superscalar design, corresponding to a width factor larger than 1, is now prevalent in most modern processors. The CDC 6600 successor, the CDC 7600, introduced pipelining [22.18], i.e., the capability for a functional unit to work on several instructions at the same time, and thus to approach the aforementioned ideal *rate* of 1 (even though completing each instruction would take multiple cycles). This CDC 7600 was to be outperformed by the Cray-1 system in the early 1980s [22.19]. The Cray-1 supercomputer relied on vector processing, a feature present today in not only graphical processing units (GPUs), but also in conventional processors (Intel MMX, PowerPC Altivec) [22.20]. Vector processing allows one instruction to be applied to multiple data at the same time, leading to even higher *width* values.

Besides introducing more parallelism within the PU, HPC has also relied on *spatial* chip- or node-parallelism to develop more powerful systems. Spatial parallelism means collocating multiple computing elements, usually all identical, within the same system, along with an interconnection network. The Cray X-MP supercomputer, which succeeded to the Cray-1 as the most powerful supercomputer in 1985, initiated the trend with two and then four interconnected vector processors. Other companies, however, quickly realized

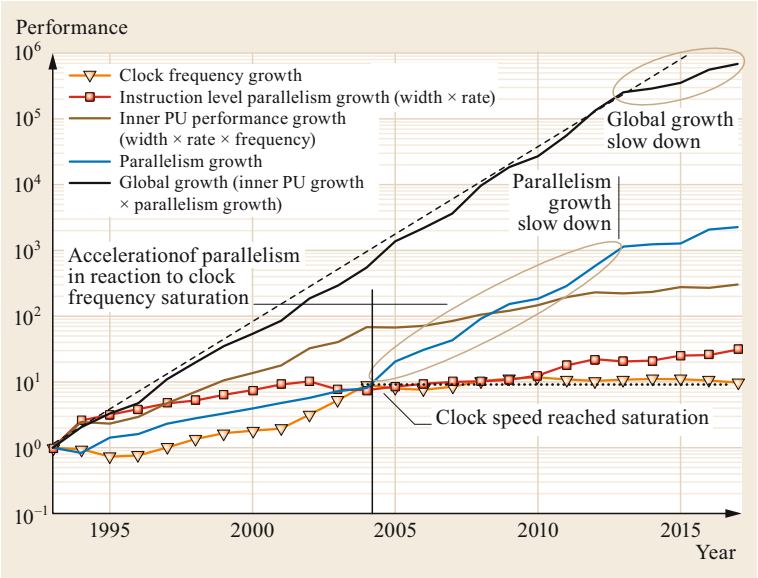


Fig. 22.2 Evolution of various contributors to overall top 20 HPC systems performance growth, relative to 1992. The *brown line*, representing the processing unit throughput, results from the product of the *orange triangle line* (processor frequency) and of the *red square line* (instruction-level parallelism). The *black line* results from the product of the *brown* and *blue lines* (processor parallelism). (Source: top500.org)

that combining many cheap components around an efficient interconnect could be more cost-effective than tailoring custom chips whose complexity was close to unmanageable. In the 1990s systems with more than one hundred processors came on the market. Fujistu's Numerical Wind Tunnel, the fastest supercomputer in 1993, incorporated 140 processors [22.21], and Thinking Machines CM-5/1024, the second-fastest machine in 1993 according to the *Top 500* ranking, had more than one thousand processors.

This was only the beginning of establishing chip parallelism as one of the main drivers of HPC performance. As shown in Fig. 22.2, the number of cores present in the 20 most powerful supercomputers steadily increased during the 1990s.

22.2.5 End of Dennard Scaling and Beyond

As apparent in Fig. 22.2, from the early 1990s to the mid-2000s, the progression in chip- or node-level parallelism was still accompanied by a steady increase in each single PU capability, through increased clock speeds and PU inner parallelism (called instruction-level parallelism ILP). Progress in the miniaturization of transistors have driven this PU performance progression. *Dennard* et al. [22.22] stated in 1974 that by decreasing the size of a transistor by a factor k, the transistor operates k times faster, while its power consumption (at the faster rate) is divided by $\approx k^2$. This implies that power density is constant, i.e., a chip of surface S consumes roughly the same power, no matter if it carries thousands, millions or billions of transistors. Moreover, the more transistors, the faster the chip operating frequency. According to what has become known as Dennard's law, each new generation of semiconductor fabrication technology thus provided an additional budget of transistors that can be used to boost the PU performance, for instance through enlarged widths or rates closer to 1, along with an intrinsic speed-up through the clock frequency. It is worth noting that Dennard's scaling provides the physical grounds for the better-known Moore's Law. Moore's Law, in its initial version of 1965, postulated that the number of transistors in densely integrated circuits doubles about every 12 months. The projected rate of growth was later revised to every 24 months in 1975.

Dennard's scaling neglects a couple of phenomena, in particular leakage current and threshold voltage, deemed marginal in 1974 when Dennard's law was minted. After three decades of fast-paced scaling, however, these phenomena started to play a role after the turn of the millennium, and currently almost dominate the transistor power consumption. These effects partly caused the inner PU performance to plateau in the mid-2000s (as highlighted in Fig. 22.2). Hence, from thereon, PU clock frequency has been curbed to compensate for the extra power consumption due to leakage current and threshold voltage. The other reason for PU performance to saturate is that it has become more advantageous to collocate two or more identical CPUs on the same chip, something called core-level parallelism, than to further augment the complexity of a single PU through increased width and improved rates.

Core-level parallelism is nowadays the main performance driver, but it might not remain such for long. Moore's law is nearing its end [22.23]: leading semi-

conductor manufacturers currently operate (in 2018) at a resolution of ≈ 10 nm, leading to ≈ 60 nm-wide transistors. Industrial fabrication platforms allowing sub-10 nm resolution are in preparation, and IBM recently presented a 5 nm prototype process [22.24]. But resolutions below 2 nm are generally considered nearly impossible to achieve: that would mean structures with only 10 atoms (an atom of silicon, measured through its covalent radius, is 220 fm large). This means that the number of transistors available on a single chip will eventually be limited to around 100 billion, a factor ten from today. This portends ultimate chip-level FLOPS performance of 50–100 TFs, as opposed to ≈ 5 TFs today.

Beyond this ultimate integration point, scheduled to happen around 2024 [22.25], further performance enhancements will almost exclusively be obtained by means of chip parallelism. In order to maintain the cost efficiency, chip designs will have to be simplified. Supercomputer housing and packaging (racks, power supply, etc.) will similarly have to be rationalized. In addition, and central to this chapter, interconnection networks will have to be tremendously improved in terms of cost and energy consumption.

22.3 Contemporary High-Performance Interconnection Networks

As shown in the previous section, installed computing power in excess of 5 TFs requires more than one chip, and requires these chips to be interconnected. For systems up to 50 TFs, it is possible to collocate up to a dozen of chips inside a single compute node (e.g., the recently introduced Nvidia DGX-1 station [22.26]), and to interconnect them using ad hoc links hardwired onto the motherboard. To reach higher installed computing power, however, there is little other option but to use multiple nodes side by side, and to introduce a high-performance interconnect to *glue* these nodes together in order to obtain a standalone system. An interconnection network is therefore a necessary and key component of any contemporary supercomputer. The goal of this section is to provide a comprehensive overview of the topology and characteristics of HPC interconnects. A correct understanding of these requirements is deemed necessary to correctly seize the role played by optical networking in HPC systems.

22.3.1 Interconnect Building Blocks

Figure 22.3 provides a high-level picture of an HPC interconnect, limiting the components to three types: adapters, routers and links.

Adapters

Adapters, also sometimes called NICs (network interface controllers), are the entry points to the network. The role of the adapter is obviously to inject packets into the interconnect access links with the appropriate format, as well as to collect incoming packets. High-performance NICs, however, also integrate advanced functionalities such as communication offloading [22.27], allowing a compute node to receive data from the network and store it into its memory while performing other computations (unrelated to the transmitted data). This capability to compute and communicate in parallel keeps the system efficiency high, as it prevents installed FLOPS from being *wasted*. Adapters are often connected to the computing chip(s) using PCIe (peripheral component interconnect express) links. In some case, the adapter is integrated on the same die along with computational units (e.g., in the BlueGene architecture from IBM [22.6, 28]).

Discrete adapters pluggable into PCIe ports are being proposed by several vendors, for a price currently spanning from ≈ $ 100 to ≈ $ 1000, depending on the data rate and format. In 2017, for 100 Gb/s speeds, mostly relevant for modern interconnects, adapter price starts at $ 450.

Routers

The *packet routers*, also denoted as *switches*, receive packets onto their N input ports and ensure these packets are forwarded with minimal delay to the correct N output ports. To minimize delays, a *cut-through* approach is adopted, i.e., the first bits of a packet are emitted onto the output port even though the last bits of the packet have not reached the router yet. High-performance routers are generally capable of extremely fast reaction times and, in the absence of traffic congestion, delay packets by tens of nanoseconds only [22.29]. Routers are also generally capable of delivering close to 100% throughput, i.e., they receive and send data on every link at maximal rate, as long as packets are adequately distributed onto output links. In addition to basic forwarding, routers are able to detect errors and apply quality-of-service (QoS) policies. QoS policing allows, for instance, traffic exchanged by computing resources, generally more sensitively to latency, to be processed with higher priority than traffic carrying output results to be stored onto the file system.

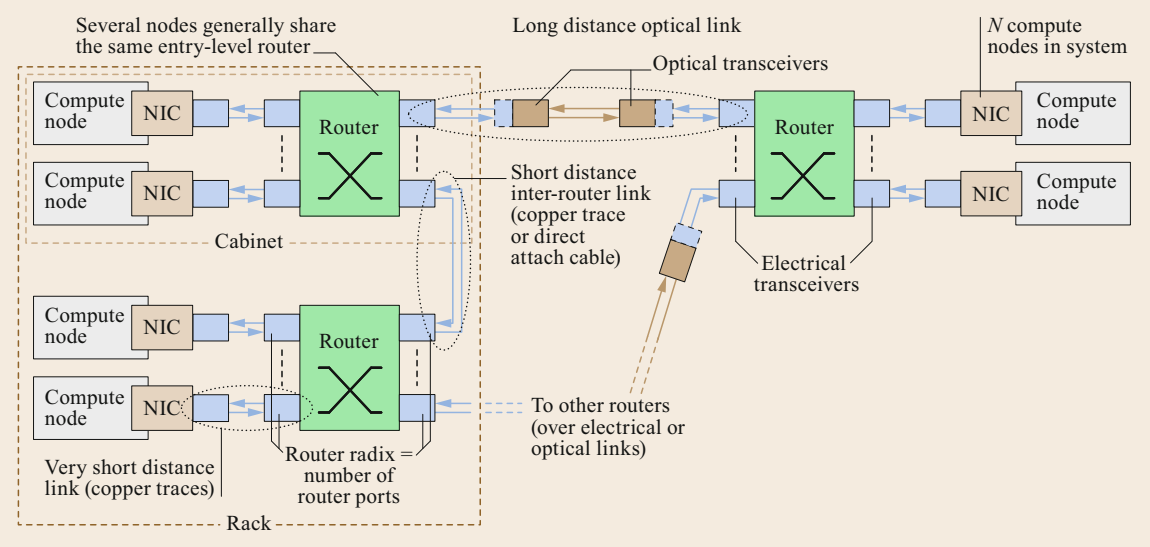


Fig. 22.3 High-level representation of the interconnect of a 2018 large-scale HPC system. Note that optical transceivers are used exclusively to connect routers not located within the same rack

From the HPC system architect perspective, routers are mainly characterized by the link-level and network-level protocols they implement, the data rate of the port, and the number of ports they provide (called the *radix*). The product of these last two figures gives the router *total switching bandwidth*, which represents the data amount that traverses the router and undergoes a switching operation every second. For example, a 36-port router operating at 100 Gb/s offers a 3.6 Tb/s total switching bandwidth. Note that router vendors sometimes present their product in terms of total I/O bandwidth, which is the sum of both ingoing and outgoing bandwidths, and thus equates to double the total switching bandwidth. In the above example, the router has a *total* I/O *bandwidth* of 7.2 Tb/s.

Modern HPC packet routers integrated on a single chip [22.29] provide up to 80 ports and link data rates up to 100 Gb/s, leading to total bandwidths reaching 5–10 Tb/s today (2017). They generally dissipate 100–200 W, leading to an energy dissipated per bit at full load of 10 to 40 pJ/bit. The router architecture reported by *Scott* et al. [22.29], which has been extensively leveraged in the recent systems from Cray [22.30, 31], a supercomputer vendor, shows a zero-load latency of 31.25 ns. This zero-load latency is the minimal amount of delay experienced by the first data bit of a packet transiting through the router.

Several vendors offer standalone boxed HPC routers including the router chip, a controller, fans, power supply, and connectors, for a price per port ranging from $100 to $500 per port and a price per Gb/s of $1.5 to $10, depending on the data rate and number of ports. Routers delivering 100 Gb/s with 48 ports currently deliver the best cost/bandwidth ratio of ≈ 1.8 $/Gb/s translating into $\approx$ $180 per port ($8640 for the full equipment). A speed of 100 Gb/s was the highest offered by vendors, although products supporting 200 Gb/s were about to emerge. Products supporting 400 Gb/s are in preparation.

Multiple router chips can be combined inside a single package to offer an increased number of ports. The so-called resulting *director switches* can offer more than five hundred ports [22.32], at the expense of a significantly larger cost, higher per switched bit power consumption, and higher latency, reflecting the fact that most bits effectively traverse two or even three chips internally.

Links

HPC interconnect links connect adapters to their *entry-point router*, and routers among themselves. Links can be *active* or *passive*. In the latter case, the signals emitted by the adapter, the router chip, or a port-dedicated retiming chip are directly sent onto the transmission medium. This medium can be a copper-based cable, most often a coaxial cable with two inner conductors (called twinaxial, and often shortened twin-ax), or simply a pair of wires on a backplane. Differential signaling [22.33] is generally employed.

Active links, in contrast, have a transceiver on each end. A transceiver prepares and transforms the data received from the adapter or router for a transmission over a (generally) longer distance than passive links. It similarly receives the data after its transmission along

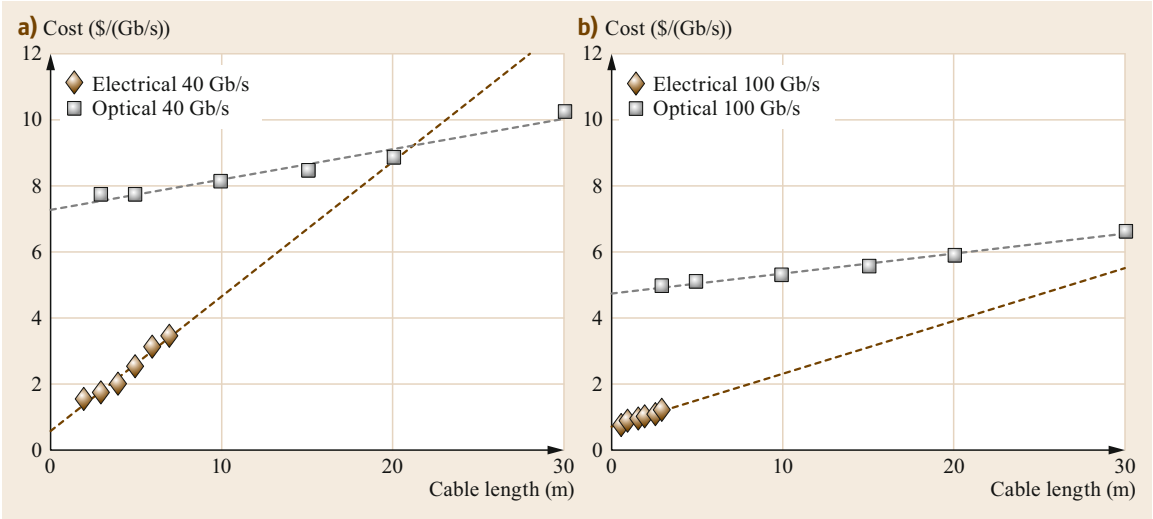


Fig. 22.4a,b Evolution of HPC links cost from 2014 **(a)** to 2017 **(b)**. Clearly apparent is the reach limitation of electrical links, leading to a cost discontinuity (in 2017, from $\approx 1\,\$/(\mathrm{Gb/s})$ for 4 m to $\approx 5\$/(\mathrm{Gb/s})$ for 5 m distance)

the link and adapts it back to the requirements of the router or adapter port. Active links can be electrical, in which case the transceiver simply retimes and/or amplifies the signal prior to launching it onto the transmission medium. Most active links, however, are based on optical communication technologies. An optical active link is either composed of a pair of transceivers and a fiber cable of arbitrary length (within the range supported by the transceiver), or prepackaged with a fixed-length fiber. The latter option is called an *active optical cable* (AOC).

Links, active or passive, are mainly distinguished by their length, price and power consumption. Figure 22.4, which partly reuses data from *Besta* and *Hoefler* [22.34] illustrates the relationship between link lengths and costs, considering the cheapest commercial on-the-shelf links available at two points in time (2014 and 2017). Electrical links are roughly five times less expensive than their optical counterparts, in $ per link terms, but are clearly limited in length. In 2014, the most economical data rate was 40 Gb/s, a speed that passive cables could carry over distances up to 7 m. In 2017, 100 Gb/s turned out to be the most economical data rate, but lengths of electrical cables were limited to 5 m. It is worth noting, however, that prices given in Fig. 22.4 correspond to discrete components, to be bought separately and individually. Cheaper electrical links can be obtained by integrating many copper traces onto a single motherboard, for instance, as well as by ordering large quantities.

The transmission technology over metallic wires causes electrical links to abruptly turn uneconomical beyond a certain point. This overall results in a price discontinuity, clearly visible in Fig. 22.4, between short connections (currently below 5 m) and longer ones. As of today, and as shown in Fig. 22.3, metallic wires are used to connect chips among themselves on the same motherboard (chip-to-chip links), motherboards among themselves within a *chassis* or *cabinet* (intracabinet links), and to connect cabinets located within the same *rack* (intrarack links). Several neighboring racks may even be interconnected with metallic wires. Optical links, in return, are used for interrack links exclusively. As will be seen in Sect. 22.3.4 describing the Dragonfly topology, the major price difference between electrical and optical links have so far encouraged interconnect designers to develop interconnection schemes that make the most use of electrical cable and limit the number of optical connections required.

It is important to note that electrical cables can be expected to become uneconomical for shorter distances as data rates evolve toward 400 Gb/s and beyond (as summarized in Sect. 22.3.6). Optical links, in contrast, are expected to become more economical thanks to higher integration, as will be evoked in Sect. 22.4. Overall, one can expect optics to eventually equip all intercabinet links (i.e., roughly longer than one meter), and possibly some intracabinet ones (> 30 cm).

In terms of power consumption, the most energy efficient active optical links consume 15–25 mW/(Gb/s) (or equivalently pJ/bit, as Watt is equivalent to J/s). A 100 Gb/s link thus consumes about 4 W, i.e., 2 W per extremity. Such an energy efficiency figure, of the order of 10 pJ/bit, might seem a lot provided that numbers in the 1 pJ/bit range or even below are frequently evoked in the literature (e.g., as summarized in the survey from

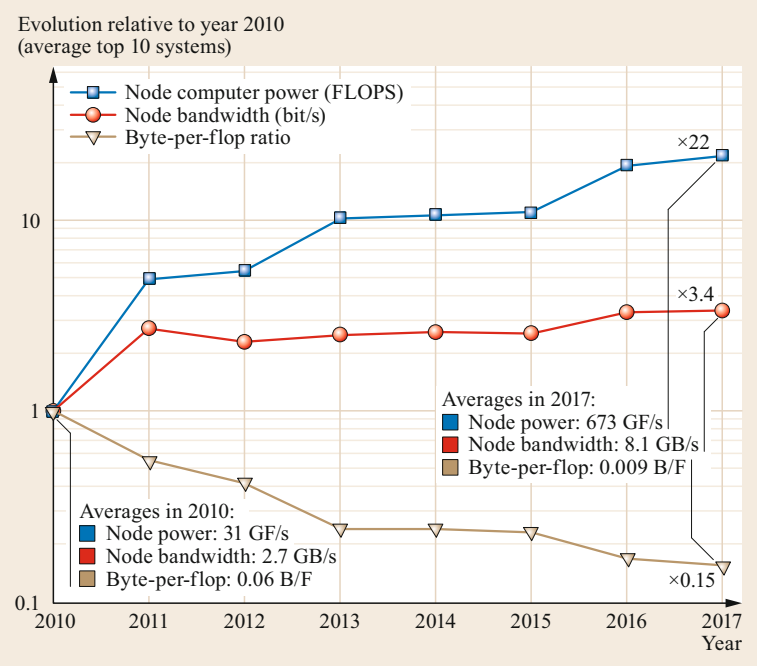


Fig. 22.5 Relative evolution of the node computer power, of the node injection bandwidth, and of their ratio over the years, for the top 10 systems. If the average node compute power has increased by a factor of 22, the injection bandwidth has risen by a factor of 3.4 only, leading to a 6.5× decrease of the byte-per-flop ratio

Thraskias et al. [22.35]). However, it is important to note that this power consumption corresponds to the amount of power that is driven from the host, and thus includes everything from the laser, through the modulator, the driver, to the serializer, and to the electrical interface. The power consumption of the latter is seldom included in the literature, in which articles often focus on the transmission subsystems, and not on the full system.

The energy dissipation of the active optical cable adds to the energy dissipated by the router and adapter to send, receive, and transmit data. The latter is generally limited to 2–3 pJ/bit [22.36].

22.3.2 Injection Bandwidth

The high-level representation of an HPC interconnect shown in Fig. 22.3 does not illustrate two very important aspects of the design of an HPC interconnect: the bandwidth of the system links, and the way routers are interconnected, i.e., its topology. We start by describing the former.

The amount of bandwidth a compute node has available to *reach out* to the other nodes through the interconnect is called the *injection bandwidth*. The injection bandwidth is generally set in proportion of the FLOPS installed on the node. A scaling factor in *byte/s/FLOPS* is thus introduced to obtain the byte/s required by a node with a given amount of installed FLOPS. Note that the byte/s/FLOPS unit is generally simplified in *byte/flop*, where flop stands again for floating point operation, a terminology that is also more intuitive. Note also that for convenience, 1 byte/flop is often approximated with 10 *bit/flop*. Fundamentally, 1 byte = 8 bits but due to transmission system overheads as parity checks or line codes, more than 8 bits are *physically* sent over a link for each byte. The 1 byte ≈ 10 bits approximation accounts for these overheads.

The injection bandwidth scaling factor of an existing system indicates how many bytes can be communicated onto the interconnect for each executed flop before reaching network congestion, which itself causes execution slowdown. Injection bandwidth scaling factors of leading supercomputers are between 0.01 and 0.001 byte/flop and show a decreasing trend over time, as shown in Fig. 22.5.

The injection bandwidth scaling factor is strongly related to the *verbosity* factor, that denotes how many bytes are *listened to* and *spoken out* each time a flop is executed. *Verbosity* factors can be defined for various components of the supercomputer node. At the very core of the PU, the *verbosity* is of 24 byte/flop at least, as an operation requires two 8-byte (64 bits) operands, produces an 8-byte result and requires an instruction. However, operands or results are often almost immediately reused for another operation, and thus do not need to be read again from or stored in the memory. Through reuse, the *verbosity* of the main memory can diminish to ≈ 0.1 bytes/flop (a figure generally considered by HPC node designers) or even less. The ability to reuse

operands, however, is highly application-dependent. The aforementioned HPCG benchmark, for instance, has a *memory verbosity* of 4 bytes/flop [22.37], which alone almost explains why system efficiencies while executing the HPCG benchmark seldom exceed 10% (Table 22.1). The *network verbosity* factor is similarly dependent on the application, but furthermore depends on the size of the problem being solved, the number of nodes solving it, and the way the application is mapped to these nodes [22.38]. As such, it is hard to predict, but can be measured experimentally [22.39]. In general, network verbosity factors tend to decrease as computational power of nodes is increased: with larger nodes, a larger proportion of the data exchanges between PUs remain concealed within the nodes, and does not emerge onto the interconnect. This effect, known as the surface-to-volume ratio, partly explains the decreasing injection bandwidth scaling factor trend shown in Fig. 22.5.

Historically, nodes have been most often equipped with a single adapter to provide the node's entire injection bandwidth. However, the recent increase in the number of installed FLOPS per node, which mainly drove the computing power progression in the last few years (Fig. 22.5), calls for a multiplication of adapters. Typically, a 40 TF node such as the Nvidia DGX-1 station needs 40 GB/s of bandwidth just to support a network verbosity factor of 0.001 byte/flop. As no single adapter on the market can deliver 40GB/s of bandwidth yet, the station integrates four independent adapters each running at $\approx$ 10 GB/s. The term *double-rail* or *quad-rail* is sometimes used to denote node architecture with multiple adapters.

Obviously, multirail node architectures put larger demand on interconnect topology scalability (as will be defined in the next subsection). A thousand-node system with quad-rail adapters requires, for instance, an interconnect supporting 4000 end-points. An end-point is a generic term that describe a *leaf* connected to a topology.

22.3.3 Topologies

Next to the node injection bandwidth and associated supported *network verbosity* factor, the *topology* is the most notable characteristic of an HPC interconnect. The topology defines how end-points are connected to their entry-point routers, and how these routers are connected among themselves. Neglecting lost packets, which seldom occur in HPC, an interconnect can be seen as an abstract *packet delayer*: any packet injected onto the interconnect is eventually delivered to its destination node with some delay. The role of the topology essentially consists of minimizing these delays, given the available hardware and injection bandwidth (dictated by a network verbosity factor).

To minimize delays, an HPC topology must first ensure that any node can reach any other node through a minimal number of hops over the network. This minimizes latency as each hop comes at the expense of some zero-load latency. To that end, the average distance of the topology should be minimized to guarantee low delays on average. However, the worst-case delay, which can be highly deleterious in parallel environments [22.40], should also be minimized. This is why topology architects pay very special attention to the topology *diameter*. The diameter is the number of hops in the longest of the minimal paths connecting one vertex to another, and thus a proxy for the worst-case zero-load latency. If the diameter of the topology is D, and realizing a hop in the topology takes about 50 ns, as reported for instance for the Blue Gene/Q interconnect [22.28], a packet will be delayed by $50D$ ns at least.

In addition to ensuring low *baseline*, zero-load latency due to propagation along links and traversal of routers, the topology must ensure that in most practical cases, as few packets as possible will experience too-long delays due to queuing. To achieve this, traffic injected should find sufficiently ample and lightly loaded paths onto the network to reach its destination.

By scaling the bandwidth available for a given topology (e.g., by doubling the bandwidth available on every link), one generally improves performances, so the overall bandwidth provided by an interconnect is an important quality factor. At the same time, two topologies comprising the same number of links and the same aggregated bandwidth (summed over all links) might yield distinct performances [22.41]. A good topology is a combination of an adequate quantity of routers and bandwidth with a deft recipe that describes how routers and end-points are connected. As for injection bandwidth, topology requirements depend on the application(s) execution in the supercomputer.

HPC topologies can be direct or indirect. In direct topologies, every router has end-points attached to it, whereas indirect topologies involve *inner* routers not directly exposed to the end-points. In the remainder of this subsection, the properties of some of the most common and exemplary topologies are detailed. Note that mathematical developments leading to scaling equations are omitted for the sake of conciseness.

Indirect Topologies

Fat-Trees. Indirect topologies are generally called fat-trees: end-points are the leaves of the tree, and traffic progresses up the root until it reaches a common ancestor with the destination, at which point it is routed downwards back to the destination end-point. The num-

ber of routers that must be traversed to reach the root of the tree (including the root router) indicates the number of levels present in the tree *minus one*. For instance, in the fat-tree constructs visible in Table 22.2 (to be defined next), three routers must be traversed to reach the top level, thus all these constructs are two-level fat-trees. The diameter of the fat-tree, excluding hops from and to end-points, is simply the double of the number of levels.

If the total bandwidth available between any consecutive levels of the tree is constant, and equal to the total bandwidth connecting the end-points to the entry-point routers, a fat-tree is said to be fully provisioned. As long as the traffic destined to one particular end-point does not exceed the end-point's ejection bandwidth, fully provisioned fat-trees are immune to congestion, even if all the traffic must progress up to the root of the tree. Fully provisioned and populated fat-trees with k levels and constructed with routers with *radix* r (thus offering r ports) have in total $r(r/2)^{k-1}$ ports available for end-points. If each compute node occupies a single end-point, the corresponding HPC system can scale up to $r(r/2)^{k-1}$ nodes. Fully provisioned fat-trees involve $(2k-1)(r/2)^{k-1}$ routers with $(k-1)r(r/2)^{k-1}$ internal links. An internal link is a link that connects two routers and thus that is not connected to an end-point. This corresponds to $k-1$ internal links per end-point, or k links in total (counting the link connecting the end-point to the interconnect). This also corresponds to $(2k-1)/r$ routers per end-point.

These expressions in *per end-point* terms are handy as they permit us to quickly evaluate the interconnect cost in proportion to the end-point's one. Consider, for example, a fully provisioned fat-tree with $k=2$ levels made of routers with $r=48$ ports, and accepting up to $r(r/2)^{k-1}=1152$ end-points. With a router and link prices of \$ 8000 and \$ 80 (passive links, Fig. 22.4), corresponding to 100 Gb/s data rate, the *cost per end-point* is in this case $(2k-1)/r\times\$\,8000+(k-1)\times\$\,80=(3/48\times\$\,8000)+(2\times80)=\$\,660$. This cost per end-point must be compared to the cost per TF discussed in Sect. 22.2.3. If the budget for a 1 TF node (interconnect included) is \$ 2000 and if this node is connected through a single adapter, the interconnect cost represents about a third of the cost. We note that connecting a 1 TF node through a single 100 Gb/s adapter yields an injection bandwidth scaling factor of 0.001 bit/flop $\approx$ 0.01 byte/flop. For a three-level fat-tree, scalable up to 27 648 end-points, the cost per end-point grows to $(\$\,8000\times5/48)+(3\times\$\,80)=\$\,1073$, in which case interconnect equipment would represent more than a half of the budget. These example calculations illustrate how interconnect equipment cost prevents the injection bandwidth scaling factor from being kept relatively high.

Slimmed Fat-Trees. To mitigate the relatively high cost of fully provisioned fat-trees, oversubscription is often applied to fat-trees, which then become *slimmed fat-trees* or *tapered fat-trees* [22.38]. Oversubscription consists of allocating more ports downward the tree than upward. For example, out of the 48 ports of a router, 36 can be used to connect end-points, while only 12 are used to connect to the higher levels. In this case the oversubscription ratio is 36 : 12 thus 3 : 1. Applying an oversubscription of $x:1$ allows us to inflate the scalability of the fat-tree by $2x/(x+1)$. The scalability of a two-level fat-tree with a 3 : 1 oversubscription ratio applied to either level thus grows by $2\times3/4=1.5$, while the scalability of a three-level fat-tree with two successive 2 : 1 oversubscription ratios is multiplied by 16/9, and thus increases by about 78%. Table 22.2 displays all possible ways to oversubscribe a two-levels fat-tree with $r=6$.

Oversubscription permits us to enlarge the fat-tree scalability for a given number of levels k and router radix r. This permits us to connect more modes while conserving a low diameter (this will be further discussed in next subsection). Oversubscription also permits us to reduce the number of routers and links per end-point. Assuming that the scalability offered by the resulting fat-tree is fully utilized, i.e., that

$$\left(\frac{r}{2}\right)\frac{2x_1}{x_1+1}\left(\frac{r}{2}\right)\frac{2x_2}{x_2+1}\cdots r\,,$$

end-points are interconnected. The number of routers required per end-point is reduced by a factor

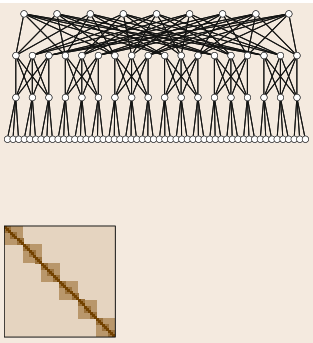
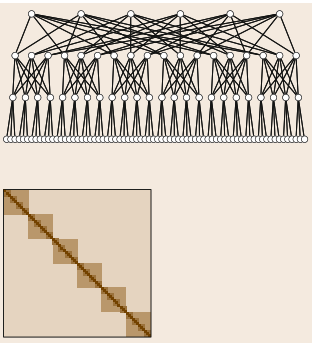
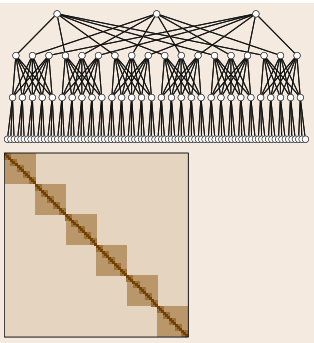
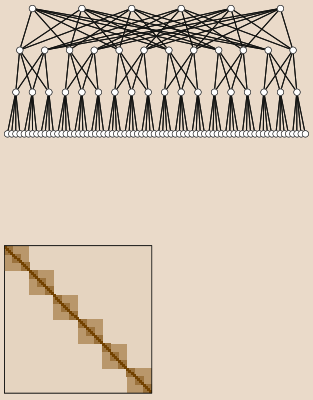
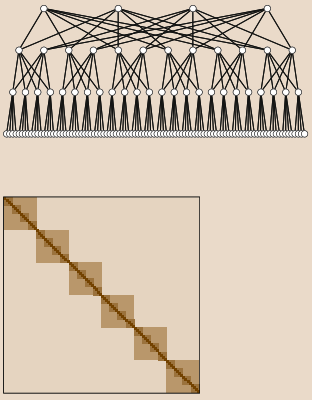
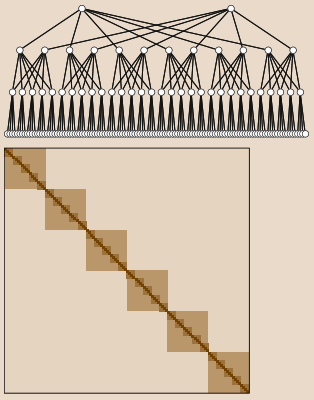
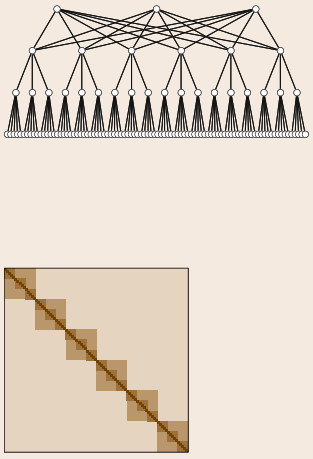
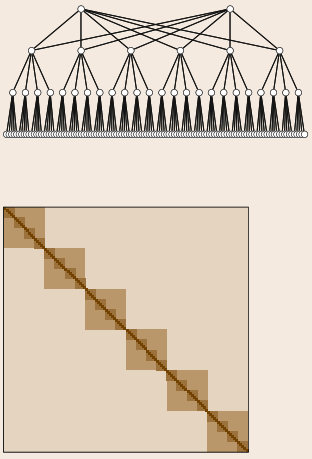
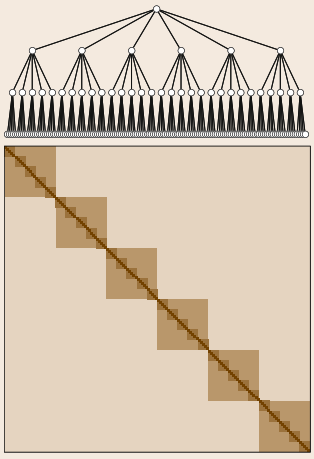
$$\frac{2k-1}{1+\frac{2}{x_1}+\frac{2}{x_1x_2}+\ldots}$$

and the number of internal links per end-point by a factor

$$\frac{k-1}{\frac{1}{x_1}+\frac{1}{x_1x_2}+\ldots}\,,$$

where x_i is the oversubscription realized between level i and $i+1$. A two times repeated 2 : 1 oversubscription thus permits us to reduce router costs by 50% and cabling by 62.5% in a three-level fat-tree. Considering a three-level fat-tree, $r=48$ and the aforementioned prices, this permits us to reduce the cost per end-point down to \$ 557 from \$ 1073 while extending the scalability to close to 50k end-points. Figure 22.6 and Table 22.2 aim at graphically illustrating the impact of oversubscription by considering the nine options available with radix $r=6$. Figure 22.6 shows how routers and link per end-point figures, and consequently cost

Table 22.2 Oversubscription options for two-level fat-trees with radix $r = 6$

	$x_2 = 1$	$x_2 = 2$	$x_2 = 5$
$x_1 = 1$	 Scalability = 54 Routers: 45 Internal links: 108	 Scalability = 72 Routers: 48 Internal links: 108	 Scalability = 90 Routers: 21 Internal links: 108
$x_1 = 2$	 Scalability = 72 Routers: 36 Internal links: 72	 Scalability = 96 Routers: 40 Internal links: 72	 Scalability = 120 Routers: 44 Internal links: 72
$x_1 = 5$	 Scalability = 90 Routers: 27 Internal links: 36	 Scalability = 120 Routers: 32 Internal links: 36	 Scalability = 150 Routers: 37 Internal links: 36

Hops separating end-point pairs: 0 | 2 | 4 | 6

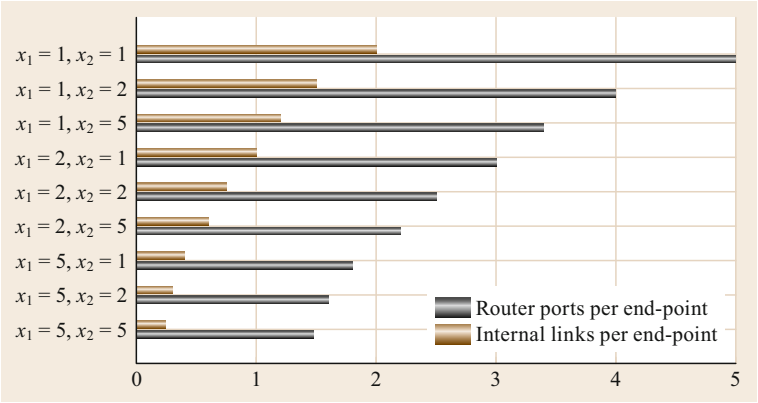


Fig. 22.6 Effect of oversubscription factors on a two-level fat-tree with six-ported routers. If no oversubscription is applied ($x_1 = x_2 = 1$), two internal links (one per level) are associated with each end-point. These two internal links occupy four router ports, in addition to the *entry* router port used by the end-point, thus five ports in total. On the contrary, if the highest level of oversubscription is applied at both levels ($x_1 = x_2 = 5$), the number of internal links per end-point falls drastically to ≈ 0.24. This demonstrates that oversubscription is a powerful tool to tailor the performance, and thus cost, of an interconnect

per end-point, decrease as oversubscription factor x_1 and x_2 are increased. Table 22.2 shows how the scalability, and the routing distances, are affected.

Substantial savings can be obtained by means of oversubscribed fat-trees, but over-subscription also imposes limits on the verbosity supported in specific cases [22.38]. Consider once again an 2 : 1 oversubscription ratio applied twice. The maximum network verbosity factor supported is divided by a factor of four if the entire traffic needs to progress up to the tree root to reach the final destination. Oversubscription factors should therefore be determined with adequate knowledge of the expected traffic patterns. If intense traffic flows are mostly concealed along the diagonal of the traffic matrix, indicating that end-points mostly exchange traffic with the end-point with the next immediate index, relatively high oversubscription ratios can be tolerated. In those cases the traffic pattern is said to show locality. In contrast, if traffic is expected to be mostly exchanged between end-points separated by two or three fat-tree levels, and therefore has to operate four or six hops to reach its destination, oversubscription will act as a bottleneck.

It is worth nothing that when the largest part of the traffic must reach the root of an overprovisioned tree, i.e., for traffic patterns without locality, the bottleneck at the root level renders the injection bandwidths underutilized. This shows that for traffic patterns without locality, it only makes sense to linearly reduce the available bandwidth everywhere in the tree if costs have to be compressed. For traffic patterns with locality, in contrast, costs can be saved by gradually tapering bandwidth across the tree levels, following the effects of locality. The effect of fat-tree network configurations on realistic workloads executed on very large systems has been analyzed in the literature, notably by *Jain* et al. [22.42]

Also note that more complex constructs of tree-inspired topologies exist, in particular the orthogonal fat-tree, which trades immunity to adversarial traffic patterns for improved scalability [22.43].

Directed Topologies

Fat-trees oblige traffic to take an even number of hops through the topology. Two hops are minimally required to ingress and egress the interconnect. However, unless two end-points share the same *entry-point router*, their communications must involve at least four hops, including hops from and to an end-point (see topology examples in Table 22.2).

It is possible to reduce this minimal number of hops for nonentry router-sharing end-points to three by considering *direct* topologies where every router acts as an entry point for end-points. The term direct topology originates from designs where compute nodes fulfill routing functionalities themselves, thus communicate *directly*. Historically, direct topologies also often tried to reproduce the locality patterns shown by specific distributed applications.

Before diving into descriptions of the most common direct topology constructs, it is worth analyzing the per end-point requirements in terms of links and routers of direct topologies.

Consider first that in the worst case, all traffic travels over as many hops as the diameter D. Under this assumption, each bit of traffic will leave a router D

times for another router, leave a router once for the final destination, and thus leave $D+1$ routers in total. Over the r ports of a router, $r(D/(D+1))$ ports should thus be assigned to links connecting routers, leaving $r(1/(D+1))$ ports available on every router to accommodate end-points.

From there, one can deduce that the number of end-points per router is $r/(D+1)$, and thus that the router per end-point is $(D+1)/r$. For example, taking $D=3$ and $r=48$, $r(D/(D+1))=48\frac{3}{4}=36$ ports out of the 48 will be devoted to internal links, while $r(1/(D+1))=12$ ports are available on each router for end-points. The number of routers per end-point is $1/12$. Considering again a price of \$ 8000 for such a router, the *router cost per end-point* is \$ 666.

Further considering all routers similar among themselves, a router offers r ports and accommodates $r/(D+1)$ end-points, so each end-point requires $r/(r/(D+1))=D+1$ ports, and thus $(D+1)/2$ links (as two ports necessarily translate into a link). As a general rule, diameter 2 and 3 direct topologies thus require ≈ 1.5 and ≈ 2 links per end-point respectively, including the link connecting the end-point to the interconnect, to offer enough bandwidth in total to carry traffic over the longest paths. Situations with fewer links per end-point denote oversubscribed direct topologies.

Toruses. A typical example of a topology mimicking a traffic pattern is the 3-D torus, that is ideally suited for so-called 3-D stencil-based parallel applications. This can be, for instance, a fluid dynamics simulation in a 3-D volume by means of finite-element methods. A portion of the 3-D volume (a cube) is assigned to each compute node. At each step of the simulation, the compute node must exchange boundary conditions with the six other nodes responsible for the six neighboring cubes. If these six nodes are directly connected to the original compute node, traffic undertakes a minimal number of hops onto the topology. IBM's BlueGene/L supercomputer series was based on a 3-D torus [22.6].

A 3-D torus requires two ports to navigate on each dimension with the six neighboring nodes, plus at least one port to connect to the end-point. A 3-D torus is thus advantageous for its low requirement in number of ports per router of seven, independent of the size of the interconnect. However, if very suited for 3-D stencils, 3-D toruses are badly suited for arbitrary traffic patterns. Hence, the worst-case number of hops required in a 3-D torus is proportional to the cube root of the number of end-points. A 3-D torus of size $11\times11\times11$ shows for instance a diameter of 15. This diameter far exceeds that of a three-level fat-tree with diameter of six only.

To offer better support to arbitrary traffic patterns, toruses can be constructed in higher dimensions. For instance, the above-mentioned K computer uses a six-dimensional torus of dimension $2\times2\times3\times5\times5\times6$ [22.44]. IBM's BlueGene/P and Q series employed a 5-D-torus interconnect [22.28]. Diameters in 5-D or 6-D toruses are reduced to fifth- or sixth-degree roots, but nevertheless remain proportional to the scale. To alleviate this scaling limitation, direct topologies offering constant diameter (within their scalability limits) have also been proposed and are introduced next.

Flattened-Butterfly Topologies. As in toruses, routers in a flattened-butterfly topology [22.45] are organized as an n-dimensional lattice (where $n>1$), but instead of connecting them with their neighboring routers exclusively, they are connected to all other routers that share at least one coordinate on the lattice. If the lattice is of dimension 4×4, the router at coordinate $(0,0)$ will be connected to $(0,1)$, $(0,2)$ *and* $(0,3)$ along the x-axis, and $(1,0)$, $(2,0)$ *and* $(3,0)$ along the y-axis. This construct, also called Hyper-X [22.46], has diameter n, as in the worst case one hop is required on each dimension. According to the above-mentioned method, $rn/(n+1)$ ports among r can be devoted to other routers ($r/(n+1)$ per dimension). The lattice can be thus of size $1+r/(n+1)$ in every dimension, leading to a number of routers of $(1+r/(n+1))^n$ and finally to a scalability of

$$\left(\frac{r}{n+1}\right)\left(1+\frac{r}{n+1}\right)^n .$$

In the one-dimensional case, which corresponds to a full-mesh topology in which any router is connected to any other router, the diameter is one, so $r/2$ ports are allocated to end-points, and the $r/2$ remaining ports are used to connect to $r/2$ other routers. 1-D flattened butterflies thus involve $r/2+1$ routers in total, leading to scalability of $r/2(r/2+1)$. With $r=48$, 600 end-points can be accommodated.

In the 2-D case, the topology is scalable to up to 4624 end-points with $r=48$. This is vastly superior to the two-level fat-tree, which scales to 1152 end-points only, for an identical number of hops. However, as shown in Fig. 22.7, if n-dimensional flattened butterflies have overall all the bandwidth required to support as many two-hop flows as end-points, this bandwidth is not necessarily available at the right place and congestion might occur, something a fully provisioned fat-tree is immune against.

Moore Bound and Scalability Limits

Being able to use a topology of diameter $D-1$ instead of D generally permits us to decrease global interconnect latencies by $D/(D-1)$, and the number of links

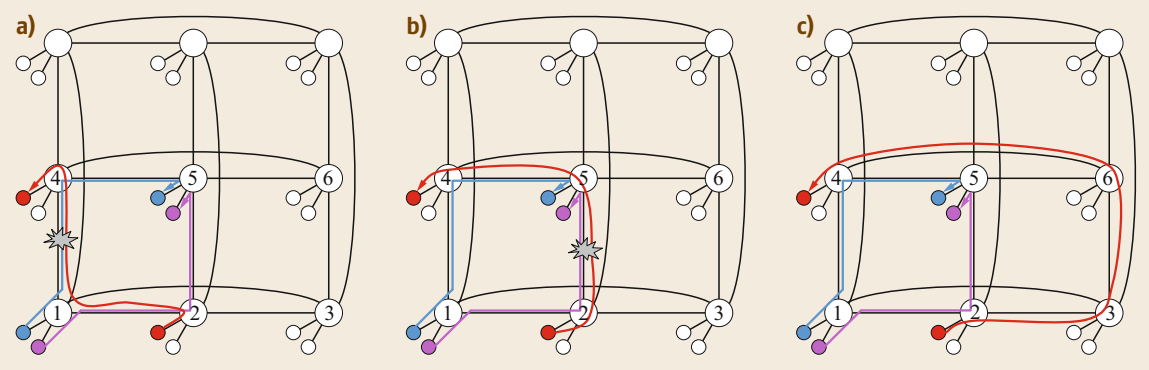


Fig. 22.7a–c Example case of blockingness in a 2-D-flattened butterfly even though only $r/(D+1) = 6/(2+1) = 2$ end-points are connected to each router. As long as two routers do not share a coordinate, there are D shortest paths available between them. The $D = 2$ shortest paths leading to router 4 from 1 are shown in *blue* and *magenta*. If the two end-points connected to router 1 need to send data at full rate to the two end-points connected to router 4, these two shortest paths can be utilized. However, this allows no bandwidth on the shortest paths leading from router 2 to router 3: neither on path 2–1–3 **(a)** nor on path 2–4–3 **(b)**. Traffic exchanged by the end-points shown in *red* can be routed over a nonminimal distance path, for example over 2-3-6-4 **(c)**. However, this path has three hops and thus consumes three units of bandwidth, one more than the $D = 2$ units allocated to each end-point

and routers per end-point by $(D+1)/D$. Specifically, achieving diameter 2 instead of 3 saves 33% in latency and 25% on link and router costs, while achieving diameter 3 instead of 4 decreases latency by 25%, and costs by 20%. There is therefore a great incentive to find low-diameter topologies with large scalability.

A straightforward way to decrease the diameter is to use routers with more ports. For instance, a three-dimensional flattened butterfly is required to support 15 000 end-points with 48 port routers. However, if routers with 72 ports are available, a 2-D flattened butterfly becomes possible (as a 25×25 lattice containing 625 routers, each supporting 24 end-points). Routers with higher numbers of ports, however, may not be available or may exhibit high cost/performance ratios, as developed in Sect. 22.3.1, *Routers*. There is therefore a strong incentive to construct topologies whose scalability is maximal for a given diameter and number of ports.

A scalability bound for topologies of diameter D built around routers with r ports can be obtained through the following steps. Assume again that $R = r(D/D+1)$ ports are allocated to connections between routers. Let us call this parameter R the internal radix, i.e., the radix of the interconnect facing ports. From a given *origin* router in the topology, $R = r(D/D+1)$ other routers are reachable in one hop; we call these *level 1* (L1). Through each of these R L1 routers, another $R-1$ L2 routers at most can be reached in two hops, thus no more than $R(R-1)$ routers in total can be reached in two hops. Following a recurrence relationship, no more than $R(R-1)^{D-1}$ routers can be reached in D hops. That generally means that a topology of diameter D cannot have more than $1+R+R(R-1)+\ldots+R(R-1)^{n-1}$ routers. Notice that each router different than the root one has one port deducted from R to connect to its *parent* router. As each router can accommodate $r/(D+1) = R/D$ end-points, no more than

$$\frac{R}{D}\left[1+R+R(R-1)+\ldots+R(R-1)^{D-1}\right] \quad (22.1)$$

end-points can be supported. For $D = 2$ and $r = 48$, $R = 32$ and the maximum scalability is 16 400. This is more than three times the scalability of the 2-D flattened butterfly. This scalability bound, evaluated for different diameters and as function of radix in Fig. 22.8, is called the Moore bound (after Edward F. Moore, not to be confused with Gordon Moore of Moore's Law). *Besta* and *Hoefler* et al. [22.34] have shown that diameter 2 topologies achieving more than 80% of the maximal Moore bound scalability could be found, as well as diameter 3 topologies achieving more than 60% of the bound.

The Moore bound, as explained above and given by (22.1), considers that always $R(R-1)^{D-1}$ routers are located at distance D of any origin router. A more relaxed definition is given by the generalized Moore bound, which allows the last *layer* to be only partially filled [22.47].

22.3.4 Dragonfly Topology

So far, topologies have been presented from a very theoretical perspective, mostly ignoring practical aspects

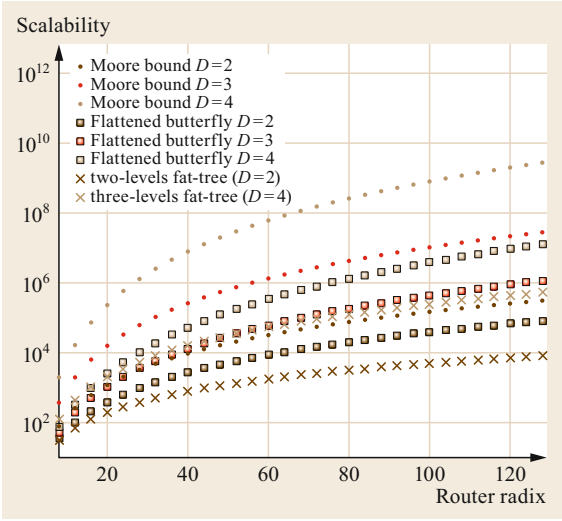


Fig. 22.8 *Dots*: Maximal theoretical topology scalability according to the Moore bound (22.1); *squares*: scalability of n-dimensional flattened butterfly; *crosses*: scalability of two- and three-level fat-trees without oversubscription, for diameter D and router radix r

of field deployments. In particular, the cost difference between electrical and optical links, already introduced in Sect. 22.3.1, *Links*, has not been taken into account. As developed earlier, this cost difference is still relatively important. There is therefore a strong incentive to take the cost of the different links into account when designing a topology, and more specifically, when making oversubscription decisions. This has led interconnect designers to design topologies establishing a clear distinction between long, intercabinet links, and short, intracabinet ones. The dragonfly is the most widespread examples of such constructs.

The dragonfly topology, introduced by *Kim* et al. [22.48], is a direct topology that strongly relies on the concept of groups. A group is composed of a collection of routers assumed to be physically located in a relatively compact volume (typically within the same cabinet), along with the end-points directly attached to them. Under this assumption, links connecting routers belonging to the same group among themselves, called intragroup links, can remain based on conventional electrical transmission technologies. Links connecting the end-points to the routers are similarly assumed to be electrical. In contrast, links connecting routers across group boundaries, called intergroup links, are optical as they generally traverse long distances.

The dragonfly topology can be seen as hierarchical. Two end-points may be connected to the same router, connected to two distinct routers but part of the same group, or to two distinct routers belonging to different groups. An optical link is utilized only in the latter case, i.e., only on the upper level of the hierarchy.

When two end-points located in two distinct groups G1 and G2 communicate, their traffic, after emerging at the *access* router, is first routed within the G1 group to a *gateway* router that provides an intergroup link toward G2 (first hop). The gateway router sends the traffic over the intergroup link (second hop) and upon arrival in the G2 group the traffic is dispatched to the *access* router of the destination end-points (third hop), to finally be distributed to the end-point itself. The dragonfly topology has therefore diameter $D = 3$.

Intragroup Level Dragonfly Connectivity

Within groups, bandwidth is generally made available in large quantities, taking advantage of the relatively low cost of electrical cables. For groups up to 10–15 routers, a full-mesh connectivity can be adopted, whereas for larger groups, a 2-D-flattened butterfly is generally considered.

In the interconnection network of Cray XC series [22.49], taken here as an example to describe a dragonfly group, 96 routers form a group. Each group is internally organized in a 2-D flattened butterfly with lattice $(x = 16) \times (y = 6)$. Each router corresponds to a blade. Sixteen blades form a chassis, which thus represents a *horizontal slice* of the flattened butterfly, and six chassis form a group. Each chassis includes $(16 \times 15)/2 = 120$ electrical links, implemented as wire traces on the chassis backplane, which form the first dimension of the flattened butterfly. A six-chassis group has 720 such links in total. The second dimension of the flattened butterfly is implemented with $16 \times ((5 \times 6)/2) = 240$ links, which are constructed using twin-ax copper cables.

Intergroup Level Dragonfly Connectivity

Groups are connected among themselves by connecting pair of routers belonging to different groups. The intergroup connectivity can take many different aspects, but in general every pair of groups have as at least one direct connection.

Consider a dragonfly composed of g groups with a routers in each group. Further consider that each router has h intergroup links attached to it. To guarantee one link between each pair of groups, a requirement to ensure diameter 3, each group must have $g - 1$ intergroup links, and thus $h \leq (g - 1)/a$. If $h \leq g - 1$, each router can be directly connected to every other group and the dragonfly becomes effectively a 2-D flattened butterfly [22.50].

Setting h to a low value may result in oversubscribed intergroup bandwidth, thus in bottlenecks and overall computational performance penalties for work-

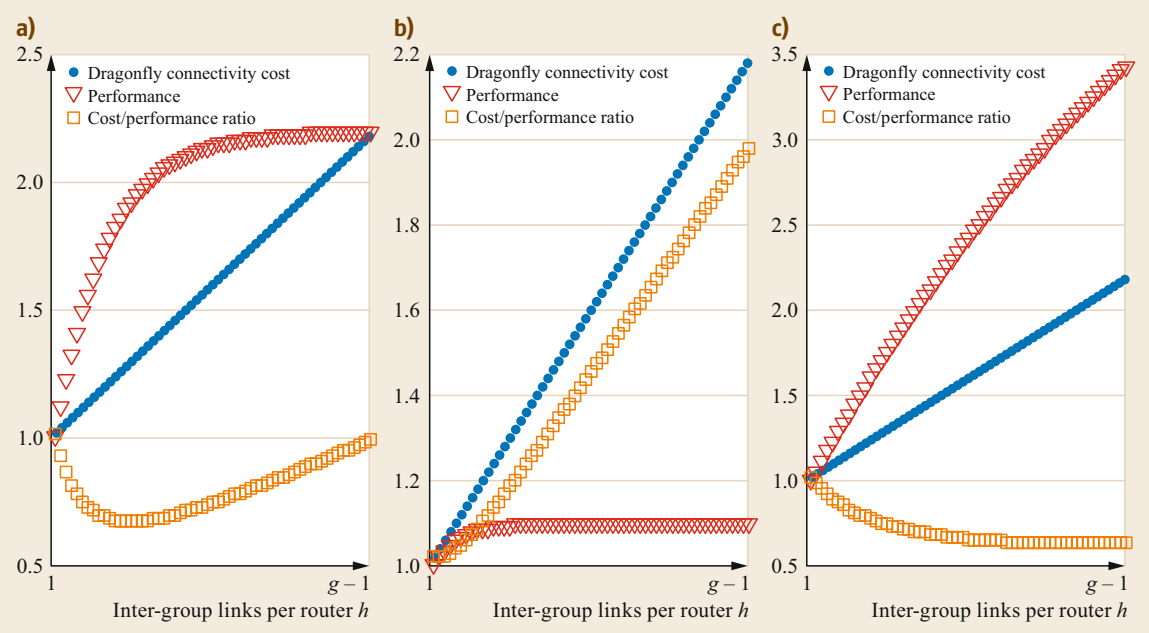


Fig. 22.9a–c Possible scenarios for dragonfly dimensioning. In the three cases, the provisioning of more intergroup links in the dragonfly topology leads to a linear increase of the interconnect cost (*blue dots* – a reference cost of 1 unit is assumed for a dragonfly with minimal connectivity, i.e., $h = 1$ intergroup link per router), until the point where every router is connected to all other groups through $g - 1$ links. Bandwidth provisioning has, however, different impacts on performance (*red triangles*). In the case **(a)**, performance grows steady until every router is connected to half of the groups, and saturates hereafter. In the case **(b)**, provisioning more bandwidth only marginally improve performance. In the case **(c)**, performance growth is steady even until reaching full provisioning, suggesting an injection bandwidth bottleneck. Finally, the cost-performance ratio curve (*orange squares*) indicates which value of h is to advantage: about $h = g/4$ in the case **(a)**, $h = 1$ in the case **(b)** and $h = g - 1$ in the case **(c)**

loads involving large amounts of intergroup traffic. In contrast, setting h to a high value has a major impact on global interconnect cost. Dragonflies based HPC architectures are designed so that the selection of h is at the discretion of the customer/operator. Being able to choose h between 1 and $g - 1$ allows the supercomputer vendor to customize the expensive global bandwidth of the dragonfly, and thus, to a large extent, the final cost, to the need and/or budget of the customer. The supercomputer operator, taking into account the traffic locality properties of the workloads that are destined to the system, makes the final call for h. Figure 22.9 shows how h should ideally be selected. The total cost of a system as a function of h, and relative to the $h = 1$ case, is represented by the dots. Cost grows linearly with h. The performance of the system, also relative to the $h = 1$ case, is illustrated by triangles. Performance is expected to grow monotonically with h, as adding bandwidth to a system should not be detrimental to performance, but is also expected to saturate as the computational resources are limited. Finally, the squares show the cost-performance ratio. Figure 22.9 shows three particular cases. In Fig. 22.9a, performance growth is steady at first but quickly fades as h grows. This results in a sweet spot in the cost-performance ratio curve, which indicates the ideal value of h. In Fig. 22.9b, more intergroup links only limitedly contribute to performance, resulting in an optimal choice of h to be 1. This situation indicates that even with minimal dragonfly connectivity, network resources are generally oversized. Finally, Fig. 22.9c shows the reverse case where cost performance/ratio is continuously improved until $h = g - 1$. This situation is synonym of undersized network resources.

Note that the applicability of this rule is limited by workload execution times prediction capabilities. Also note that the same economical approach can be applied to any part of the interconnect, and more generally, to any design variable of the system. However, changing the amount of memory available inside an end-point, or the bandwidth of the link connecting an end-point to its router, generally requires motherboards to be re-

designed, something impossible to do at procurement time. In contrast, the number of optical links can be provisioned almost independently of the end-point and cabinet design. The dragonfly topology capability to expose an important design variable to the end-user is one of the main factors that drove its adoption.

22.3.5 Higher Networking Layers

Next to offered bandwidth and topologies, HPC interconnects are also characterized by specific operational modes at the link and network layers. The most important of these aspects are mentioned here.

Lossless Flow Control

Intuitively, it is relatively easy to understand that to keep sending packets to a destination that is overwhelmed by previous packets is likely to be counterproductive. In order to prevent gridlocks and subsequent performance collapses [22.51], the amount of traffic that is being injected into a packet-based network must be controlled, something generally described as *flow control*.

Historically, packet-based computer networks have predominantly relied on lossy schemes to achieve flow control. Indeed, the distributed and loosely coupled nature of the Internet has been partly based on the possibility for intermediate routers to drop packets [22.52]: packets that cannot be placed in the queue associated to a link because this queue is full are simply discarded. The recipient, noticing that one packet is missing or that no packets are received at all, informs the sender that congestion is likely underway and that the rate at which data is sent must be reduced.

Lossy flow-control-based networks are well suited for large-scale and poorly structured networks like the Internet but are largely unsuited for HPC. First, the loss of a single packet might cause one of the supercomputer PUs to stall while waiting for the dropped packet to be resent and delivered. The overall parallel performance is very often determined by the performance of the slowest end-point [22.40]. Packet losses can thus have dire consequences in the context of a massively distributed computation. Second, possible packet drops oblige network interfaces to maintain a copy of all packets already sent but not yet acknowledged, which may result in a large memory footprint [22.53]. In addition to these shortcomings directly related to packet loss, lossy flow-control schemes have been shown to lead to poor utilization of bottlenecked links [22.54].

Lossless flow-control schemes have therefore been developed, the most widespread in the HPC context being the *credit-based* one [22.55]. Each emitter port maintains a counter indicating how many free slots remain available at the corresponding remote reception port. Each time a packet is sent, this counter is decremented. If the counter reaches 0, packet injection is interrupted. Injection resumes upon reception of credits. Whenever a slot is made free at the remote port, this port sends a credit back to the sender port. Upon reception of this credit, the counter is incremented [22.56]. A delay exists between the moment a slot is effectively made free, and the point the sender receives the corresponding credit. The situation where the counter reaches 0 even though free slots are available (therefore with credits on their way) is called a credit stall. To avoid credit stalls, buffer depths at the receiver side must be made proportional to the link latency [22.55]. The use of credit-based flow control is thus limited to local links with relatively low latency.

Routing

A supercomputer being a closed world managed by a single entity, end-points and routers alike can be numbered in independent and consistent ways. This allows us to use much shorter address fields in packets: a 16 bits address field can be used to distinguish up to 65 536 end-points, which is sufficient in most cases (to be compared with 48 bits MAC addresses [22.57], or 128 bits IPv6 ones). This also greatly simplifies routing decisions at routers.

The closed-world vision offered by HPC systems also enables routers, in case of congestion, to adapt their forwarding decisions to exploit the path diversity available in the topology. Figure 22.10 illustrates adaptive routing applied to the 2-D torus and dragonfly cases. Adaptive routing, also known as deflection routing, has been studied in the context of torus [22.58, 59], fat-tree [22.60] and dragonfly-based networks [22.48, 61]. Notable implementations are reported in references [22.28, 30, 62]. Because each router unilaterally and independently decides to deflect part of the traffic to an alternate route, adaptive routing does not lead to minimal congestion levels as if routes were calculated by means of a multicommodity flow formulation [22.63]. In return, routers, when performing adaptive routing, are able to react to congestions in microseconds, if not less. This high reactivity is crucial to prevent counterproductive, day-late-and-dollar-short routing decisions. Recently, adaptive routing has been shown to be competitive compared to per-flow (i.e., per source-destination pair) route optimization and resource management using software defined networking (SDN) [22.63].

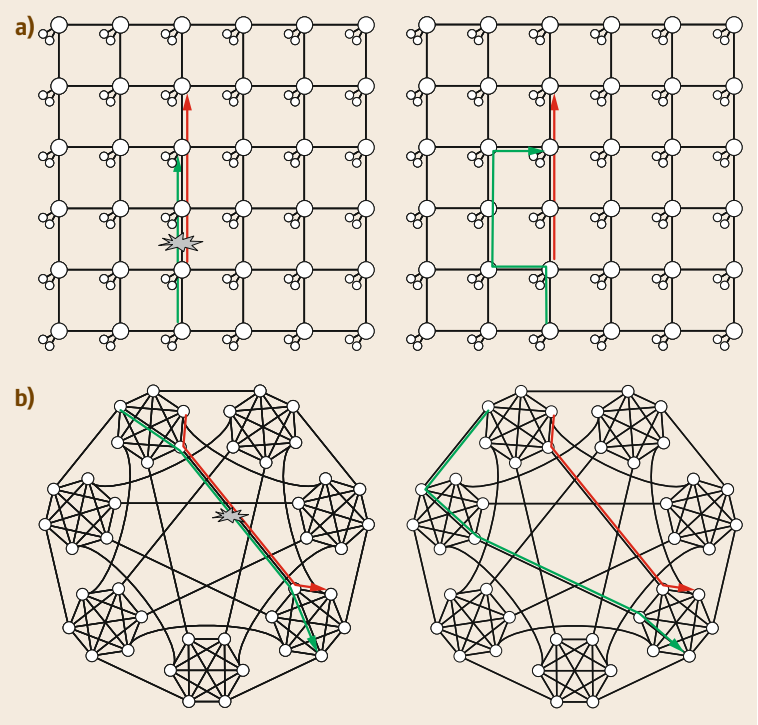


Fig. 22.10a,b Adaptive routing applied to torus **(a)** and dragonfly **(b)** topologies. In the first case, the path is lengthened by two hops (from three to five), and in the second case by one hop (from three to four)

22.3.6 Contemporary HPC Networks: Summary

In this section, the characteristics of modern HPC interconnects have been reviewed. The main rules of thumb to retain are the following:

- The standard link-level bandwidth is 100 Gb/s in 2017. This same year some 200 Gb/s products are being introduced, while 400 Gb/s ones are announced for 2018 or 2019.
- Routers routinely provide between 30 and 80 ports, delivering close to 10 Tb/s of switching bandwidth.
- FLOPS installed in a compute node must be multiplied by a factor 0.001–0.01 to obtain the node injection bandwidth (in bytes). A compute node offering 10 TF thus requires between 100 Gb/s and 1 Tb/s of injection bandwidth.
- Due to mismatch between link-level bandwidth ($\approx$ 100 Gb/s) and injection bandwidth, end-points are increasingly being connected through multiple independent links.
- Diameter is considered the important metric for HPC topology, as it guarantees low latency between any pair of end-points. Keeping diameter 2 for 10 000 end-points requires 42 ports at least (Moore bound). A slim-fly [22.34] construct requires 43 ports, a 2-D flattened butterfly 63 and a 2-D fat-tree with no oversubscription, 142 ports.
- Cost difference between electrical and optical cables must be taken into account in the topology design process. Optical links, the most expensive components, must be oversubscribed in priority. The dragonfly topology is specifically thought to ease optical link oversubscription.
- HPC specific flow-control protocol and routing schemes are deployed in many HPC interconnects and permit high utilization and low latency. End-to-end transfer times are often inferior to the microsecond [22.28]. Transmission delay routinely dominates transfer times.

Owing to the very important role that interconnects play in supercomputers, the HPC community has very high expectancies in terms of performance, in particular in term of latency. At the same time, costs and power budgets must be severely controlled.

It is with the rules of this challenging environment in mind that optical technologies for next-generation interconnects must be developed, as addressed in the next section.

22.4 Future of Optics in HPC Interconnects

Most metallic wire-based transmission lines have a cutoff frequency inferior to the GHz. This causes high bandwidth signals with modulated rates superior to the GBd to be severely attenuated across distances and suffer signal integrity issues. As supercomputers got larger and required more interconnect bandwidth, designers turned to optics for the longest cables to enable longer reach. The ASCI Purple supercomputer, introduced in 2005, is notably known to be the first system to include optical links [22.64]. Since then, most supercomputers have incorporated optical links. As supercomputers head into the exascale era, and as signaling rates are about to transition from 25 Gb/s to 50 Gb/s to support Tb/s-order link bandwidths required by heavyweight end-points, optical cables are more than ever a constituent of HPC interconnects [22.15]. From its debut in HPC, photonics has also constituted a way to reduce the volume occupied by bulky copper cables, in order to facilitate cooling and simplify maintenance. In 2017, optics is principally seen as a means to limit the bulk around computing equipment, and more precisely to reduce the footprint of metallic traces connecting router ASICs (application-specific integrated circuits) to transceiver connectors (such as QSFP (quad small form-factor pluggable)). Hence, these traces increasingly act as a bottleneck to the total switching bandwidth of routers, which, as discussed in Sect. 22.3.1, *Routers*, need to scale beyond 10 Tb/s. To that aim, solutions to bring tens of Tb/s of *optical bandwidth* directly into the router ASIC or in its direct proximity (e.g., on an underlying interposer die or substrate), and offering large bandwidth densities, are under intense investigation.

In this section, the two main technologies envisioned to realize architectures involving *optically connected router ASICs*, VCSELs (vertical-cavity surface-emitting lasers) with multimode fibers and silicon photonics with single-mode fiber-based links, are reviewed. The capability of these technologies to deliver the required levels of *escape bandwidth density*, in particular, is analyzed. More forward-looking technologies such as plasmonics, optical switching, and free-space optics, are also reviewed and their likeliness to address HPC interconnect challenges analyzed.

22.4.1 VCSEL/MMF Link Technology

The VCSEL technology consists of VCSEL arrays that use both space (i.e., fiber parallel) and wavelength dimensions to increase the total link bandwidth [22.65]. An n-fiber, m-wavelength link requires m dies each consisting of n VCSELs, arranged as a one-dimensional array. Each array is taken from a separate wafer with a specific quantum well epilayer design, dedicated to the emission of a specific wavelength. By adhering to typical coarse wavelength division multiplexing (CWDM) wavelength spacing, the separation between each wavelength of emission of subsequent VCSEL channels is on the order of 25 nm in the range of 850–1100 nm. Note that the VCSEL-specific acronym SWDM (shortwave wavelength division multiplexing) has been introduced for this technology [22.66]. In addition to the m VCSEL array, each link extremity also includes arrays of photodiodes fabricated on the same quantum well epiwafers, with typical aperture of 25 µm, 20–25 GHz 3 dB bandwidth, and 0.8 A/W responsivity. Wavelength multiplexing and demultiplexing is realized by means of passive thin-film filters [22.67]. VCSELs and photodiodes are attached to the carrier also supporting the ASIC, while the filters are attached to the male connector.

An SWDM 200 Gb/s link ($m = 4$ wavelengths and 25 GHz PAM-4 modulation) with 50 m reach has been demonstrated [22.68]. A full design with $n = 6$ fibers per direction, $m = 4$ wavelengths, and performing 25 GHz modulation on each wavelength, resulting in 600 Gb/s full-duplex links, has been reported [22.69].

Packaging, Bandwidth, and Bandwidth Density

VCSEL dies successfully passing wafer-level testing are flip-chip assembled through solder reflow. This assembly process is not damaging and guarantees high yield. The solder reflow technology keeps impedances well controlled as no *hanging* wires are involved. This allows for leaner control circuitries. The injection molded female part of the connector is computer vision aligned to the VCSELs and attached to the carrier with solder reflow as well. This guarantees a satisfying alignment of the female part of the connector with the carrier, and consequently of the VCSELs and photodiodes with the WDM filters and multimode fibers. As of November 2017, fibers were spaced 250 µm apart; 126 µm is considered for the future but is subject to yield issues due to reliable stripping and cleaving of the fiber coating at these narrower dimensions. This limitation is an area of advanced development by fiber manufacturers. Connectors can be clipped in and removed by hand, theoretically allowing manual maintenance operations at the connector level.

The optomechanical package is a major limiting factor for bandwidth density. Connector footprint is

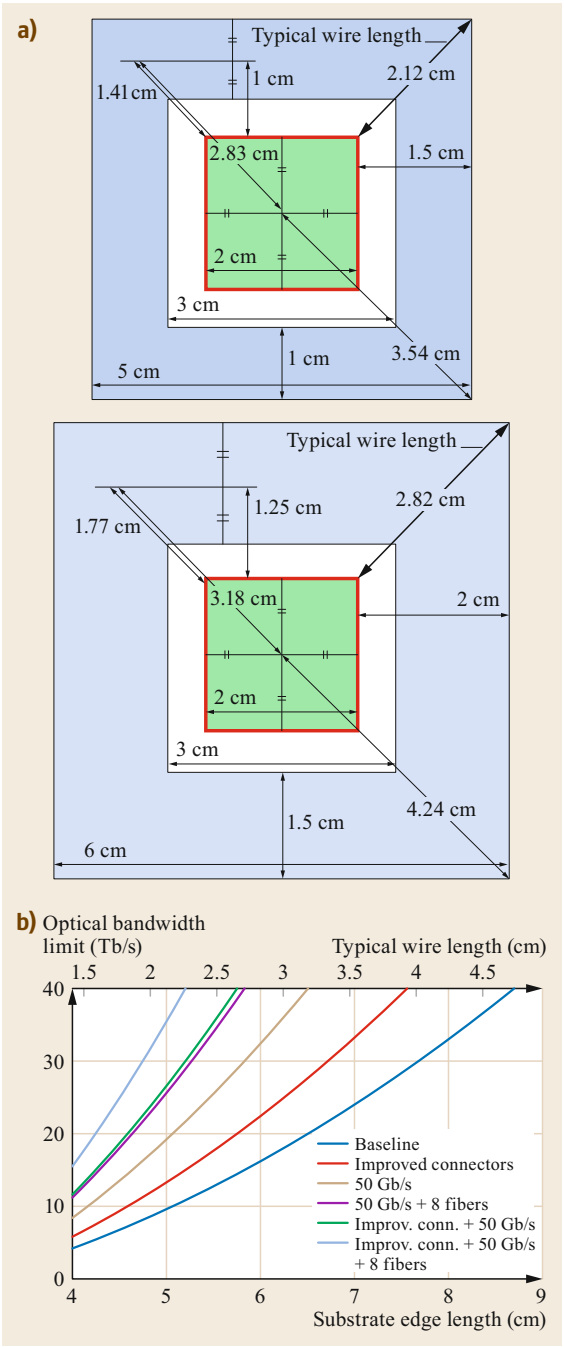


Fig. 22.11 **(a)** Geometrical analysis of area available for VCSEL around router chip, and of typical wire distances. **(b)** Impact of substrate size on wire length (*upper x-axis*) and resulting optical bandwidth ►

typically 1 cm^2 [22.69], which results in a bandwidth density of 6 $(Gb/s)/mm^2$ in the 600 Gb/s full-duplex example case (12 $(Gb/s)/mm^2$ if only one direction is considered). Considering a 8×8 cm large substrate, carrying a 2×2 cm large ASIC, about 56 cm^2 remain available for VCSEL connectors, enabling $50 \times 0.6 = 30$ Tb/s of bandwidth for the router ASIC. The bandwidth density of 6 $(Gb/s)/mm^2$ is likely to scale as:

- PAM-4 signaling is used instead of NRZ (2×)
- Signaling rates evolve from 25 Gbd to 50 Gbd (2×)
- Wavelengths are increased from 4 to 6 (1.5×)
- The number of fibers per connector is doubled potentially using 125 µm spacing (2×).

Altogether, the technology has the potential to scale to ≈ 7 Tb/s per per connector, resulting in $\approx 70\,(Gb/s)/mm^2 = 7\,(Tb/s)/cm^2$. This would result in extremely compact carriers of roughly 4×4 cm if 50 Tb/s of ASIC bandwidth are required. Figure 22.11 illustrates how the available optical bandwidth is affected by the substrate size, considering various scenarios. Also visible on the figure is the relationship between substrate size (bottom *x*-axis graduation) and length of metallic wires (top *x*-axis graduation) connecting the ASIC to the VCSELs and photodiodes.

Electrical escape bandwidth densities shall also be considered in conjunction with the optical. Hence, to reach the VCSELs and their associated connectors, electrical signals must

- Escape the main CMOS (complementary metal oxide semiconductor) die to reach the substrate (*vertical escape bandwidth*)
- Escape the chip reticle (*lateral escape bandwidth*).

Considering a 150 µm pitch between the solder bumps connecting the ASIC chip to the substrate (45 bumps/mm^2), three bumps per lane (VCSEL drive, photodiode, ground), an ASIC die size of 2×2 cm, and assuming that 20% of the bumps dedicated to IO, the *vertical escape bandwidth* is typically

$$\frac{20\% \times 400\,\mathrm{mm}^2 \times 45\,\mathrm{mm}^{-2} \times 25\,\frac{\mathrm{Gb}}{\mathrm{s}}}{3} = 30\,\frac{\mathrm{Tb}}{\mathrm{s}}\,.$$

As for the lateral escape bandwidth, if a 100 µm pitch between two wires on the substrate, a 80 mm long chip edge, and four metal layers in substrate are considered, the total lateral bandwidth is

$$\frac{80\,\mathrm{mm}}{10\,\mathrm{mm}} \times 4 \times \frac{25\,\frac{\mathrm{Gb}}{\mathrm{s}}}{3} = 26.6\,\frac{\mathrm{Tb}}{\mathrm{s}}.$$

Vertical and lateral escape bandwidths may thus constitute another obstacle to scalability of router bandwidth envelope. They can be scaled if a larger ASIC

chip is considered, although the lateral bandwidth limitation will only scale as the square root of the chip surface. All these calculations are based on a 25 Gb/s bit-rate. Transitioning to 50 Gb/s, for example using PAM-4 modulation, automatically doubles bandwidth.

Transmission Quality and Reach

At 25 Gb/s per wavelength, VCSEL technology can reach distances up to 300 m with standard OM3 fibers. Specialty fibers with optimized modal bandwidth ($\approx$ 5 GHz/km) may be required to carry higher speed signals (50 Gbd) over distances exceeding 50 m, which is sufficient for HPC applications.

With only four to six wavelengths per fiber, wavelength spacing can be kept above 20 nm, which is large enough to tolerate VCSEL wavelength drift due to temperature variations. Thin-film filters can be designed with a band pass wide enough to adequately capture the signal even under changing temperature conditions. Outside of the band of interest, their filtering effect is high and guarantees high crosstalk suppression.

Beyond six wavelengths, the tolerance to temperature variation decreases sharply due to the reduced wavelength spacing. Cascaded losses associated with the multiple reflections taking place within the thin-film filters become significant.

Cost

The VCSEL packaging option allows all components to be aligned and attached through a standard solder reflow process. VCSELs, in particular, are treated as simple *capacitors* by integrated circuit packaging houses. The yield of the assembly process has been shown to be excellent and does not affect costs significantly. VCSELs can be wafer tested and show yields $> 80\%$. Injection molded plastic connectors can be produced at very low cost. For these reasons, it is reasonable to expect the cost of the VCSEL-based solution to drop largely under the critical 1 \$/(Gb/s) mark. This cost is subject to improvement as bandwidth is scaled and/or large quantities are produced.

Power

The power consumption of a directly modulated VCSEL driver is of the order of ≈ 20 mW [22.71], which translates into ≈ 0.8 pJ/bit for 25 Gb/s. Reaching higher modulation speeds requires slightly raised bias currents only, keeping the driving power almost flat. From a pure VCSEL perspective, high symbol rates are thus advantageous for power efficiency.

The VCSEL driving power represents, however, an almost marginal fraction of the total power consumption of an end-to-end link. Pre-emphasis is required before the final driver stage to overcome VCSEL diode bandwidth and equalization is similarly required after the transimpedance amplification (TIA) and limiting amplifier stages at receiver side. Clock recovery mechanisms must also be accounted for, as well as serialization-deserialization operations. The consumption of each of these elements is of the order of 1 pJ/bit [22.71].

Reliability

The effect of temperatures up to 70 °C as well as of elevated current levels on VCSELs' reliability and wear out have been evaluated. VCSEL lifetimes superior to ten years have been extrapolated. Thermal simulations have shown that temperature should not exceed 70 °C in the close proximity of the VCSELs. Aluminum being a very reactive element, aluminum-free designs should be favored to reach high reliability. VCSELs with active regions comparable to the ones used in pump lasers, showing lifetimes > 25 years, have been designed. The thin-film filters used to demultiplex SWDM signals have routinely been deployed in satellites over the past years and are known to be robust to temperatures well beyond 70 °C, being deposited and annealed at high temperatures. All these elements should concur with excellent reliability figures for VCSEL-based interconnects.

22.4.2 Silicon Photonics Technology

Silicon photonics technology is based on the idea of manipulating light by means of nanoscale structures imprinted in a silicon die using lithographic processes inspired by [22.72, 73] or identical to CMOS [22.74]. After 50 years of silicon semiconductor industrial developments, the electrical properties of silicon are well understood. Silicon photonics builds on this experience: changing electrical properties results in changing optical properties as well. Silicon photonics is currently exploited by several vendors including Mellanox to produce 100 Gb/s active optical cables and short distance transceivers for HPC applications. External modulation is based on Mach–Zehnder interferometers [22.75] or electroabsorption [22.76]. Microring resonators with an integrated PN junction have also been proposed to perform modulation [22.77] and WDM links offering 320 Gb/s on a single fiber have been demonstrated using eight ring resonators and eight ring filters [22.78]. A comprehensive analysis of ring resonator-based links is provided in [22.79].

Packaging, Bandwidth, and Bandwidth Density

Silicon photonics naturally meshes with single-mode optics and thus requires single-mode fibers to be cou-

pled to the optical die. Single-mode optical inputs/outputs allow us to take advantage of the long reach enabled by the low dispersion of single-mode fibers. However, single-mode optics pose a far greater challenge in terms of packaging than multimode optics. An alignment precision of the order of 10% of the mode field diameter, i.e., around one micron, is required, compared to 5–10 μm for multimode fibers in the VCSELs technology.

For packaging solutions that use vertical coupling, grating couplers with approximately 10 micron spot size are the de facto standard, but fiber coupling that is cost effective and scalable is still lacking in industry for these alignment tolerances. Existing solutions still suffer from lack of maturity and result in long, thus expensive, coupling processes, as well as significant optical losses. Advanced, vision-based optical IO connector alignment equipment is under development. Alternative techniques, among others evanescent coupling, have also been investigated [22.70]. Plastic ferrules for connectors are in practice excluded as they cannot tolerate temperature changes observed on the chip edge. Ferrules should thus be made of metal or of a custom material with a coefficient of thermal expansion matching that of silicon. In comparison, edge-coupled solutions that use nitride layers to expand the beam profile and use spot-size converters are also under exploration. This latter technique has the advantage of being polarization insensitive but comes at the expense of larger footprint, increased process steps and requires surface preparation such as dice/cleave or dry etch to expose the correct facet for fiber coupling. Once this facet is prepared, pick-and-place machines can be used to quickly and efficiently align fiber ribbons to the chip to enable high bandwidth density. It should also be noted that such optical chips are thus not a candidate for wafer-level testing and require proxy measurements to make sure the photonic circuits are accurate, which requires the use of vertical grating couplers.

With a projected per-fiber bandwidth of the order of 200–500 Gb/s for silicon photonics, connectors accommodating fiber ribbons with 8–16 fibers are envisioned to reach tens of Tb/s of optical escape bandwidth. Eight connectors each offering eight fibers in each direction (thus 64 fibers in each direction) and 360 Gb/s per fiber (as 24×15 Gb/s) offers for instance 23.04 Tb/s of total bandwidth envelope. The optical escape bandwidth can be scaled to ≈ 50 Tb/s by considering more numerous channels (40 or more, as in DWDM (dense wavelength division multiplexing) applications), higher modulations rates, and/or more advanced modulation formats [22.80]. The number of connectors attached to a chip can also be reasonably scaled. Waveguide pitch is not a limitation as distances as short as 20 μm between couplers have been demonstrated. As for fiber pitches, assuming 250 μm pitch between two fibers in a cable, 480 fibers can theoretically be attached around the 120 mm long edge of a 3×3 cm interposer. Connector mechanical stability is the eventual limiting factor. Couplers must be separated in connectors of reasonably small size to limit mechanical stress. Consider a 16-fiber connector, whose fibers occupy 4 mm of the edge and with ferrule/attachments that occupy 4 mm as well. Three such 8 mm diameter connectors can be mounted along a > 2.4 cm long edge, four along a > 3.2 cm long one. The maximum number of fibers affixed to a SiP chip can thus be estimated to be 256. This results in 128 full-duplex connections. Combined with a per-fiber bandwidth of 500 Gb/s, a total optical escape bandwidth of 64 Tb/s can be foreseen.

Similar to the VCSEL-based approach, the SiP solution is also subject to electrical escape bandwidth limitations if a dual chip approach is selected and the CMOS drivers are not incorporated into the same die (as shown in Fig. 14 of [22.81], for instance). A ring resonator is typically associated with six control pads: heater control plus ground, diode junction control plus ground, and photodiode plus ground (the photodiode being required to provide feedback for thermal stabilization [22.82]). Since two rings are required per wavelength (neglecting the last wavelength, which requires no ring for filtering), a *lane* in the SiP solution requires up to 12 bumps. However, since the *driver die* is likely to also be made of silicon, and thus to exhibit similar thermal expansion coefficients, finer bump pitches (40–70 μm) than for the VCSEL solution (150 μm) are allowed. This translates into vertical escape bandwidth in excess of 40 Tb/s:

$$\frac{20\% \times 400\,\mathrm{mm}^2\ \text{(ASIC area devoted to IO)}}{0.07 \times 0.07\,\mathrm{mm}^2\ \text{(area per bump)}} \times \frac{15\ \frac{\mathrm{Gb}}{\mathrm{s}}\ \text{per lane}}{6\ \text{bumps per lane}} \simeq 40\ \frac{\mathrm{Tb}}{\mathrm{s}}\,.$$

As in the VCSEL case, the area associated with the bumps dominates the silicon real estate requirements of the IO functions. The same holds at the photonic layer: the four pads required to control a ring occupy an area (with 50 μm pitch) of 200 μm × 200 μm. This guarantees sufficient spacing between the rings and other optical components. Silicon photonics, being based on single-mode fibers, allows links to span over distances superior to 100 m. Pending appropriate link budgets, quasi error-free operation (BER $< 10^{-15}$) can be achieved.

Ring-resonator-based modulators should allow energy efficiencies around 1 pJ/bit (not counting additional circuitries). Mach–Zehnder-based modulators are

slightly less efficient. Similar to the VCSEL-based solution, additional circuits are required to operate the transition between the digital and optical formats. However, the silicon photonic solution being able to support a potentially large number of wavelengths, lower bitrates can be envisioned, which slightly reduces the requirements for pre-emphasis and SERDES (serializer/deserializer). This is expected to result in slightly better energy efficiencies than in the VCSEL case.

22.4.3 Other Forward-Looking Technologies for HPC Interconnects

Optical interconnects for HPC systems based on directly modulated VCSELs combined with MMF (multimode fiber) are expected to emerge around the year 2020. These VCSEL-based interconnects will likely be the ones equipping the first exascale systems. Interconnects based on integrated silicon photonics will likely emerge a couple of years later as packaging methods will have matured and cointegrated photonics-specific issues will have been solved.

Beyond this point it is not clear which other photonic technologies can further contribute to HPC interconnect performance. One of the major limitations of photonics resides in the relative bulkiness of the components: a waveguide in silicon has a cross-section close to a tenth of μm^2. The footprint of a ring resonator in silicon is close to $1000\,\mu m^2$; that of a photodiode only slightly inferior [22.83] while VCSELs are slightly larger. As wavelengths of interest are in the μm region, guiding structures made of conventional materials cannot be made much smaller than a tenth of μm while devices interacting with light must be at least ten times longer or wider than the wavelength. This is respectively one and three order of magnitude larger than transistors in the most advanced CMOS processes. The important footprint of photonics devices not only limits the bandwidth densities as calculated above. It also prevents scaling of electrical parameters as the resistance (R) and capacitance (C) of electro-optical components in charge of converting electrical signals into optical ones and vice versa [22.84]. The receiver sensitivity, for instance, which is one of the main determinants of the energy efficiency of an optical link as it dictates in most cases the amount of optical power the laser must supply, is strongly determined by the capacitance of the photodiode [22.85].

The footprint drawback of photonics could be alleviated by turning to plasmonics [22.86]. Plasmons represent oscillations of the free electrons present in a metallic volume. Plasmons propagating along the surface of a metallic volume are called surface plasmons. Surface plasmons are very sensitive to environment changes, can convey energy in a very dense form, and can couple with light. This results in the capacity of plasmonic devices to confine light in subwavelength dimensions [22.87], a feature that can be exploited to minimize footprint [22.88].

Recent research has shown that compact and ultrafast modulators based on plasmonics can be realized [22.89]. The application of plasmonics for photodetectors is reviewed by *Brongersma* [22.84]. Plasmonics have attracted major attention from academia. However, industrial applications of plasmonics, as for instance in HPC interconnects, remain far looking and major gaps preventing adoption must be addressed. In particular, the inclusion of plasmonic materials, such as graphene or polymers, in lithographic processes such as CMOS in a scalable cost-effective way remains to be demonstrated. Moreover, the capability of plasmonic devices to operate at high rates can only be exploited if electrical driving circuits supporting the same rates are available. For speeds above 100 GHz, the use of a lithography process optimized for extremely high frequencies (EHF) such as silicon-germanium is almost mandatory, which drift apart from the high integration goal described earlier in this section. Finally, all demonstrations of communication subsystems involving plasmonics have so far been characterized by high optical losses. Unless gain can be incorporated in plasmonics-based devices [22.90], these losses will severely limit industrial applicability.

A handful of technologies have also been evoked to complement HPC optical interconnects as described in Sect. 22.3. Optical spatial switches have been proposed to reshuffle the fiber connectivity to achieve a better matching between connectivity and traffic matrices over relatively long periods of time [22.91–93]. Technologies underlying the realization of such spatial optical switches are reviewed by *Cheng* et al. [22.94]. Optical switching has also been evoked as a possible replacement for electrical packet switches. However, as summarized in reference [22.15], optical switching at the packet or even flow level is heavily limited by the lack of buffering capabilities in the optical domain.

Instead of reshuffling the topology by means of optical switches, *Fujiwara* et al. [22.95] and *Hu* et al. [22.96] have proposed connecting racks with free-space optical (FSO) links in lieu of optical fibers. By adapting the beam direction, the connectivity can be fully reconfigured. Free-space connections also benefit from a faster propagation time of light in air than in silicon, and from Euclidean distance instead of typical Manhattan routing. Altogether, FSO-based interconnects could reduce the time-of-flight latency by a factor of two. Given that this time-of-flight latency typically represents more than half of the total latency, this can

enable nonnegligible performance gains. FSO links, however, are affected by scintillation and mechanical vibrations. Losses pertaining to this effect must be compensated at the expense of energy efficiency. Moreover, a path to fabrication of FSO links at a price comparable to that of optical links as detailed in Sect. 22.3.1, *Links*, is to be demonstrated. FSO terminals, finally, must be placed at the top of the rack and can thus be located more than one meter apart from routers or adapters. Communicating data over this distance at 100 Gb/s speeds incurs additional cost and energy consumptions.

The use of FSO links for interchip communications has also been proposed as a way to extend transmission capabilities beyond that of a single fiber [22.97].

22.5 Summary

As high-performance computing applications proliferate, there is a strong global appetite for HPC systems of increasing capability. As shown in Sect. 22.2 of this chapter, the supercomputers required to tackle today's largest computational challenges necessarily rely on chip parallelism to reach tens of petaFLOPS, and soon one or more exaFLOPS, of installed computing power. This places the interconnect at the center of the technical and scientific effort to increase HPC performance.

Building large-scale HPC interconnects is primarily achieved by minimizing the costs, as well as, and increasingly, power consumption. The performance of an interconnect, along with its cost and power consumption, must be related to those of the end-points. These relationships have been detailed in Sect. 22.3. By considering a cost of $2000 for one installed TF, and a network verbosity factor of 0.01 byte/flop, the cost of the 10 GB/s = 100 Gb/s of interconnection bandwidth required per end-point can be upper bound by several hundreds of dollars. The analyses and explanations provided on the topology level in Sect. 22.3.3 justify this rule of thumb. More importantly, the understanding of simple rules of thumb, and of scalability bounds such as the Moore bound (Sect. 22.3.3, *Moore Bound and Scalability Limits*), allows one to remark that no massive cost savings should be expected from particularly innovative topologies as long as no oversubscription is applied. Costs can be tapered by means of oversubscription, but tapering ratios should be selected knowing the impact they will have on HPC application performance.

A budget of $100–200 for 100 Gb/s directly sets a price objective of ≈ 1 $/(Gb/s), a figure that is thus far out of reach for optical technologies. The most immediate and stringent requirement of HPC in terms of optical technologies thus consists today of ways and means to rationalize transceiver development costs. As detailed in Sect. 22.4, multiwavelength VCSEL-based CWDM links and silicon photonic integration are among the main technologies envisioned to attain the cost and power efficiency metrics dictated by HPC trends.

Highly integrated optical technologies are also required to allow total switching bandwidths of packet routers to scale beyond 10 Tb/s. High FLOPS density compute nodes totaling tens of TFs have an urgent need for link bandwidths of 400 Gb/s and beyond. At such rates, the reach of passive cables will likely be reduced to ≈ 1 m, obliging packet routers to communicate optically on most of their ports. Here as well, VCSEL/MMF and silicon photonics technologies are the best positioned to solve the packet router bandwidth density challenge.

References

22.1 J.S. Dietrich: Cosmic cubism, Eng. Sci. **47**(4), 17–20 (1984)

22.2 J. Gao, H. Cheng, H.C. Wu, G. Liu, E. Lau, L. Yuan, C. Krause: Thunderbolt Interconnect – Optical and copper, J. Lightwave Technol. **35**(15), 3125–3129 (2017)

22.3 T. Mikolov, K. Chen, G. Corrado, J. Dean: Efficient estimation of word representations in vector space, arXiv:1301.3781 [cs.CL] (2013)

22.4 D.A. Reed, J. Dongarra: Exascale computing and big data, Commun. ACM **58**(7), 56–68 (2015)

22.5 J.S. Vetter: Contemporary high performance computing. In: *Contemporary High Performance Computing: From Petascale Toward Exascale* (Chapman & Hall, CRC, New York 2013)

22.6 N.R. Adiga, G. Almasi, G.S. Almasi, Y. Aridor, R. Barik, D. Beece, R. Bellofatto, G. Bhanot, R. Bickford, M. Blumrich, A.A. Bright, J. Brunheroto, C. Cascaval, J. Castanos, W. Chan, L. Ceze, P. Coteus, S. Chatterjee, D. Chen, G. Chiu, T.M. Cipolla, P. Crumley, K.M. Desai, A. Deutsch, T. Domany, M.B. Dombrowa, W. Donath, M. Eleftheriou, C. Erway, J. Esch, B. Fitch, J. Gagliano, A. Gara, R. Garg, R. Germain, M.E. Giampapa, B. Gopalsamy, J. Gunnels, M. Gupta, F. Gustavson, S. Hall, R.A. Haring, D. Heidel, P. Heidelberger, L.M. Herger, D. Hoenicke,

R.D. Jackson, T. Jamal-Eddine, G.V. Kopcsay, E. Krevat, M.P. Kurhekar, A.P. Lanzetta, D. Lieber, L.K. Liu, M. Lu, M. Mendell, A. Misra, Y. Moatti, L. Mok, J.E. Moreira, B.J. Nathanson, M. Newton, M. Ohmacht, A. Oliner, V. Pandit, R.B. Pudota, R. Rand, R. Regan, B. Rubin, A. Ruehli, S. Rus, R.K. Sahoo, A. Sanomiya, E. Schenfeld, M. Sharma, E. Shmueli, S. Singh, P. Song, V. Srinivasan, B.D. Steinmacher-Burow, K. Strauss, C. Surovic, R. Swetz, T. Takken, R.B. Tremaine, M. Tsao, A.R. Umamaheshwaran, P. Verma, P. Vranas, T.J.C. Ward, M. Wazlowski, W. Barrett, C. Engel, B. Drehmel, B. Hilgart, D. Hill, F. Kasemkhani, D. Krolak, C.T. Li, T. Liebsch, J. Marcella, A. Muff, A. Okomo, M. Rouse, A. Schram, M. Tubbs, G. Ulsh, C. Wait, J. Wittrup, M. Bae, K. Dockser, L. Kissel, M.K. Seager, J.S. Vetter, K. Yates: An Overview of the BlueGene/L supercomputer. In: *Proc. ACM Conf. Supercomput.* (2002)

22.7 J. Dongarra: Report on the Sunway TaihuLight System, Univ. Tennessee Comput. Sci. Rep. UT-EECS-16-742 (2016)

22.8 R. Dolbeau: Theoretical Peak FLOPS per instruction set on less conventional hardware, http://www.dolbeau.name/dolbeau/publications/peak-alt.pdf (2016)

22.9 TOP500.org: Top 500 supercomputer sites, https://www.top500.org/ (2019)

22.10 The High Performance Conjugate Gradients (HPCG) Benchmark Project: http://www.hpcg-benchmark.org/

22.11 S. Rumley, D. Nikolova, R. Hendry, Q. Li, D. Calhoun, K. Bergman: Silicon photonics for exascale systems, J. Lightwave Technol. **33**(3), 547–562 (2015)

22.12 Graph 500: Large-scale benchmarks, https://graph500.org/ (2017)

22.13 R.C. Murphy, K.B. Wheeler, B.W. Barrett, J.A. Ang: Introducing the graph 500. In: *Proc. CUG* (2010)

22.14 E. Joseph, S. Conway: *Major Trends in the Worldwide HPC Market* (IDC, Framingham 2017)

22.15 S. Rumley, M. Bahadori, R. Polster, S.D. Hammond, D.M. Calhoun, K. Wen, A. Rodrigues, K. Bergman: Optical interconnects for extreme scale computing systems, Parallel Comput. **64**(Suppl. C), 65–80 (2017)

22.16 U.S. Energy Information Administration: Electricity monthly update, https://www.eia.gov/electricity/monthly/update/wholesale\protect_markets.php (2019)

22.17 J.E. Thornton: The CDC 6600 project, Ann. Hist. Comput. **2**(4), 338–348 (1980)

22.18 A. Aiken, U. Banerjee, A. Kejariwal, A. Nicolau: Overview of ILP architectures. In: *Instruction Level Parallelism* (Springer, New York 2016) pp. 9–42

22.19 J.S. Vetter, B.R. de Supinski, L. Kissel, J. May, S. Vaidya: Evaluating high-performance computers, Concurr. Comput. **17**(10), 1239–1270 (2005)

22.20 R. Espasa, M. Valero, J.E. Smith: Vector architectures: Past, present and future. In: *Proc. Int. Conf. Supercomp.* (1998) pp. 425–432

22.21 A.H. Karp, M. Heath, D. Heller, H. Simon: 1994 Gordon Bell Prize winners, Computer **28**(1), 68–74 (1995)

22.22 R.H. Dennard, F.H. Gaensslen, V.L. Rideout, E. Bassous, A.R. LeBlanc: Design of ion-implanted MOSFET's with very small physical dimensions, IEEE J. Solid-State Circuits **9**(5), 256–268 (1974)

22.23 M.M. Waldrop: The chips are down for Moores law, Nature **530**, 144–147 (2016)

22.24 N. Loubet, T. Hook, P. Montanini, C.W. Yeung, S. Kanakasabapathy, M. Guillom, T. Yamashita, J. Zhang, X. Miao, J. Wang, A. Young, R. Chao, M. Kang, Z. Liu, S. Fan, B. Hamieh, S. Sieg, Y. Mignot, W. Xu, S.C. Seo, J. Yoo, S. Mochizuki, M. Sankarapandian, O. Kwon, A. Carr, A. Greene, Y. Park, J. Frougier, R. Galatage, R. Bao, J. Shearer, R. Conti, H. Song, D. Lee, D. Kong, Y. Xu, A. Arceo, Z. Bi, P. Xu, R. Muthinti, J. Li, R. Wong, D. Brown, P. Oldiges, R. Robison, J. Arnold, N. Felix, S. Skordas, J. Gaudiello, T. Standaert, H. Jagannathan, D. Corliss, M.H. Na, A. Knorr, T. Wu, D. Gupta, S. Lian, R. Divakaruni, T. Gow, C. Labelle, S. Lee, V. Paruchuri, H. Bu, M. Khare: Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET. In: *Symp. VLSI Technol.* (2017) pp. T230–T231

22.25 IEEE: IRDS More Moore white paper, https://irds.ieee.org/images/files/pdf/2016\protect_MM.pdf (2016)

22.26 D. Foley, J. Danskin: Ultra-performance Pascal GPU and NVlink interconnect, IEEE Micro **37**(2), 7–17 (2017)

22.27 S. Derradji, T. Palfer-Sollier, J.P. Panziera, A. Poudes, F.W. Atos: The BXI interconnect architecture. In: *IEEE 23rd Proc. Ann. Symp. High Perform. Interconnects* (2015) pp. 18–25

22.28 D. Chen, N.A. Eisley, P. Heidelberger, R.M. Senger, Y. Sugawara, S. Kumar, V. Salapura, D.L. Satterfield, B. Steinmacher-Burow, J.J. Parker: The IBM Blue Gene/Q interconnection network and message unit. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* (2011), Art. 26

22.29 S. Scott, D. Abts, J. Kim, W.J. Dally: The BlackWidow high-radix Clos network, ACM SIGARCH Comput. Archit. News **34**(2), 16–28 (2006)

22.30 G. Faanes, A. Bataineh, D. Roweth, T. Court, E. Froese, B. Alverson, T. Johnson, J. Kopnick, M. Higgins, J. Reinhard: Cray Cascade: A scalable HPC system based on a Dragonfly network. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* (2012), Art. 103

22.31 Cray: Cray® XC50™ supercomputer, https://www.cray.com/sites/default/files/Cray-XC50-NVIDIA-Tesla-P100-GPU-Accelerator-Blade.pdf (2017)

22.32 Mellanox: InfiniBand switch systems, http://www.mellanox.com/page/switch\protect_systems\protect_overview (2019)

22.33 S.H. Russ: Differential signaling. In: *Signal Integrity: Applied Electromagnetics and Professional Practice* (Springer, Berlin 2016) pp. 101–109

22.34 M. Besta, T. Hoefler: Slim Fly: A cost effective low-diameter network topology. In: *Proc. Int. Conf.*

High Perform. Comput. Netw. Storage Anal. (2014) pp. 348–359

22.35 C.A. Thraskias, E.N. Lallas, N. Neumann, L. Schares, B.J. Offrein, R. Henker, D. Plettemeier, F. Ellinger, J. Leuthold, I. Tomkos: Survey of photonic and plasmonic interconnect technologies for intra-datacenter and high-performance computing communications, IEEE Commun. Surv. Tutor. **20**(4), 2758–2783 (2018)

22.36 S. Rumley, R.P. Polster, S.D. Hammond, A.F. Rodrigues, K. Bergman: End-to-end modeling and optimization of power consumption in HPC interconnects. In: *45th Int. Conf. Parallel Process. Workshops* (2016) pp. 133–140, https://doi.org/10.1109/ICPPW.2016.33

22.37 V. Marjanović, J. Gracia, C.W. Glass: Performance modeling of the HPCG benchmark. In: *High Performance Computing Systems. Performance Modeling, Benchmarking, and Simulation. PMBS 2014*, Lecture Notes in Computer Science, Vol. 8966, ed. by S. Jarvis, S. Wright, S. Hammond (Springer, Cham 2015) pp. 172–192

22.38 G. Michelogiannakis, K.Z. Ibrahim, J. Shalf, J.J. Wilke, S. Knight, J.P. Kenny: APHiD: Hierarchical task placement to enable a tapered fat tree topology for lower power and cost in HPC networks. In: *7th IEEE/ACM Int. Symp. Cluster Cloud Grid Comput.* (2017) pp. 228–237

22.39 S. Lammel, F. Zahn, H. Fröning: SONAR: Automated communication characterization for HPC applications. In: *High Performance Computing. ISC High Performance 2016*, Lecture Notes in Computer Science, Vol. 9945, ed. by M. Taufer, B. Mohr, J. Kunkel (Springer, Cham 2016) pp. 98–114

22.40 J. Dean, L.A. Barroso: The tail at scale, Commun. ACM **56**(2), 74–80 (2013)

22.41 N. Jain, A. Bhatele, S. White, T. Gamblin, L.V. Kale: Evaluating HPC networks via simulation of parallel workloads. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* (2016) pp. 154–165

22.42 N. Jain, A. Bhatele, L.H. Howell, D. Böhme, I. Karlin, E.A. León, M. Mubarak, N. Wolfe, T. Gamblin, M.L. Leininger: Predicting the performance impact of different fat-tree configurations. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* (2017), Art. 50

22.43 G. Kathareios, C. Minkenberg, B. Prisacari, G. Rodriguez, T. Hoefler: Cost-effective diameter-two topologies: Analysis and evaluation. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* (2017), Art. 36

22.44 Y. Ajima, Y. Takagi, T. Inoue, S. Hiramoto, T. Shimizu: The tofu interconnect. In: *IEEE 19th Ann. Symp. High Perform. Interconnects* (2011) pp. 87–94

22.45 J. Kim, W.J. Dally, D. Abts: Flattened butterfly: A cost-efficient topology for high-radix networks, ACM SIGARCH Comput. Archit. News **35**(2), 126–137 (2007)

22.46 J.H. Ahn, N. Binkert, A. Davis, M. McLaren, R.S. Schreiber: HyperX: Topology, routing, and packaging of efficient large-scale networks. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* (2009), Art. 41

22.47 S. Rumley, M. Glick, S.D. Hammond, A. Rodrigues, K. Bergman: Design methodology for optimizing optical interconnection networks in high performance systems. In: *High Performance Computing. ISC High Performance 2015*, Lecture Notes in Computer Science, Vol. 9137, ed. by J. Kunkel, T. Ludwig (Springer, Cham 2015) pp. 454–471

22.48 J. Kim, W.J. Dally, S. Scott, D. Abts: Technology-driven, highly-scalable dragonfly topology, ACM SIGARCH Comput. Archit. News **36**(3), 77–88 (2008)

22.49 B. Alverson, E. Froese, L. Kaplan, D. Roweth: Cray XC series network, WP-Aries01-1112 (2012)

22.50 M.Y. Teh, J.J. Wilke, K. Bergman, S. Rumley: Design space exploration of the dragonfly topology. In: *High Performance Computing. ISC High Performance 2017*, Lecture Notes in Computer Science, Vol. 10524, ed. by J. Kunkel, R. Yokota, M. Taufer, J. Shalf (Springer, Cham 2017) pp. 57–74

22.51 P. Gevros, J. Crowcroft, P. Kirstein, S. Bhatti: Congestion control mechanisms and the best effort service model, IEEE Netw. **15**(3), 16–26 (2001)

22.52 V. Jacobson: Congestion avoidance and control, SIGCOMM Comput. Commun. Rev. **18**(4), 314–329 (1988)

22.53 P. Grun: *Introduction to InfiniBand for End Users* (InfiniBand Trade, Beaverton 2010)

22.54 A. Falk, T. Faber, J. Bannister, A. Chien, R. Grossman, J. Leigh: Transport protocols for high performance, Commun. ACM **46**(11), 42–49 (2003)

22.55 W. Dally, B. Towles: *Principles and Practices of Interconnection Networks* (Morgan Kaufmann, New York 2003)

22.56 T.M. Pinkston, J. Duato: Appendix F: Interconnection networks. In: *Computer Architecture: A Quantitative Approach*, 5th edn., ed. by J.L. Hennessy, D.A. Patterson (Morgan Kaufmann, Waltham 2011)

22.57 IEEE Standards Association: Standard group MAC addresses: A tutorial guide, https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/tutorials/macgrp.pdf (2019)

22.58 S.L. Scott, G.M. Thorson: The cray t3e network: Adaptive routing in a high performance 3d torus. In: *Proc. Symp. HOT Interconnects IV* (1996) pp. 147–156

22.59 A. Singh, W.J. Dally, A.K. Gupta, B. Towles: GOAL: A load-balanced adaptive routing algorithm for torus networks. In: *Proc. 30th Ann. Int. Symp. Comput. Archit.* (2003) pp. 194–205

22.60 J. Kim, W.J. Dally, D. Abts: Adaptive routing in high-radix clos network. In: *Proc. ACM/IEEE Conf. Supercomput.* (2006), Art. 92

22.61 N. Jiang, J. Kim, W.J. Dally: Indirect adaptive routing on large scale interconnection networks, ACM SIGARCH Comput. Archit. News **37**(3), 220–231 (2009)

22.62 R. Alverson, D. Roweth, L. Kaplan: The Gemini system interconnect. In: *IEEE 18th Symp. High Perform. Interconnects* (2010) pp. 83–87

22.63 P. Faizian, M.A. Mollah, Z. Tong, X. Yuan, M. Lang: A comparative study of SDN and adaptive routing on

dragonfly networks. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* (2017), Art. 51
22.64 B. Barney: Using ASC Purple, https://computing.llnl.gov/tutorials/purple/index.html (2019)
22.65 S. Benjamin, K. Hasharoni, A. Maman, S. Stepanov, M. Mesh, H. Luesebrink, R. Steffek, W. Pleyer, C. Stömmer: 336-Channel electro-optical interconnect: Underfill process improvement, fiber bundle and reliability results. In: *IEEE 64th Electron. Compon. Technol. Conf.* (2014) pp. 1021–1027
22.66 SWDM Alliance: http://www.swdm.org/
22.67 T. Saeki, S. Sato, M. Kurokawa, A. Moto, M. Suzuki, K. Tanaka, K. Tanaka, N. Ikoma, Y. Fujimura: 100 Gbit/s compact transmitter module integrated with optical multiplexer. In: *IEEE Photonics Conf.* (2013) pp. 307–308
22.68 T.N. Huynh, F. Doany, D.M. Kuchta, D. Gazula, E. Shaw, J. O'Daniel, J. Tatum: 4×50Gb/s NRZ shortwave-wavelength division multiplexing VCSEL link over 50m multimode fiber. In: *Opt. Fiber Commun. Conf. Exhib.* (2017), https://doi.org/10.1364/OFC.2017.Tu2B.5
22.69 M.R.T. Tan, P. Rosenberg, W.V. Sorin, B. Wang, S. Mathai, G. Panotopoulos, G. Rankin: Universal photonic interconnect for data centers, J. Lightwave Technol. **36**(2), 175–180 (2018)
22.70 L. Carroll, J.-S. Lee, C. Scarcella, K. Gradkowski, M. Duperron, H. Lu, Y. Zhao, C. Eason, P. Morrissey, M. Rensing, S. Collins, H.Y. Hwang, P. O'Brien: Photonic packaging: Transforming silicon photonic integrated circuits into photonic devices, Appl. Sci. **6**(12), 426 (2016)
22.71 N. Kohmu, T. Takai, N. Chujo, H. Arimoto: A 25.78-Gbit/s × 4-ch active optical cable with ultra-compact form factor for high-density optical interconnects, Appl. Sci. **8**(1), 137 (2018)
22.72 Y. Arakawa, T. Nakamura, Y. Urino, T. Fujita: Silicon photonics for next generation system integration platform, IEEE Commun. Mag. **51**(3), 72–77 (2013)
22.73 P. Dumon: Towards foundry approach for silicon photonics: Silicon photonics platform ePIXfab, Electron. Lett. **45**(1), 581–582 (2009)
22.74 M.T. Wade, J.M. Shainline, J.S. Orcutt, C. Sun, R. Kumar, B. Moss, M. Georgas, R.J. Ram, V. Stojanović, M.A. Popović: Energy-efficient active photonics in a zero-change, state-of-the-art CMOS process. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.Tu2E.7
22.75 L. Liao, A. Liu, D. Rubin, J. Basak, Y. Chetrit, H. Nguyen, R. Cohen, N. Izhaky, M. Paniccia: 40 Gbit/s silicon optical modulator for highspeed applications, Electron. Lett. **43**(22), 1196 (2007)
22.76 Y. Tang, H.-W. Chen, S. Jain, J.D. Peters, U. Westergren, J.E. Bowers: 50 Gb/s hybrid silicon traveling-wave electroabsorption modulator, Opt. Express **19**(7), 5811–5816 (2011)
22.77 Q. Xu, B. Schmidt, S. Pradhan, M. Lipson: Micrometre-scale silicon electro-optic modulator, Nature **435**(7040), 325–327 (2005)
22.78 Y. Liu, R. Ding, Q. Li, Z. Xuan, Y. Li, Y. Yang, A.E. Lim, P.G. Lo, K. Bergman, T. Baehr-Jones, M. Hochberg: Ultra-compact 320 Gb/s and 160 Gb/s WDM transmitters based on silicon microrings. In: *Opt. Fiber Commun. Conf.* (2014), https://doi.org/10.1364/OFC.2014.Th4G.6
22.79 M. Bahadori, S. Rumley, D. Nikolova, K. Bergman: Comprehensive design space exploration of silicon photonic interconnects, J. Lightwave Technol. **34**(12), 2975–2987 (2016)
22.80 M. Bahadori, S. Rumley, R. Polster, A. Gazman, M. Traverso, M. Webster, K. Patel, K. Bergman: Energy-performance optimized design of silicon photonic interconnection networks for high-performance computing. In: *Design Autom. Test Eur. Conf. Exhib.* (2017) pp. 326–331
22.81 A.V. Krishnamoorthy, O. Torudbakken, S. Müller, A. Srinivasan, P. Decker, H. Opheim, J.E. Cunningham, X. Zheng, M. Dignum, K. Raj: From chip to cloud: Optical interconnects in engineered systems for the enterprise. In: *IEEE Opt. Interconnects Conf.* (2016) pp. 34–35
22.82 K. Padmaraju, D.F. Logan, T. Shiraishi, J.J. Ackert, A.P. Knights, K. Bergman: Wavelength locking and thermally stabilizing microring resonators using dithering signals, J. Lightwave Technol. **32**(3), 505–512 (2014)
22.83 M. Piels, J.E. Bowers: 40 GHz Si/Ge uni-traveling carrier waveguide photodiode, J. Lightwave Technol. **32**(20), 3502–3508 (2014)
22.84 M.L. Brongersma: Plasmonic photodetectors, photovoltaics, and hot-electron devices, Proc. IEEE Inst. Electr. Electron Eng. **104**(12), 2349–2361 (2016)
22.85 D.A.B. Miller: Attojoule optoelectronics for low-energy information processing and communications, J. Lightwave Technol. **35**(3), 346–396 (2017)
22.86 S.A. Maier: *Plasmonics: Fundamentals and Applications* (Springer, New York 2007)
22.87 D.K. Mynbaev, V. Sukharenko: Plasmonics for optical communications: The use of graphene for optimizing coupling efficiency. In: *Int. Caribb. Conf. Devices Circuits Syst.* (2014), https://doi.org/10.1109/ICCDCS.2014.7016180
22.88 M.L. Brongersma, V.M. Shalaev: The case for plasmonics, Science **328**(5977), 440–441 (2010)
22.89 M. Ayata, Y. Fedoryshyn, W. Heni, B. Baeuerle, A. Josten, M. Zahner, U. Koch, Y. Salamin, C. Hoessbacher, C. Haffner, D.L. Elder, L.R. Dalton, J. Leuthold: High-speed plasmonic modulator in a single metal layer, Science **358**(6363), 630–632 (2017)
22.90 D. Dai, Y. Shi, S. He, L. Wosinski, L. Thylen: Gain enhancement in a hybrid plasmonic nanowaveguide with a low-index or high-index gain medium, Opt. Express **19**(14), 12925–12936 (2011)
22.91 S. Kamil, L. Oliker, A. Pinar, J. Shalf: Communication requirements and interconnect optimization for high-end scientific applications, IEEE Trans. Parallel Distrib. Syst. **21**(2), 188–202 (2010)
22.92 K. Wen, P. Samadi, S. Rumley, C.P. Chen, Y. Shen, M. Bahadroi, K. Bergman, J. Wilke: Flexfly: Enabling a reconfigurable dragonfly through silicon photonics. In: *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.* (2016) pp. 166–177

22.93 C. Minkenberg, G. Rodriguez, B. Prisacari, L. Schares, P. Heidelberger, D. Chen, C. Stunkel: Performance benefits of optical circuit switches for large-scale dragonfly networks. In: *Opt. Fiber Commun. Conf.* (2016), https://doi.org/10.1364/OFC.2016.W3J.3

22.94 Q. Cheng, S. Rumley, M. Bahadori, K. Bergman: Optical technologies for data centers: A review, Optica **5**(11), 1354 (2018)

22.95 I. Fujiwara, M. Koibuchi, T. Ozaki, H. Matsutani, H. Casanova: Augmenting low-latency HPC network with free-space optical links. In: *IEEE 21st Int. Symp. High Perform. Comput. Archit.* (2015) pp. 390–401

22.96 Y. Hu, I. Fujiwara, M. Koibuchi: Job mapping and scheduling on free-space optical networks, IEICE Trans. Inform. Syst. **E99.D**(11), 2694–2704 (2016)

22.97 J.M. Kahn, D.A.B. Miller: Communications expands its space, Nat. Photonics **11**(1), 5–8 (2017)

Sébastien Rumley

Dept. of Electrical Engineering
Columbia University
New York, NY, USA
sebastien.rumley@olympe.ch

Sébastien Rumley received his MS and PhD degrees in Communication Systems from EPFL, Switzerland, in 2005 and 2011. He is currently Research Scientist with the Lightwave Research Laboratory, Columbia University, New York. His research interests include cross-scale modeling and optimization of large-scale interconnection networks, with particular focus on optical technologies and silicon photonics.

Keren Bergman

Dept. of Electrical Engineering
Columbia University
New York, NY, USA
bergman@ee.columbia.edu

Keren Bergman received a BS from Bucknell University in 1988 an MS in 1991, and a PhD in 1994 from MIT, all in Electrical Engineering. Bergman leads the Lightwave Research Laboratory encompassing multiple cross-disciplinary programs at the intersection of computing and photonics. Bergman serves on the Leadership Council (AIM) Photonics. She is a Fellow of the Optical Society of America (OSA) and IEEE.

M. Ashkan Seyedi

Hewlett Packard Labs
HPE
Palo Alto, CA, USA
ashkan.seyedi@hpe.com

Ashkan Seyedi received a dual Bachelor's in Electrical and Computer Engineering from the University of Missouri-Columbia and a PhD from the University of Southern California working on photonic crystal devices, high-speed nanowire photodetectors, efficient white LEDs, and solar cells. While at Hewlett Packard Enterprise as a research scientist, he has been working on developing high-bandwidth, efficient optical interconnects for exascale and high-performance computing applications.

Marco Fiorentino

Hewlett Packard Labs
HPE
Palo Alto, CA, USA
marco.fiorentino@hpe.com

Marco Fiorentino is a Distinguished Technologist at Hewlett Packard Labs. Before joining HPE in 2005 he worked at Northwestern University, the University of Rome, and MIT. Marco received a PhD in Physics from the University of Napoli, Italy in 2000, working on quantum optics. He is currently working on silicon photonics technologies for computercom. He is a Senior Member of the Optical Society of America.

23. Intra-Datacenter Network Architectures

Roberto Proietti, Pouya Fotouhi, Sebastian Werner, S.J. Ben Yoo

This Handbook chapter provides an overview of existing interconnection topologies and architectures for large-scale cloud computing systems. We discuss (a) tree-based solutions (also referred to as indirect topologies in this chapter), largely adopted in most of today's data center systems, as well as (b) directly connected topologies, such as full mesh (all-to-all), flattened butterfly, and HyperX. This chapter also touches upon some emerging interconnect solutions enabled by recent advances in photonic integrated technologies and switches, while more details about these aspects can be found in the Handbook Chaps. 21, 24, and 25.

Data centers have become the center of how the world functions today. While billions of consumers rely on the availability of data centers in their daily lives for services such as navigation, personal banking, search engines, or social networking, commercial businesses exploit the computing and storage capabilities of data centers to increase productivity through data analytics and statistics. To provide satisfying and reliable services to all users and their application domains, data centers must offer high performance, cost efficiency, sustainability, and scalability. In particular, energy efficiency significantly affects operating costs, sustainability, and scalability. Hence, the cost-effective design of a data center must consider both operating expenses and capital expenses.

23.1 Trends and Future Challenges of Cloud Data Centers

With IPv6, data centers can now address every appliance and sensor on earth and exoplanet. A vast set of data from everywhere will be networked, processed, and accessed on virtual platforms (commonly referred to as cloud architectures) that consist of data systems, networks (optical, electrical/wireless and wireline), and client interfaces (e.g., terminals or hand-held devices). At the physical layer, these virtual platforms run in data centers, which are essentially a collection of internetworked servers designed to store, access, and process data for the clients—a system architecture aptly termed warehouse-scale computing systems. In fact, the explosive growth of data that needs to be stored, accessed, and processed, has led to the current trend of turning warehouse-scale computing systems into increasingly larger and deeply networked hyper-scale data centers.

Such system scales have created three main problems that, if not solved, will prevent data centers from supporting the projected data growth in the future. Firstly, the power consumption of data centers limits their scalability. Secondly, the intradata center interconnection limits its performance. Thirdly, the interdata center interconnection limits the performance and utilization of the cloud. In particular, the energy efficiency of the cyberinfrastructure links all three issues together. Readers should also refer to Chap. 21 *Evolving require-*

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_23

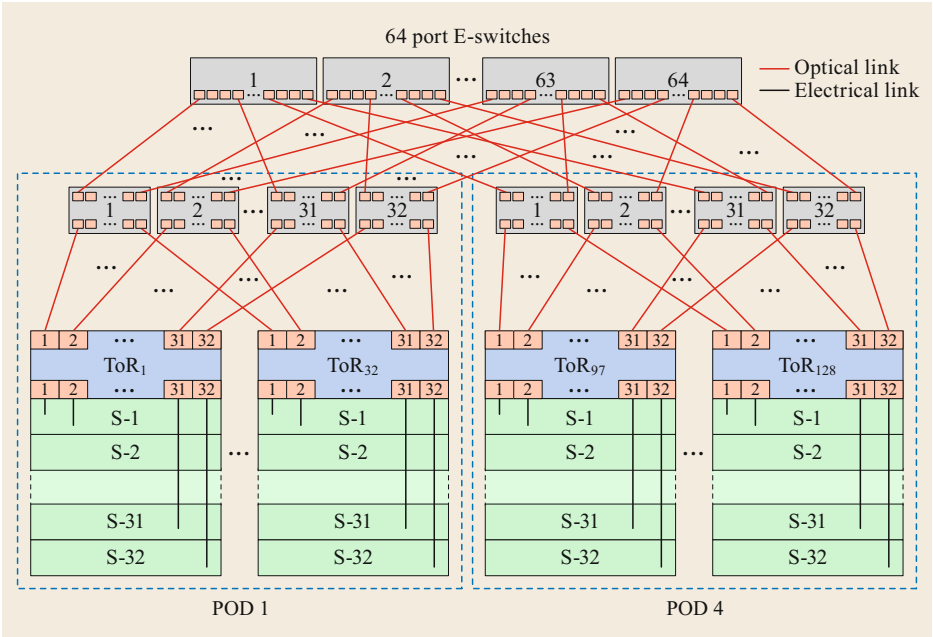


Fig. 23.1 128 rack data center using electrical switches connected by optical links in a spine-leaf or tree architecture (after [23.1])

ments and trends of data centers networks for more details.

Today's data centers already consume megawatts of power and require a large power distribution and cooling infrastructure. Global data center IP traffic is expected to grow threefold over the next 5 years, at a compound annual growth rate (CAGR) of 25% from 2016 to 2021. At the same time, the energy consumption in US data centers reached 91 TWh in 2013 and is expected to increase at a rate that doubles about every 8 years [23.2]. The exponential trend of data growth has led to the emergence of optical communication between racks in data centers to overcome the bandwidth and energy limitations of electrical interconnects. While this transition from electrical interconnection to optical interconnection has greatly increased communication capacity and improved energy efficiency, the energy efficiency is still not good enough to support scalability for future data centers for several reasons, which we discuss below.

First, as illustrated in Fig. 23.1, typical data centers deploy networks with a very large number of power-hungry electronic switches cascaded in numerous stages according to a spine-leaf tree topology with PODs (portable data centers). Due to limitations in radix (port count) and bandwidth of the electronic switches, the inefficiency of the cascaded switch stages accumulates, especially in terms of latency, throughput, and power consumption.

Modular data centers have become increasingly popular among web service providers. Microsoft, Google, and Facebook [23.3] constructed data centers based on self-contained shipping containers (often called a POD) with up to thousands of servers, network infrastructure, and cooling system attached. Providing a nonblocking switch to connect all servers at the POD level is feasible as the number of servers scales up to thousands. However, at the whole data center scale with up to thousands of PODs, achieving energy efficiency and high performance for the network is nontrivial.

Second, although the demand for high-capacity communications brought optics to data centers, today's solutions based on pluggable optical modules offer only partial savings in the communication chain. Historically, integrated circuits and systems have improved by collapsing functions into a single integrated circuit and eliminating interfaces. For instance, embedded solutions proposed by COBO (Consortium for On-Board Optics) [23.4] fail to eliminate any intermediate electronic interfaces such as equalizers and serializers/deserializers (SERDES). Readers should also refer to Chap. 24 *System aspects for optical interconnect transceivers* for more details.

The transmission distances and bandwidth on electrical wires without repeaters are severely limited due to losses (skin effects or bulk resistivity) and distortion imposed on the signals by the impedance of the electrical wires [23.6]. According to *Miller* and *Ozaktas* [23.6], the transmission distance limit is

$$\text{lim} = \sqrt{\frac{B_0}{B}\frac{1}{A}}\,,$$

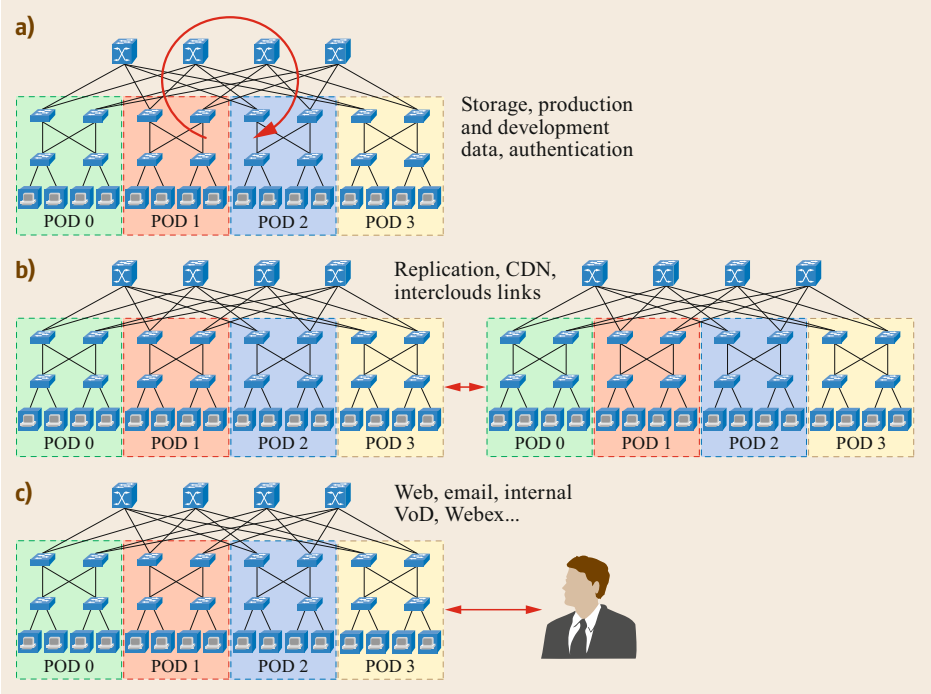


Fig. 23.2a–c Global data center traffic by destination in 2020 (Source: Cisco Global Index, 2015–2020; Sinergy Research, after [23.5]). **(a)** Intra data center (77%), **(b)** inter data center (9%), **(c)** data center to users (14%)

where A is the cross-sectional area of the electrical wire, B the line rate, and $B_0 = 10^{15}$ (inductive capacitive (LC) lines) to 10^{16} bit/s (resistive capacitive (RC) lines). This indicates a < 1 cm transmission limit at 25 Gb/s line rates for typical modern on-chip electrical interconnects. On the other hand, optical interconnects are free of such impedance effects and become advantageous over their electrical counterparts beyond a certain distance at a given line rate. *Naeemi* et al. [23.7] defined this distance as a partition length and *Beausoleil* et al. [23.8] calculated these lengths according to ITRS (International Technology Roadmap for Semiconductors) [23.9], where they found the partition length for a wire or waveguide width of 1 μm to be less than 2 mm. Readers should also refer to the survey paper by *Thraskias* et al. [23.10] for a comprehensive overview and comparison of electrical and photonic interconnects.

Third, today's cloud computing architectures are designed with a fixed topology with fixed patterns of data movements at fixed data rates, while actual computations have large peak-to-average ratios in processing, bursty data traffic, dynamically changing data movement patterns, and heterogeneous processing threads.

To achieve low-service latency within modular data centers, inter-POD network requirements in terms of bandwidth and latency should not be ignored. A search engine, for instance, requires hundreds of servers for computation. These servers are not necessarily located within a single POD and most often are distributed among multiple PODs. Moreover, there are periodic traffic patterns due to back-up operations on virtual machines, as well as large-scale database management operations. Different groups of nodes may have different traffic characteristics. Thus, dynamically providing higher bandwidth for strongly coupled nodes within data centers is desirable, yet realizing such a dynamic bandwidth provisioning in the data center is by no means trivial.

Unlike telecommunications where typically ≈ 80% traffic bypasses and ≈ 20% traffic add/drops locally, data traffic in computing systems is ≈ 80% internal and ≈ 20% external, with traffic bandwidth amounts in the order of Pb/s. The statistics from *Cisco* [23.5] shown in Fig. 23.2 support this argument by showing that 77% of the traffic is internal and 23% is external. Therefore, in order to sustain this ever-increasing traffic inside data centers, the symbiotic integration of photonics and electronics must happen at every level—between racks, boards, cards, chips, and memories. The reader can also refer to the following references that provide interesting insights for the traffic characteristics in data centers [23.11–13].

Certainly, the topologies and architectures of future data centers that make heavy use of photonics technologies should be revisited to take into account the benefits of optical interconnects discussed above, while, at the same time, account for the limitation of switching func-

tionalities in the optical domain (for more details, see Sect. 23.3.2 and [23.14–29]).

While many elements and aspects define a particular data center architecture (interconnect topology, network and link protocols, routing schemes [23.30–32], physical layer implementations), this chapter focuses on presenting an overview of existing topologies as well as emerging architectural solutions enabled by photonic technologies.

The remainder of this chapter is organized as follows. Section 23.2 covers the legacy interconnect topologies widely adopted in many deployed data centers as well as other topologies more suitable for high performance computing (HPC) systems or demonstrated at a research level. Section 23.3 provides an overview of existing and emerging photonic technologies enabling new possible architectural solutions. Section 23.4 concludes the chapter.

23.2 Data Center Topologies

A network topology defines the arrangement of the various elements (links, nodes, etc.) of a computer network. Essentially, it is the topological structure of a network and may be depicted physically or logically. Several recent surveys provide thorough investigations of data center network topologies, each with a different focus [23.33–35]. This section focuses on a representative subset of the wide range of topologies that have been proposed for data centers to date. The notations used in this section are as follows, k is the number of ports in a switch, m the number of servers connected to a ToR, n the number of ports in a server, and S the total number of servers within a data center network.

Network topologies can be classified in different ways. One classification divides topologies into tree-based and non-tree-based topologies. *Liu* et al. [23.35] classify topologies into two categories: fixed and flexible. A topology that can be modified after being deployed is referred to as a flexible topology, and fixed otherwise. *Wu* et al. [23.34] present a classification based on the connections between end point servers. A switch-based network, also referred to as an indirect network, consists of multiple levels of switches to connect the end servers. On the other hand, a direct network connects servers to each other without any switch, routers, or network devices. A hybrid network employs a combination of these two methods. This chapter divides topologies into two categories: direct and indirect. Our definition of a direct or indirect network is different from that of *Wu* et al. [23.34]. An indirect topology is defined as a topology with some switches connecting only other switches (and not compute nodes) to each other. We define a direct topology as a topology where all the switches are connected to at least one compute node. In other words, there are no intermediate levels of switches in the topology. Please note that a compute node here could refer to either a single server or a rack of servers. In the latter case, it is represented by a top-of-the-rack (ToR) switch.

23.2.1 Indirect Topologies

Tree

Although often considered simple topologies, tree topologies are broadly used in data center networks. Tree topologies may have either two or three levels. In a two-level tree, switches at the core level connect the switches at the lower level, which are referred to as edges. Servers are located at the leaves and are connected to the switches at the edge level. In a three-level tree, an intermediate level called aggregation is added between the core and edge levels. Figure 23.3 illustrates a 4-ary three-level tree topology with 16 servers, 8 edge switches, 4 aggregation switches, and a single core switch.

Note that in tree topologies, a switch can be connected only to the parent switch and its children. In other words, switches in the same tier or in nonneighboring tiers are not connected to each other.

Tree topologies are intuitive and simple but entail a number of drawbacks. First, the number of servers a tree topology can accommodate is bounded by the number of ports in the switches. Second, performance and reliability of switches should increase as we go higher in the topology levels, which would require different switches for different levels and, in turn, a move away from using commodity switches for such systems.

Fat Tree

As an extension to tree topologies, fat trees are widely used in commercial data center networks as well as research prototype interconnections [23.36, 37]. Fat trees were originally proposed by *Leiserson* in 1985 [23.38]. While there have been several attempts to formalize and use fat trees since then, we found the method used by *Al-Fares* et al. [23.39] to be the most intuitive and comprehensive.

Similar to three-level tree topologies, fat trees also consist of three levels: core, aggregation, and edges. *Al-*

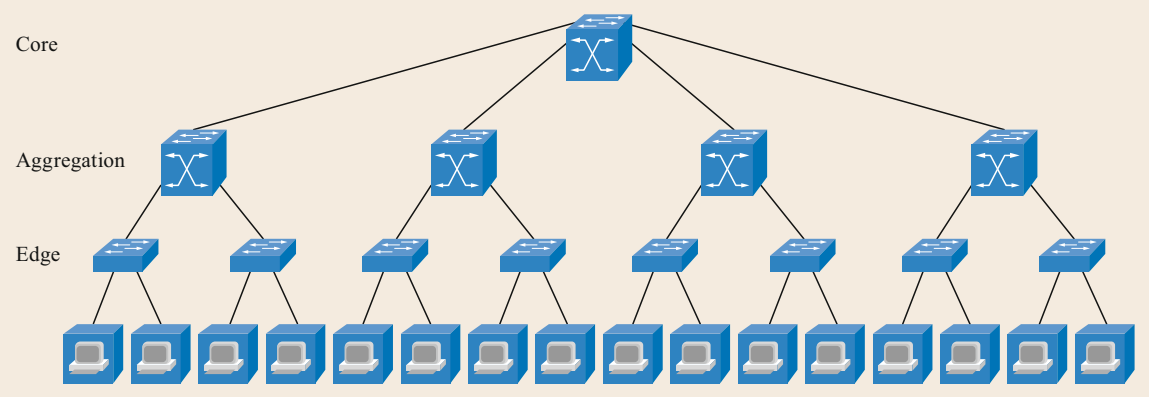


Fig. 23.3 Illustration of a 4-ary tree topology

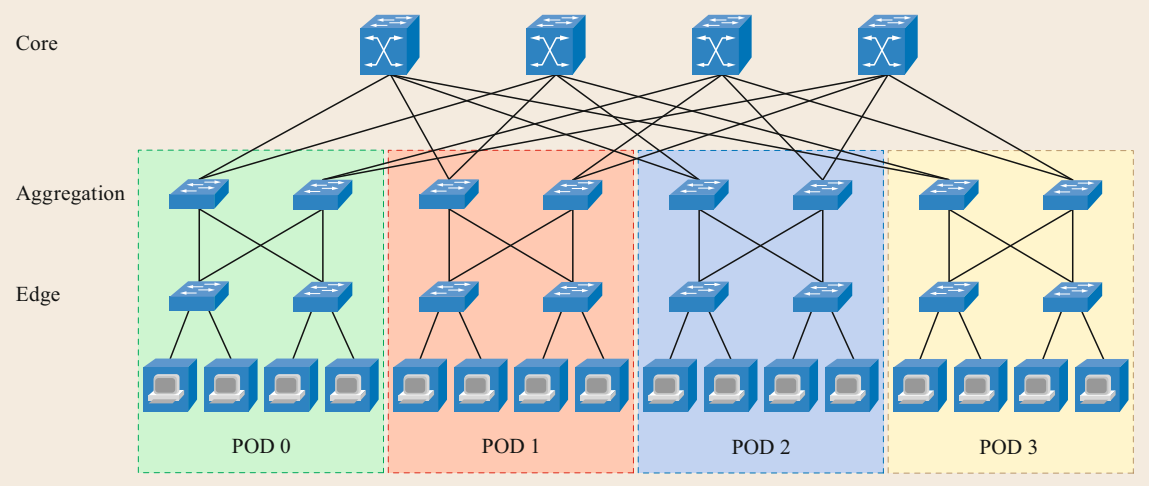


Fig. 23.4 Illustration of a 4-ary fat tree topology

Fares et al. [23.39] adds the POD to the terminology, which consists of aggregation switches, edge switches, and servers. A k-ary fat tree contains $(k/2)^2$ k-port core switches and k PODs. Each core switch is connected to the aggregation layer of each POD through one port (the i-th port of any core switch is connected to POD i). Each POD has two layers of $k/2$ k-port switches. Switches in the upper layer (i.e., aggregation layer) are connected to core switches using $k/2$ of their ports. The remaining $k/2$ ports are used to connect to $k/2$ switches at the lower layer (i.e., edge layer).

$$\left(\frac{k}{2}\right)_{\text{agg. switches/POD}} \times \left(\frac{k}{2}\right)_{\text{ports/switch}} = \left(\frac{k^2}{4}\right)_{\text{core switches}} .$$

Each k-port switch at the edge layer uses $k/2$ of its ports to connect to the aggregation layer and the remaining ports to directly connect to $k/2$ hosts. Therefore, a k-ary fat-tree provides support for a total of $k^3/4$ hosts using k port switches. Figure 23.4 shows a 4-ary fat-tree topology as described before

$$k_{\text{PODs}} \times \left(\frac{k}{2}\right)_{\text{edge switches/POD}} \times \left(\frac{k}{2}\right)_{\text{servers/switch}} = \left(\frac{k^3}{4}\right)_{\text{servers}} .$$

Unlike tree topologies, fat trees can use commodity off-the-shelf switches at all levels. Thus, fat trees offer more flexibility and can be more cost effective compared to tree topologies.

23.2.2 Directly Connected Topologies

This section describes topologies and architectures which are directly connected, meaning that each ToR

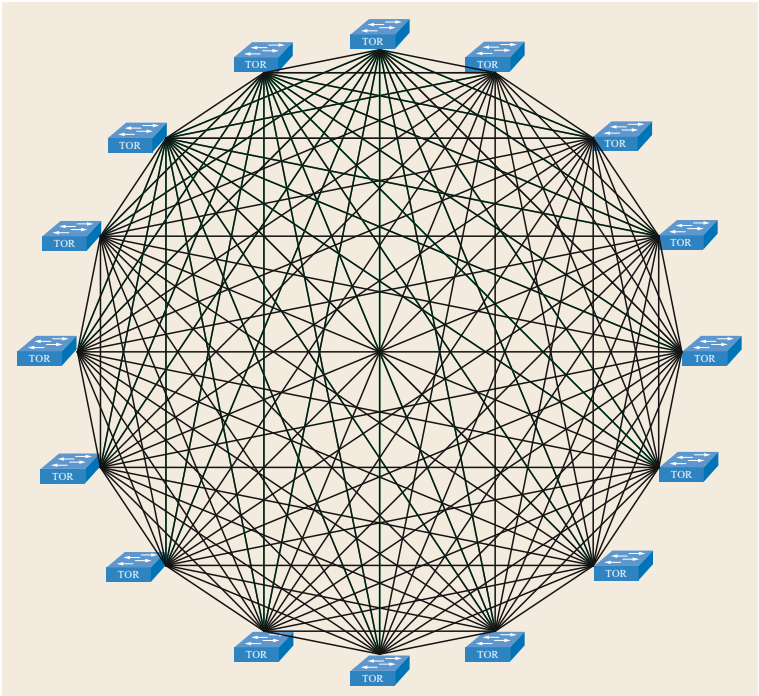


Fig. 23.5 All-to-all interconnection

switch always has a group of ports connecting to a group of nodes.

All-to-All

The all-to-all architecture (also known as full-mesh topology) is a fully connected topology, where each element (ToR in this case) connects to all the other elements in the network (Fig. 23.5a). All-to-all is the simplest topology to realize and offers the lowest diameter (minimum number of hops); however, its inherently limited scalability represents a significant drawback. In fact, the number of ToRs that can be interconnected is restricted by the number of ToR ports ($k-m$, according to the notation used) available for rack-to-rack interconnection. Another drawback is cabling, which requires a very high number of cables that grows with $2\times[(k-m)-1]^2$. Luckily, this problem can be alleviated by using the unique properties of wavelength routing in arrayed waveguide grating routers (AWGRs) [23.40], as is described later in Sect. 23.3.2 and illustrated in Fig. 23.16.

HyperX

The HyperX topology, introduced for the first time by *Ahn* et al. [23.41], can be derived from the all-to-all topology by dividing the nodes into groups (or dimensions) and by removing some intergroup links. Within a group, all nodes are fully connected. Between the groups, each node connects to all the nodes in the other groups with the same positions or coordinates. Figures 23.6 and 23.7 show two examples of HyperX topologies in two dimensions. This architecture allows us to increase the scalability of the system at the expense of a larger diameter and a higher average hop count compared to the all-to-all. In the examples of Figs. 23.6 and 23.7, the ToR switch has four ports for inter-rack communication. This would allow us to interconnect only five racks in a full-mesh topology. By using HyperX, up to eight or nine racks can be interconnected at the expense of increasing the network diameter.

Flattened Butterfly Topology

Another popular topology is the flattened butterfly (FB) topology [23.42, 43], typically represented in rows and columns, as shown in Fig. 23.8. Each node (or ToR) is fully connected with all the nodes in the same row and column. In practice, we can consider the rows and columns as two different dimensions of a HyperX.

Torus

The torus topology [23.43, 44] can also be derived from the HyperX topology when assuming that each dimension is not fully connected but simply interconnected based on a ring topology. A 2-D torus (two-dimensional torus) can then be obtained by taking the flattened but-

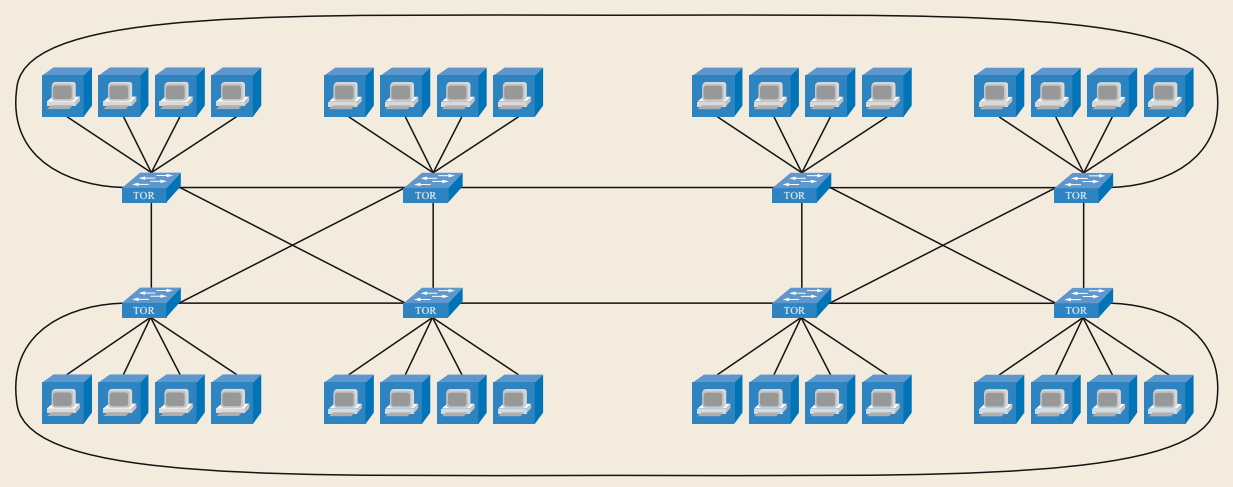


Fig. 23.6 HyperX. CONF (configuration) A topology

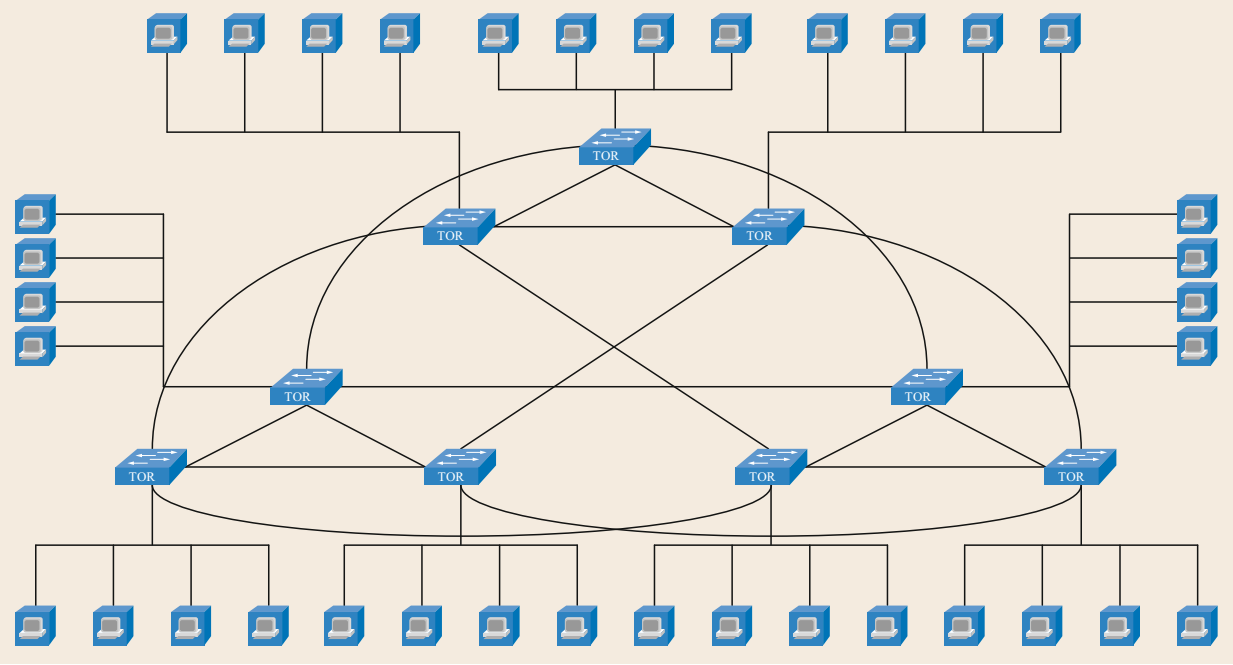


Fig. 23.7 HyperX. CONF B topology

terfly topology described before and keeping only the links interconnecting with the neighboring nodes (see Fig. 23.9). Additional dimensions can be added to obtain a multidimensional torus (e.g., 3-D torus as shown in Fig. 23.10). As we can see, in each coordinate (x, y, or z) the nodes are interconnected in a ring. For this reason, the torus is certainly the most scalable topology, but it should be noted that the diameter and average number of hops in the network can become high even at small scales.

DCell and BCube

DCell [23.45] consists of a basic element DCell_0, which connects k servers to one k-port switch. A R-level DCell (DCell_R) is built up recursively as follows: DCell_1 comprises $k+1$ DCell_0, where each of the k servers of every DCell_0 connects to a server in another DCell_0. Consequently, full connectivity is provided within DCell_1, with exactly one link between each DCell_0 pair. Figure 23.11 exemplifies a DCell_1 consisting of 5 DCell_0 elements with four servers each. Continuing this con-

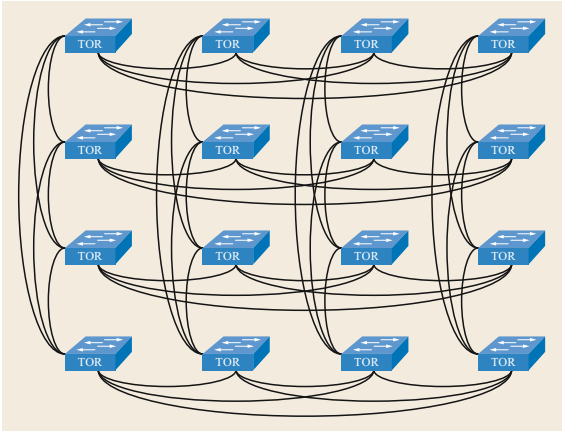


Fig. 23.8 Flattened butterfly topology

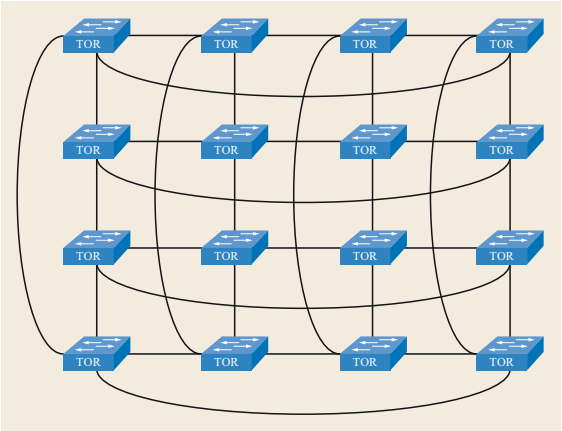


Fig. 23.9 2-D-torus topology

nectivity pattern up to a level R will result in DCell_R—constructed out of multiple DCell_{R-1}—in which each server will have $R+1$ links: one link to connect to the switch at level 0 and one link on each level i to connect to a server in the same DCell_i (but different DCell_{i-1}).

The number of servers in DCell thus has a double-exponential scalability: for instance, for t_{R-1} servers in each DCell_{R-1}, a DCell_R will consist of t_R DCell_{R-1} and, in turn, $t_{R-1} \times t_R$ servers. The number of servers on level R is $t_R = t_{R-1} \times (t_{R-1}+1)$. The number of NICs (network interface card) (k) available on each server limits the number of levels of a DCell. One key benefit of DCell is its ability to scale to a very large number of servers with only very few ports required in the servers and switches. For instance, it has been reported that for $k = 6$ and $R = 3$, DCell can accommodate more than 3 million servers [23.45].

BCube [23.46] is a topology specifically designed to facilitate the shipping and portability of container-based modular data centers. Its construction begins similarly to DCell in the sense that a basic element BCube_0 connects k servers to one k-port switch; however, its scaling is different: to construct BCube_1, k switches are added and connected to one server in each BCube_0. Therefore, BCube_1 consists of k BCube_0 and requires in total $2k$ switches, or more generally: A BCube_k consists of k BCube_{R-1} and k^{R} additional k-port switches, each of which are connected to one server in each BCube_{R-1}.

Both BCube and DCell require servers to have $R+1$ NICs; however, the additional switches per level in

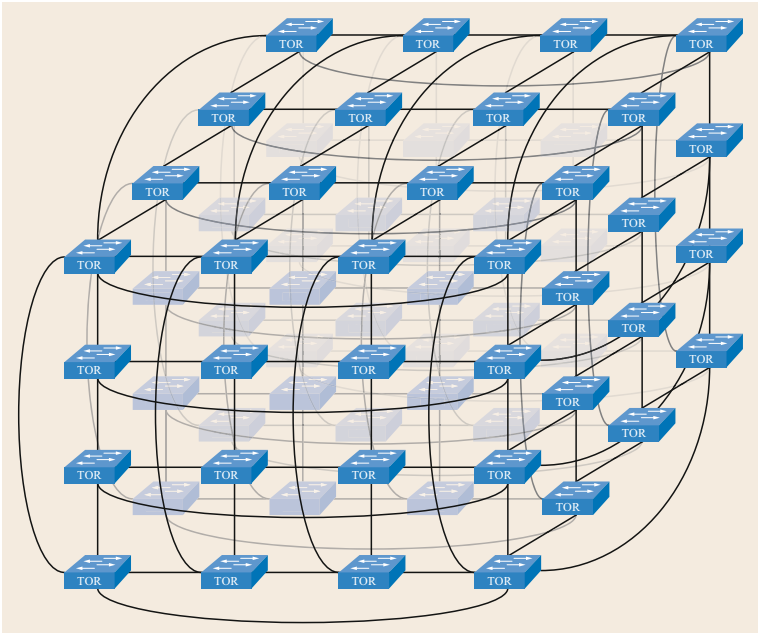


Fig. 23.10 3-D-torus topology

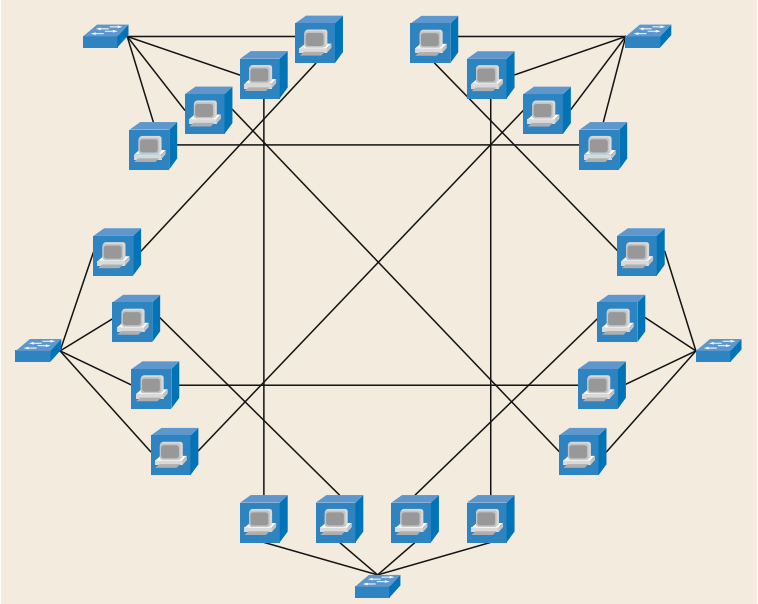


Fig. 23.11 DCell topology

BCube lead to less efficient scalability (although it is sufficient for container-based data centers [23.46]).

23.2.3 Comparison of Topologies

To illustrate the impact of topologies on data center networks, this section compares a subset of the topologies discussed earlier.

It should be noted that among the subset of topologies considered for the comparison, the nodes in tree and fat tree topologies are assumed to have a single-port network interface, while the other four topologies require nodes with multiple ports embedded. Also, it should be noted that the all-to-all and torus topologies considered for the comparison are different from the illustrations presented in the sections above. For this comparison section, these topologies are assumed to connect compute nodes directly, and not ToR switches as discussed before. The following measures are taken into consideration to compare different topologies (note that the comparison method used in this section uses similar terminology to that of *Liu* et al. [23.35]).

Total Number of Servers

In order to compare the scalability among topologies, the maximum number of servers one topology can accommodate (S) is calculated. Our analyses were made under the same number of ports in each switch (k) and the same number of ports in a server (n), if applicable, for a given topology. For all-to-all, the number of nodes connected to each ToR switch is represented by m; D represents the dimension of a torus topology. For the tree topology, we assumed a three-level topology. Parameter R represents the recursion level in a recursive topology (i.e., DCell and BCube).

Total Number of Switches

The number of switches one topology requires provides an estimation of the cost of the network to be deployed. Unless otherwise stated, we assume that all switches used in every topology are identical.

Diameter of the Topology

To elaborate the efficiency of routing for a given topology, we considered the diameter of the network. The diameter can be defined as the longest of the shortest paths between two nodes in a network. The transmission latency is highly correlated with the diameter of the topology.

Bandwidth

As a measure of performance, the bandwidth (in the number of links) provided by each topology is presented in Table 23.1. Our analyses are based on two scenarios. One-to-one shows the maximum bandwidth offered by each topology in a scenario where one arbitrary node communicates with another node in the network. The second scenario, labeled as all-to-all, shows the maximum bandwidth each topology offers when every node actively communicates with all other end points in the network.

Bisection Width

This parameter is taken into consideration as a measure of fault resilience for each topology. It is defined as the

Table 23.1 Comparison results

	Tree	Fat tree	All-to-all	2-D torus	DCell	BCube
Total servers	$(k-1)^3$	$\frac{k^3}{4}$	$(k-m+1)\times m$	$(k-4)\times D^2$	$\left(k+\frac{1}{2}\right)^{2^R}-\frac{1}{2}\le(k+1)^{2^R}-1$	k^{R+1}
Total switches	$\frac{k^2+k+1\times S}{k^3}$	$\frac{5S}{k}$	$\frac{S}{m}$	D^2	$\frac{S}{k}$	$\frac{S}{k}\times\log_k S$
Diameter	$2\log_{k-1} S$	6	1	$2\times\left\lfloor\frac{D}{2}\right\rfloor$	$2^{R+1}-1$	$\log_k S$
One-to-one	1	1	$k-m$	4	$R+1$	$R+1$
All-to-all	k	S	$\frac{(k-m)\times(k-m+1)}{2}$	$2S$	$>\frac{S}{2^R}$	$\frac{k\times(S-1)}{k-1}$
Bisection width	$\frac{k}{2}$	$\frac{S}{2}$	$\frac{(k-m)\times S}{4\times m}$	$2-D$	$>\frac{S}{4\log_k S}$	$\frac{S}{2}$

minimum number of links to be removed in order to divide the network into two equal subsets. The larger the bisection width, the higher failure rate the topology can tolerate.

Table 23.1 shows a comparison of the parameters discussed above. For the total number of servers, DCell provides higher scalability as the number of servers this topology can accommodate grows double-exponentially with the number of ports in each server (n). However, it should be noted that increasing the number of ports in a server will eventually degrade performance, since the bandwidth provided by a given node is limited. Tree topology suffers from poor hardware redundancy, and BCube is expected to have a higher cost for the network, as it uses more switches. All-to-all, torus, and fat tree offer a small constant diameter, outperforming the exponentially growing (with n) diameter in DCell as well as the logarithmically growing diameter in BCube and tree topologies. Among the topologies considered for comparison, as one may expect, all-to-all offer highest all-to-all bandwidth, as all the links would be simultaneously used to transfer the data. Fat tree topology also provides considerable all-to-all bandwidth using redundant switches in the aggregation and edge levels. Tree-based topology provides the lowest all-to-all bandwidth among the topologies under study, since the bandwidth is limited by the number of ports in one switch (k). Moreover, the tree topology provides the lowest fault resilience, as it offers the smallest bisection width.

23.3 Emerging Data Center Solutions with Photonics Technologies

Power consumption in electronics is a serious roadblock, even if we can keep up with Moore's law regarding integrating more devices on each chip to meet the demands of future applications. At the device level, *Dennard*'s law [23.47], which describes the simultaneous improvements in transistor density, switching speed, and power dissipation [23.47] to follow *Moore*'s law [23.48], was already obsolete in 2004. As Fig. 23.12 illustrates, while Moore's law continued for more than four decades, the clock speeds and the power efficiencies have flat-lined since 2004. Multicore solutions emerged shortly after 2004 as a new processor-level solution for power efficiency and scaling, sometimes called the new Moore's law. While we believe that multicore and chip-level parallelism solutions will continue to expand, communication and data movements will continue to be a challenge. For this reason, over three successive generations, the performance/Watt has improved only marginally.

Obviously, electronics alone cannot provide solutions for all the challenges of massively parallel data processing. As mentioned already in Sect. 23.1, while electronics accompany skin effects, capacitance, electromagnetic interference (EMI), and distortion/dispersion, photonics supports nearly distance-independent parallel transport across the vast optical bandwidth [23.49]. As computing nodes are evolving to multicore and multiprocessor systems with very high bandwidth requirements between processors and memory banks on the same board, interchip optical interconnects could also provide significant benefits in terms of energy per bit. In [23.49] a comparison between optical interconnects with on-chip and off-chip laser and electrical interconnects is shown, which has the promising advantages of the optical solution for distances above few tens of millimeters. More information regarding optical versus electrical interconnects is provided in [23.50–53].

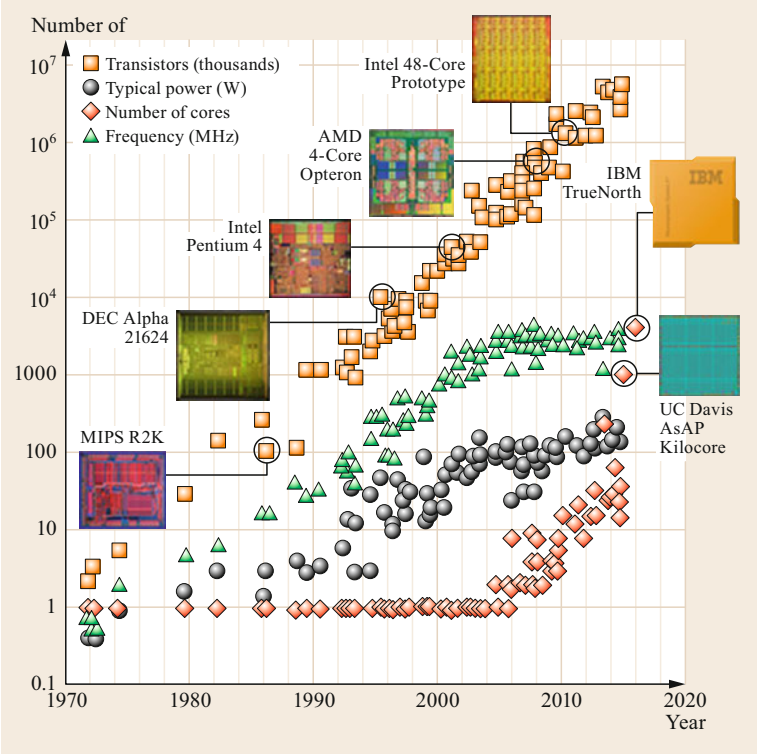


Fig. 23.12 A 45-year trend of the number of transistors per integrated circuit, clock speed (MHz), power (W), performance per clock (ILP (instructions level parallelism)), and the number of core per processor die (figure created based on data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, M. Horowitz, F. Labonte, O. Shacham, and Christopher Batten) (after [23.1])

23.3.1 Embedded Photonics, Silicon Photonics, and Heterogeneous 2.5-D Integration

We envision that future data centers will exploit photonics embedded with electronics through close integration everywhere; in chip-to-chip, board-to-board, and rack-to-rack interconnections.

While monolithic co-integration of CMOS (complementary metal–oxide–semiconductor) and silicon photonics in the same fabrication runs is attractive, the yield and the required technological compatibility challenges make it impractically expensive at the moment (e.g., 14 and 7 nm electronic CMOS compatibility with silicon photonics). Optical interposers and OE-PCBs (optoelectronic printed circuit boards) (Fig. 23.13) represent more practical and effective short-term technologies that could enable reduced parasitic, low power consumption, dense optical interconnects, and close integration of photonics and electronics, while allowing flexible combinations of heterogeneous technologies with reasonable yield.

Figure 23.14a–d illustrates a method of embedded photonics utilizing active silicon photonic interposers (optical interposer with silicon photonic modulators and detectors) interfacing with electronic integrated circuits (ICs) and OE-PCBs. Figure 23.14a and b show a side-view and a top-view schematic of 2.5-D integration of the electronic ICs, active silicon photonic interposers, and OE-PCBs achieved by this method. Figure 23.14c shows an assembly process using evanescent coupling between the silicon photonic waveguides and the OE-PCB waveguides, and Fig. 23.14d illustrates the case with such multiple optical interposers assembled on a larger OE-PCB. OE-PCBs and OE-backplanes will exploit low-loss optical waveguide layers laminated on conventional electrical PCBs.

OE-PCBs and OE-Backplanes

PCBs with embedded optical layers could offer a cost-effective opportunity to reduce energy consumptions and latency induced by electrical wires at high data rates [23.54–59]. Successful OE-PCBs will eliminate any need for high-speed electrical interconnections on board, and electrical connections will only support power and low-speed control and programming. The majority of past efforts [23.54–59] focused on multimode polymer optical waveguides within FR4 PCBs, where multimode dispersions and high losses limited the performance and energy efficiency improvements. There has been recent advances in single-mode optical polymer PCBs [23.60] and multimode glass waveguide PCBs [23.61–63] to pursue single-mode glass optical waveguides embedded in electrical PCBs. Initial efforts

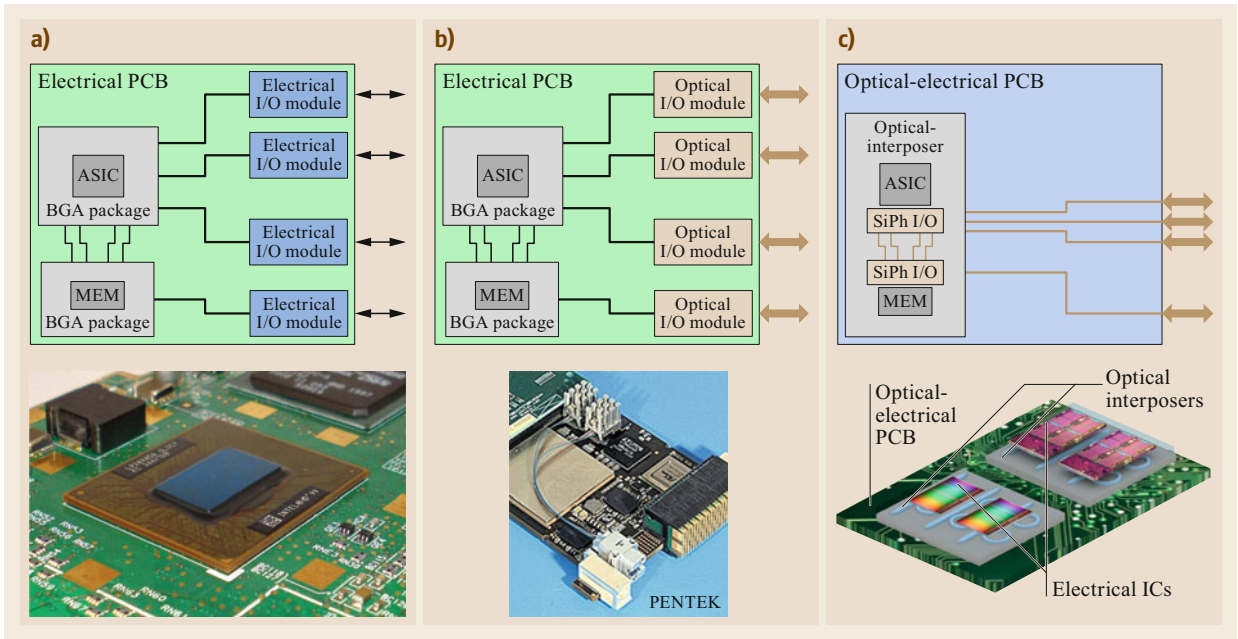


Fig. 23.13a–c Comparisons of **(a)** today's electronic interfaces with electronic I/Os. **(b)** Today's embedded optics with standard pluggable optical interfaces to BGA (ball grid array) packaged ASICs and **(c)** the proposed embedded photonics with 2.5-D and 3-D integration using silicon-photonic interposers (after [23.1])

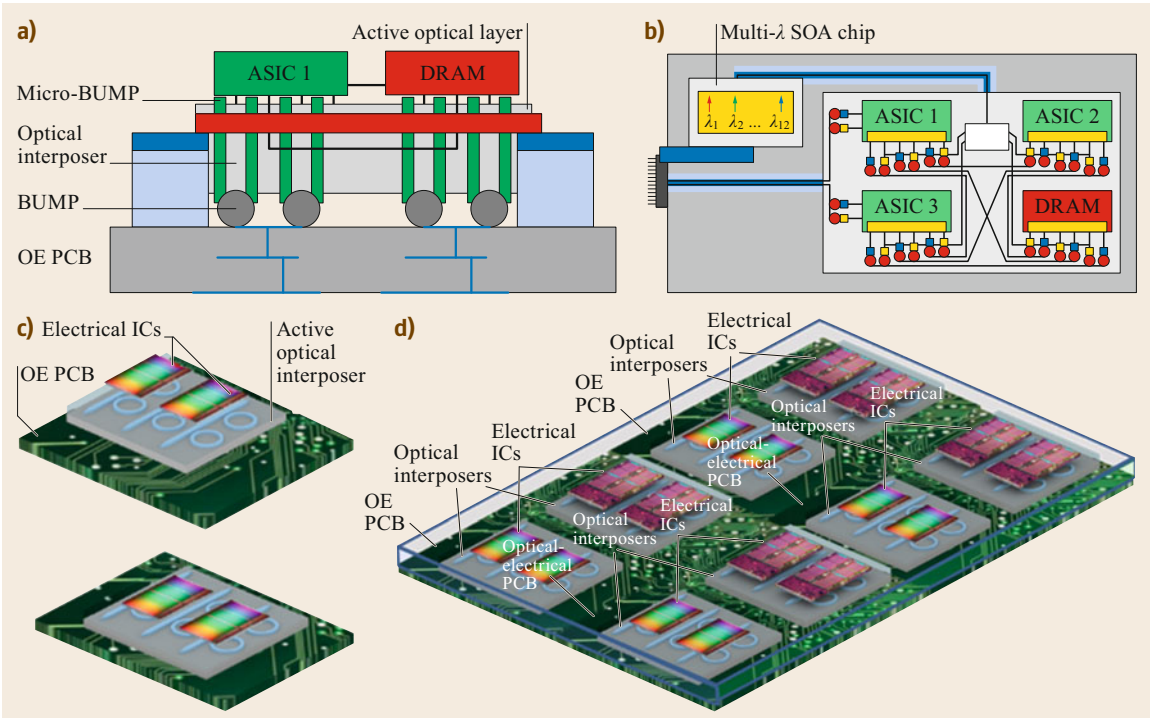


Fig. 23.14 **(a)** and **(b)** show a side-view and a top-view schematic of 2.5-D integration of the electronic ICs, active silicon photonic interposers, and OE-PCBs, and **(c)** shows an assembly process using evanescent coupling between the silicon photonic waveguides and the OE-PCB waveguides, and **(d)** illustrates an OE-PCB containing multiple silicon photonic optical interposers and electronic ICs (SOA: semiconductor optical amplifier; ASIC: application-specific integrated circuit; DRAM: direct random-access memory) (after [23.1])

will utilize the glass lamination technology mentioned in [23.62] to embed ion-exchanged silica waveguide layer in between two FR4 electrical PCBs [23.62]. This method offers relatively sturdy operation, which somewhat mitigates the difference in thermal coefficients of expansion (TCEs) between the glass and the FR4; however, it requires the opening of the FR4 in shape to drop in the optical interposer modules. Similar openings should be made on the FR4 on the other side to balance the stress and TCE difference. Successful progress in developing OE-PCBs with optical and electrical connectors allow realization of OE-PCBs and building of servers and switches interconnected with optical waveguides.

23.3.2 High-Radix Optical Switches

As mentioned in Sect. 23.1, the limited radix and bandwidth of electronic switches severely affect data center scalability in terms of latency, throughput, and power consumption. Optical switching can potentially overcome these limitations, and many optical switch architectures have been investigated and reported in the literature. Table 23.2 summarizes the main all-optical switching technologies highlighting the pros and cons of each solution. The readers should also refer to Chap. 25 *Optical switching for data center networks* for more details.

Despite the significant differences highlighted in Table 23.2, all these optical switches share a common aspect: they are bufferless (there is no buffering operation at the switch input and output ports) and, therefore, cannot be cascaded. Therefore, they could be used only as core switches in folded-CLOS type architectures (i.e., fat tree [23.39, 64]) or they could be used in directly connected architectures like torus [23.65], flattened butterfly [23.66, 67], or dragonfly [23.68], where they can directly interconnect computing nodes or ToR switches. Among the solutions in Table 23.2, MEMS switches are the only ones currently commercially available. However, due to the slow switching time of the order of milliseconds, MEMS can only switch so-called elephant flows [23.37, 69, 70], and they cannot replace the packet-switching features of electronic switches.

We envision that the next-generation high-radix high-bandwidth data center switches will make use of multiple electronic switches optically interconnected via silicon interposer and OE-PCB technologies with 2.5-D hybrid electro-optic integration platforms. The low crosstalk, very low loss and high energy efficiency provided by photonic interconnects could enable unprecedented switch bandwidth and radix. Figure 23.15 shows an example of such a hybrid approach currently under development at UC Davis NGNS laboratories. This switch is called RH-LIONS (reconfigurable hierarchical low-latency interconnect optical network switch).

RH-LIONS makes use of the electronic-photonic integration technologies discussed above to implement a switching architecture with small electrical switches at the edges that are all-optical interconnected via wavelength routing in an arrayed waveguide grating router.

An arrayed waveguide grating router (AWGR) [23.40, 71] is an example of a device with such wavelength routing capability. As Fig. 23.16 illustrates, the well-known wavelength routing property of an AWGR allows any input port to communicate with any output port simultaneously using different wavelengths without contention. Thus, an $N \times N$ AWGR

Table 23.2 Different all-optical switching technologies

	Optical 3-D MEMS all-optical switch	Piezoelectric Directlight® Beam-steering all-optical switch	Wavelength routing switch (e.g., AWGR with tunable lasers)	SOA-based broadcast and select	SOA-based crossbar	MRR-based crossbar
Switching time	≈ 1 ms	≈ 1 ms	≈ 1 ns	≈ 1 ns	≈ 1 ns	≈ 1 µs
Scalability	312 × 312	384 × 384	512 × 512	64 × 64	16 × 16	4 × 4
Applications	OCS, OPS	OCS, OPS	OCS, OPS	OCS, OPS	OCS, OPS	OCS
Level of transparency	Strict	Strict	Wavelength-dependent	Strict	Strict	Wavelength-dependent
Crosstalk	< −55 dB	< −55 dB	< −35 dB	< −45 dB	< −45 dB	< 20 dB
Polarization dependent loss (PDL)	< 0.1 dB	< 0.1 dB	Wavelength dependent	< 1 dB	< 1 dB	< 0.1 dB
Switching	Space	Space	Wavelength	Space	Space	Wavelength
References	[23.5, 11, 12]	[23.24]	[23.17–20, 26]	[23.13–16]	[23.25]	[23.21, 22]

MEMS: microelectro-mechanical system; SOA: semiconductor optical amplifier; AWGR: arrayed waveguide grating router; MRR: microring resonator; OCS: optical circuit switching; OPS: optical packet switching

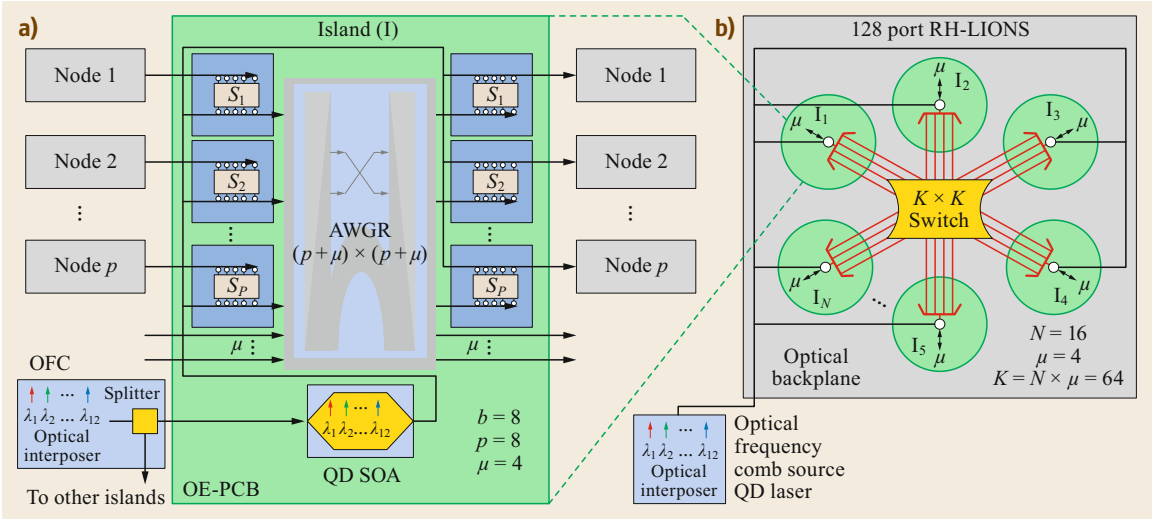


Fig. 23.15a,b A two-hierarchy RH-LIONS switch of size $pN \times pN$. Shown is an example of a 128-port switch with $N = 16$, $b = p = 8$, $\mu = 4$, $K = N \times \mu = 64$

intrinsically provides all-to-all communication among N nodes in a flat topology using N wavelengths. This realization is called a passive AWGR switch or a passive low-latency interconnect optical network switch (LIONS), since no optical reconfiguration is necessary. This is true under the assumption that each node connected to the AWGR has N transceivers (TRXs) and an embedded switch forwarding the packets to the proper TRX based on the destination.

RH-LIONS can scale beyond 128 ports and 50 Tb/s capacity. Figure 23.15 shows a two-hierarchy RH-LIONS switch of radix $p \times N$. Shown is an example of a 128-port switch with $N = 16$, $p = 8$, $b = p$, $\mu = 4$, and $K = N \times \mu = 64$. In general, a two-hierarchy RH-LIONS switch would include N islands (green circles in Fig. 23.15b). Each island is composed of p electronic switches (S in Fig. 23.15) and connects to p nodes; p also represents the number of required wavelengths and AWGR ports for intra-island all-to-all communication; μ is the number of AWGR-ports and the number of wavelengths reserved for interisland communications. Therefore, $p + \mu$ is the total number of AWGR ports and wavelengths required. To reach a very high bandwidth per switch port, we assume using a WDM optical value of b (number of wavelengths per port), where $b =$ *aggregate-switch-BW/#switch-port/line-rate/2*. Finally, a $K \times K$ optical switch (circuit switching) allows using fewer optical transceivers μ for interisland communication, while enabling topology reconfiguration between the islands. For instance, if $\mu = 4$, we can create a baseline mesh, and then the $K \times K$ optical switch can be used to modify the topology according to the traffic patterns. To build a 128-port RH-LIONS switch, we need 16 12-port AWGRs (one AWGR per island, $N = 16$ is the total

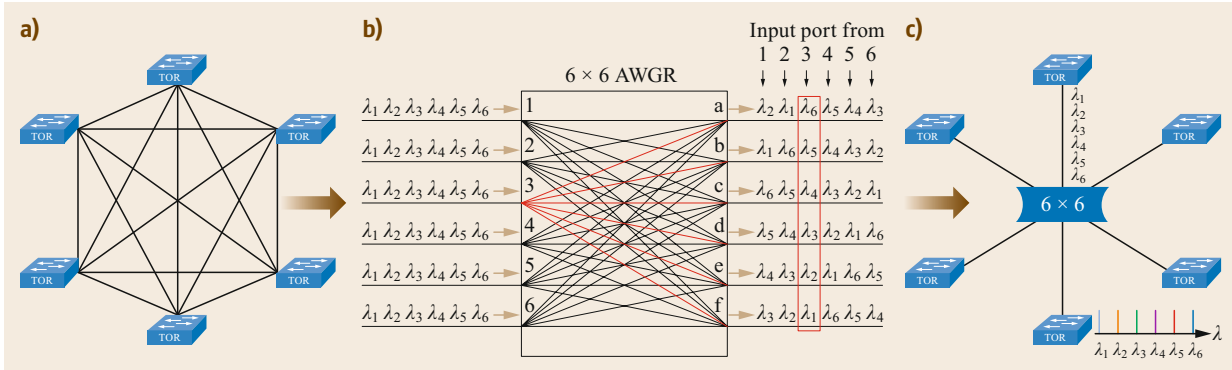


Fig. 23.16 **(a)** All-to-all interconnection with $N(N-1)$ point-to-point links ($N = 6$). **(b)** Wavelength routing in a 6-port AWGR. **(c)** All-to-all interconnection with only N links and N wavelengths

number of islands), 128 20-port E-switches ($p+\mu+b=20$ ports for S in Fig. 23.15), and one circuit-based optical switch (e.g., MEMS K-port, $K=\mu\times N$, where $\mu=4$ and $N=16$). In Fig. 23.15, each port of the switch can support up to 200 Gb/s (board-level I/O, $b=8$ WDM, 25 GHz optical frequency), for a total aggregate bandwidth of 51.2 Tb/s. The E-switch could be an energy-efficient commodity switch die (e.g., 300 mW per port with up to 24 ports at 25 Gb/s [23.72] for the packaged chip).

In the following sections, we report a few examples of architectural solutions recently published and leveraging some of the photonic interconnect and switching technologies discussed above. In particular, HALL (hierarchical all-to-all) and ARON (application-driven reconfigurable optical network), which were developed at UC Davis NGNS networking laboratories.

23.3.3 Emerging Architectures with Optical Switching

HALL

This section discusses the hierarchical all-to-all optical interconnect architecture (HALL) [23.73] (also called Hi-LIONS: hierarchical low-latency interconnect optical network switch), which implements all-to-all connection in hierarchical fashion to reach scalability levels well beyond what is possible with a single AWGR [23.74]. Figure 23.17 illustrates the architecture. The lower hierarchy is a computing blade with the same architecture as the island described in Fig. 23.15 (only replacing switches with processors). However, in general, the lower hierarchy could also be represented by a ToR switch with p wavelength-specific TRXs. The architecture in Fig. 23.17 scales as $p\times(p-1)\times(p-2)\times(p-3)$. There are $p-1$ blades in a rack, $p-2$ racks in a cluster, and $p-3$ clusters in a POD. The architecture can scale to a high number of nodes while using a limited number of wavelengths (<32). Inter-POD communication can still be achieved using another level in the hierarchy or using a different approach.

ARON

To achieve faster and agile reconfigurations, researchers at UC Davis investigated an application-driven optically-reconfigurable architecture (ARON [23.75]) by combining wavelength routing in AWGRs and nanosecond-scale wavelength-tunable transceivers [23.76]. As Fig. 23.18 shows, corresponding to the user demands or communication patterns, the full system is divided into multiple application regions (AppRegions), which are reconfigured into proper topologies (such as dragonfly, torus, etc.) by tuning the wavelengths of transceivers via a software-defined control plane.

Figure 23.18b shows the physical architecture of ARON, which consists of a basic subnetwork and a reconfigurable network. The basic subnetwork is a mesh topology that guarantees minimum full-system connectivity, while the reconfigurable network is used to perform topology reconfiguration. The proposed application-driven POD architecture is based on AWGR. A POD consists of m racks, each one containing n nodes, one AWGR ($AWGR_f$) for the basic subnetwork, and one AWGR ($AWGR_t$) for the reconfigurable network. Each node contains an r-port router with s fixed-lambda transceivers (TRXs) for the basic subnetwork and $(r-s)$ tunable TRXs for the reconfigurable subnetwork. The control plane contains an application manager (AM), an OpenFlow controller (OC), and several OpenFlow agents (OAs) responsible for topology reconfiguration via wavelength tuning. There are also n inter-rack AWGRs ($AWGR_T$).

Each $AWGR_T$ has its channels on a wavelength grid that is interleaved with the channel grid of $AWGR_t$ (i.e., both grids have a channel spacing of 0.8 nm, but they are offset to 0.4 nm). This way, each node can use the

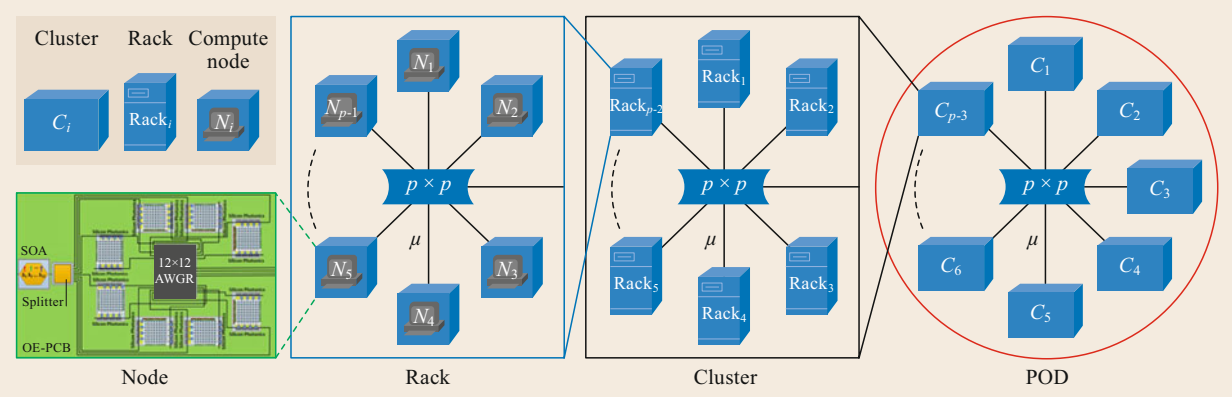


Fig. 23.17 A HALL topology

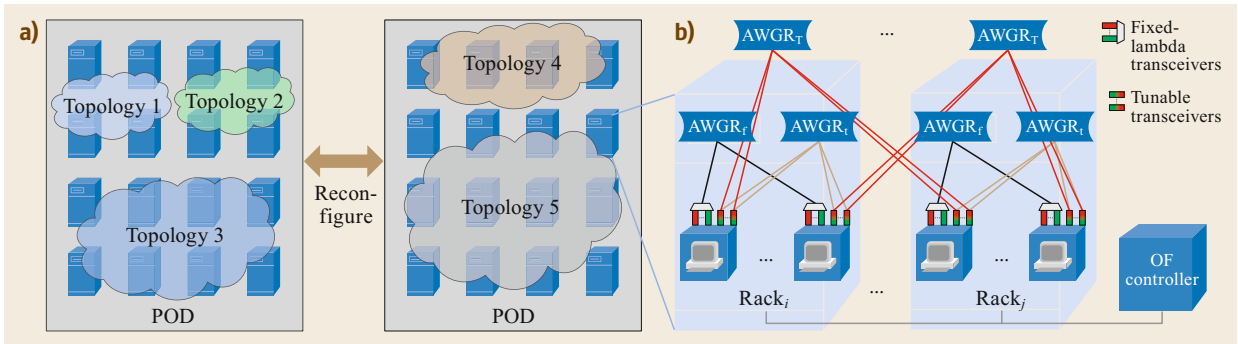


Fig. 23.18 (a) The concept of application region partitioning and (b) physical architecture of the POD

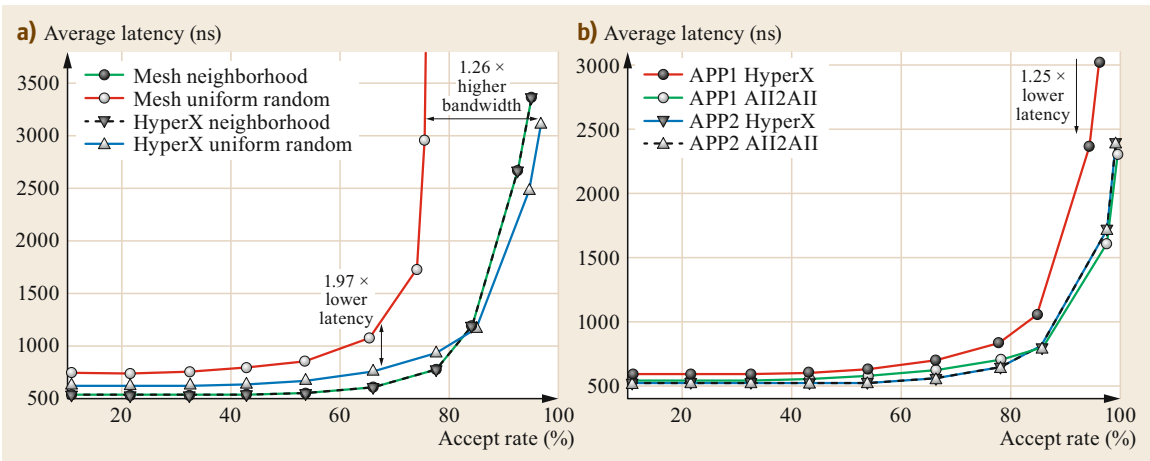


Fig. 23.19 (a) Performance of running a single application and (b) performance of running two applications (after [23.75])

same $r-s$ tunable TRXs for intra or inter-rack reconfiguration by working on a different grid [23.75].

Let us assume 64 as the typical value for the AWGR port count. Then, a POD is composed as follows:

- $m \times n$ nodes, each one with r transceivers, s of which are at fixed lambda.
- $n = 64/(r-s)$ nodes per rack. If $(r-s) = 2$, then $n = 32$.
- $m = 64/(r-s)$ racks with two AWGRs per rack. If $(r-s) = 2$, $m = 32$, and the number of AWGRs in the first layer is 64.
- A second layer of n AWGRs ($AWGR_T$).

Figure 23.19 shows an example of what could be achieved with such reconfigurability. As Fig. 23.19a shows, mesh is the suitable topology under neighborhood traffic, since it can achieve similar performance with much fewer transceivers than HyperX. When the traffic changes to uniform random, we configured to HyperX to achieve 1.26× higher throughput and 1.97× lower latency.

c-Through

To improve the network in the current data centers, *Wang* et al. [23.46, 77] proposed a hybrid electrical-optical architecture called c-Through. In c-Through, each ToR switch is connected to two networks: the electrical packet-switched network and the optical circuit-switched network. The optical network should connect ToR pairs with high traffic in between, as the optical switch only provides matching on the ToR graph [23.77].

Each host includes a traffic monitoring unit to record the bandwidth requirements across the network. This information is handed to an optical configurations manager to apply appropriate configurations to the optical switch.

A maximum weight perfect matching problem is formulated based on the traffic demand and the connected links. The perfect matching problem in c-Through is solved using *Edmonds'* algorithm [23.78].

The optical manager configures the optical circuit switch and updates the ToR switches accordingly. A VLAN-based routing with two different virtual local

area networks (VLANs) is used to route the traffic in each ToR. One VLAN is used for the packet-switched network and one is used for the circuit-switched network. For a given packet, the second VLAN is used if the sending ToR is connected to the destination ToR through the optical circuit.

Simulation results based on microbenchmarks as well as real-world applications shows that c-Through is well suited for slowly changing traffic patterns. For such applications, c-Through significantly reduces the latency between the two ToRs connected through the optical circuit and reduces the completion time [23.33].

Helios

Helios [23.79] is a hybrid electrical/optical switch architecture for modular data centers and exhibits a similar architecture to c-Through. However, different from c-Through, Helios relies on WDM links and its scheme follows the design of typical two-layer data center networks, ToR switches (referred to as POD switches), and core switches. Conventional electrical packet switches can be used for POD switches. For core switches, either optical circuit switches or electrical packet switches can be used.

Helios takes advantage of both electrical and optical networks, each for different scenarios. For POD switch pairs requiring high-bandwidth links, the optical circuit switch is used. On the other hand, all-to-all traffic patterns among different POD switches are serviced using electrical packet switches.

The high-level architecture of Helios is illustrated in Fig. 23.20. Each POD is required to have both colorless and WDM transceivers. POD switches are connected to electrical packet switches at the core level through colorless transceivers. WDM transceivers are used to connect the POD to the core optical circuit switches. Using a passive optical multiplexer, W links from WDM transceivers are multiplexed to form a superlink with a capacity of W times the capacity of the transceiver. Assuming the same number of WDM and colorless transceivers on each POD, half of the total bandwidth can be shared among PODs. The remaining half can be allocated according to the traffic pattern.

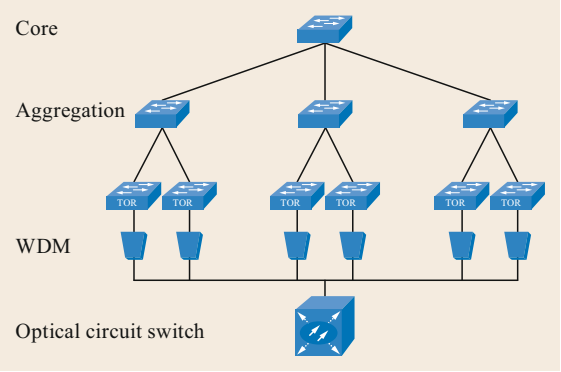


Fig. 23.20 The Helios architecture

In the Helios architecture, the control unit consists of three units: the topology manager (TM), the circuit switch manager (CSM), and the POD switch manager (PSM). The topology manager seeks the optimal configuration for the optical circuit switch by monitoring the traffic. According to the decisions made by the TM unit, the POD switch manager decides whether to use the optical circuit switch or the electrical packet switch to route each packet. The circuit switch manager configures the Glimmerglass MEMS switch based on the connections graph it receives.

In the Helios architecture, the main advantage comes from using off-the-shelf optical components and transceivers. The optical transceivers used in POD switches are wavelength division multiplexing (WDM) transceivers. A switch from Glimmerglass [23.79, 80] is used for the optical circuit switches. For the performance analysis provided, a Glimmerglass crossbar optical circuit switch has been used. Helios also offers a trade-off between cost and the number of wavelengths per link. The more expensive dense WDM units can support more wavelengths, while coarse WDM units offer support for a limited number of wavelengths at a lower cost [23.33].

The shortcoming of this approach lies in the reconfiguration time of the MEMS switches used. A reconfiguration time of several milliseconds make Helios a good candidate for communication patterns with connections longer than a few seconds. The significant increase in throughput as the stability parameter increases further reveals the impact of the reconfiguration overhead.

OSA

OSA [23.15] is another architecture taking advantage of commercially available MEMS switches and also makes use of wavelength selective switches to achieve dynamic topology reconfiguration by exploiting spatial and wavelength switching.

Assume that each ToR switch is connected to MEMS through a single port out of N ports available. Therefore, considering the bipartite port-matching in MEMS, a ToR may communicate with only one other ToR at any time. Thus, the graph at ToR level is disconnected. Alternatively, each ToR can communicate with K other ToRs at any instant, if K ports are used to connect each ToR to the MEMS. However, the total number of ToRs connected to one MEMS is reduced from N to N/K. Note that here, K is not the port count but the degree of each ToR within the ToR graph. OSA ensures the connectivity of the ToR graph through MEMS con-

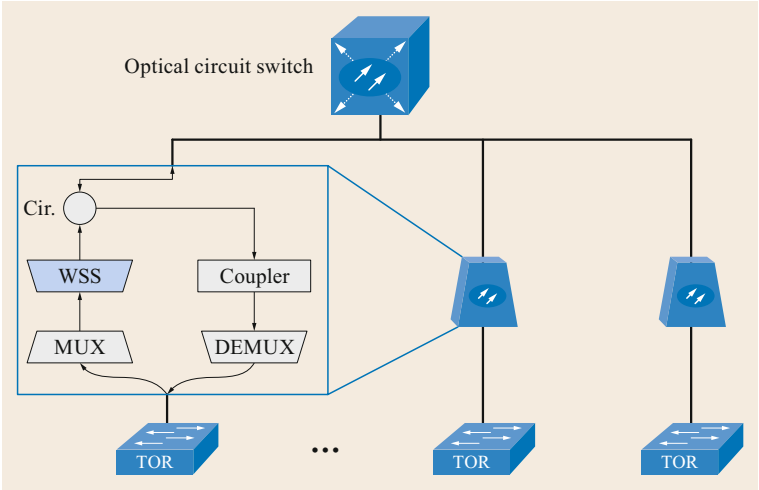


Fig. 23.21 The OSA architecture

figuration, as connected sets of ToRs are defined though MEMS configuration.

In the OSA architecture, to achieve connectivity across the network, communicating with the ToRs beyond the topology connected by the local MEMS requires traversing through remote MEMS hops. Such packets should get through optical–electrical–optical (O–E–O) conversion in order to read the packet header and identify the destination at each hop. The O–E–O conversion-including protocol overheads imposes considerable latency to packet transmission. Moreover, the bandwidth offered by each MEMS port is limited. Therefore, OSA should maintain a minimum number of hops for connections where significant traffic is expected.

For communication with higher data rates than what a single ToR link offers, OSA uses wavelength demultiplexing (WDM), as optical fibers are capable of transmitting multiple wavelengths simultaneously. As Fig. 23.21 illustrates, each ToR connects to the MEMS through a wavelength selective switch (WSS) unit and a multiplexer (MUX). If the communication between one ToR and another requires W times the rate of a single port, W ports (each dedicated to a unique wavelength) are used to service this request. Using WDM, these W wavelengths are multiplexed into a single fiber feeding the WSS unit. Each wavelength is mapped to the corresponding port in the MEMS by the WSS unit. The OSA architecture provides a dynamic link capacity at runtime by choosing appropriate values for parameter W.

OSA requires bidirectional circuits in MEMS in order to efficiently use all MEMS ports. Thus, optical circulators are needed between MEMS ports and the ToR. A send channel of the transceiver from ToR passes though the MUX and WSS and then connects to MEMS port through the optical circulator. Incoming traffic from MEMS port to the ToR is delivered simultaneously through the coupler and the demultiplexer (DEMUX). It should be noted that the capacity links in each direction are independent of each other, even though the links within MEMS are bidirectional.

23.4 Conclusions

This chapter presented an overview of the intradata center interconnect topologies and architectures used and currently deployed from cloud computing service providers as well as other architectures proposed and studied in research institutions. Among the various solutions discussed, some provide the better scalability often associated with an increased average number of hops or cabling complexity. Architectures employing reconfigurability provide the flexibility needed for hosting diverse applications and high-performance computations.

Like in any engineering problem, the oracle solution highly depends on the application of the system. Hence, the design choice should take into account the characteristics of the applications, as well as the deployment, maintenance issues, and capital and operating costs. As our demands for various data services grow, we expect that new data center architectures exploiting the benefits of emerging photonic technologies will emerge to meet the ever-changing requirements of various applications.

References

23.1 S.J. Ben Yoo, R. Proietti, P. Grani: Photonics in data centers. In: *Optical Switching in Next Generation Data Centers*, ed. by F. Testa, L. Pavesi (Springer, Cham 2018)

23.2 A. Shehabi, S. Smith, D. Sartor, R. Brown, M. Herrlin, J. Koomey, E. Masanet, N. Horner, I. Azevedo, W. Lintner: *United States Data Center Energy Usage Report* (LBNL, Berkeley 2016)

23.3 L.A. Barroso, J. Clidaras, U. Hölzle: The datacenter as a computer: an introduction to the design of warehouse-scale machines, Synth. Lect. Comput. Archit. **8**, 1–154 (2013)

23.4 COBO: Consortium for On-Board Optics, http://cobo.azurewebsites.net (2015)

23.5 Cisco: *Cisco Global Cloud Index: Forecast and Methodology, 2015–2020*, White Paper, 2016)

23.6 D.A.B. Miller, H.M. Ozaktas: Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture, J. Parallel Distributed Comput. **41**, 42–52 (1997), 02/25 1997

23.7 A. Naeemi, X. Jianping, A.V. Mule, T.K. Gaylord, J.D. Meindl: Optical and electrical interconnect partition length based on chip-to-chip bandwidth maximization, IEEE Photonics Technol. Lett. **16**, 1221–1223 (2004)

23.8 R.G. Beausoleil, P.J. Kuekes, G.S. Snider, S.Y. Wang, R.S. Williams: Nanoelectronic and nanophotonic interconnect, Proceedings IEEE **96**, 230–247 (2008)

23.9 International Roadmap Committee: *International Technology Roadmap for Semiconductors: 2013 Edition Executive Summary* (Semiconductor Industry Association, San Francisco 2013), https://www.semiconductors.org/wp-content/uploads/2018/08/2013ExecutiveSummary.pdf

23.10 C.A. Thraskias, E.N. Lallas, N. Neumann, L. Schares, B.J. Offrein, R. Henker, D. Plettemeier, F. Ellinger, J. Leuthold, I. Tomkos: Survey of photonic and plasmonic interconnect technologies for intra-datacenter and high-performance computing communications, IEEE Commun. Surv. Tutor. **20**(4), 2758–2783 (2018)

23.11 T. Benson, A. Anand, A. Akella, M. Zhang: Understanding data center traffic characteristics. In: *Proc. 1st ACM Workshop Res. Enterp. Netw.* (2009) pp. 65–72

23.12 T. Benson, A. Akella, D.A. Maltz: Network traffic characteristics of data centers in the wild. In: *Proc. 10th ACM SIGCOMM Conf. Internet Meas.* (2010) pp. 267–280

23.13 A. Roy, H. Zeng, J. Bagga, G. Porter, A.C. Snoeren: Inside the social network's (datacenter) network, SIGCOMM Comput. Commun. Rev. **45**, 123–137 (2015)

23.14 S. Han, T.J. Seok, N. Quack, B.-W. Yoo, M.C. Wu: Large-scale silicon photonic switches with movable directional couplers, Optica **2**, 370–375 (2015)

23.15 K. Chen, A. Singla, A. Singh, K. Ramachandran, L. Xu, Y. Zhang, X. Wen, Y. Chen: OSA: an optical switching architecture for data center networks with unprecedented flexibility, IEEE/ACM Trans. Netw. **22**, 498–511 (2014)

23.16 N. Farrington, G. Porter, S. Radhakrishnan, H.H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, A. Vahdat: Helios: a hybrid electrical/optical switch architecture for modular data centers. In: *Proc. ACM SIGCOMM 2010 Conf., New Delhi* (2010), Paper

23.17 R. Hemenway, R.R. Grzybowski, C. Minkenberg, R. Luijten: Optical-packet-switched interconnect for supercomputer applications, J. Opt. Netw. **3**(12), 900–913 (2004)

23.18 C. Hawkins, B.A. Small, D.S. Wills, K. Bergman: The data vortex, an all optical path multicomputer interconnection network, IEEE Trans. Parallel Distributed Syst. **18**, 409–420 (2007)

23.19 O. Liboiron-Ladouceur, A. Shacham, B.A. Small, B.G. Lee, H. Wang, C.P. Lai, A. Biberman, K. Bergman: The data vortex optical packet switched interconnection network, J. Lightwave Technol. **26**(13), 1777–1789 (2008)

23.20 K. Bergman, D. Keezer, S. Wills: Design, demonstration and evaluation of an all optical processor memory-interconnection network for petaflop supercomputing. In: *ACS Interconnects Workshop* (2010), http://lightwave.ee.columbia.edu/?s=research&p=high-performance_computing_systems#dv.

23.21 P. Grani, R. Proietti, S. Cheung, S.J. Ben Yoo: Flat-topology high-throughput compute node with AWGR-based optical-interconnects, J. Lightwave Technol. **34**, 2959–2968 (2016)

23.22 R. Proietti, Y. Yin, R. Yu, C.J. Nitta, V. Akella, C. Mineo, S.J. Ben Yoo: Scalable optical interconnect architecture using AWGR-based TONAK LION switch with limited number of wavelengths, J. Lightwave Technol. **31**, 4087–4097 (2013)

23.23 R. Proietti, Y.W. Yin, R.X. Yu, X.H. Ye, C. Nitta, V. Akella, S.J. Ben Yoo: All-optical physical layer NACK in AWGR-based optical interconnects, IEEE Photonics Technol. Lett. **24**, 410–412 (2012)

23.24 R. Proietti, Y. Yawei, Y. Runxiang, C.J. Nitta, V. Akella, C. Mineo, S.J. Ben Yoo: Scalable optical interconnect architecture using AWGR-based TONAK LION switch with limited number of wavelengths, J. Lightwave Technol. **31**, 4087–4097 (2013)

23.25 A.S.P. Khope, A.A.M. Saleh, J.E. Bowers, R.C. Alferness: Elastic WDM crossbar switch for data centers. In: *2016 IEEE Opt. Interconnects Conf. (OI)* (2016) pp. 48–49, https://doi.org/10.1109/OIC.2016.7483001

23.26 N. Binkert, A. Davis, N.P. Jouppi, M. McLaren, N. Muralimanohar, R. Schreiber, J.H. Ahn: The role of optics in future high radix switch design. In: *38th Proc. Annual Int. Symp. Comput. Archit. (ISCA)* (2011) pp. 437–447

23.27 S. Cheung, T. Su, K. Okamoto, S. Yoo: Ultra-compact silicon photonic 512×512 25 GHz arrayed waveguide grating router, IEEE J. Sel. Top. Quantum Electron. **20**, 310–316 (2014)

23.28 Huber+Suhner: Single mode network optical switch from 4×4 to 192×192 ports, https://www.hubersuhner.com/en/documents-repository/

technologies/pdf/data-sheets-optical-switches/polatis-6000-network-data-sheet (2017)
23.29 R. Stabile, A. Albores-Mejia, A. Rohit, K.A. Williams: Integrated optical switch matrices for packet data networks, Microsyst. Nanoeng. **2**, 15042 (2016)
23.30 Y. Shang, D. Li, M. Xu: Green routing in data center network: modeling and algorithm design. In: *Proc. First ACM SIGCOMM Workshop Green Netw.* (2010), https://doi.org/10.1145/1851290.1851292
23.31 A. Hammadi, L. Mhamdi: Review: a survey on architectures and energy efficiency in data center networks, Comput. Commun. **40**, 1–21 (2014)
23.32 M. Al-Fares, S. Radhakrishnan, B. Raghavan, N. Huang, A. Vahdat: Hedera: Dynamic flow scheduling for data center networks. In: *NSDI '10 Proc. 7th USENIX Conf. Netw. Syst. Des. Implement.* (2010) p. 19
23.33 C. Kachris, I. Tomkos: A survey on optical interconnects for data centers, IEEE Commun. Surv. Tutor. **14**, 1021–1036 (2012)
23.34 K. Wu, J. Xiao, L.M. Ni: Rethinking the architecture design of data center networks, Frontiers Comput. Sci. **6**, 596–603 (2012)
23.35 Y. Liu, J.K. Muppala, M. Veeraraghavan, D. Lin, M. Hamdi: *Data Center Networks: Topologies, Architectures and Fault-Tolerance Characteristics*, SpringerBriefs in Computer Science (Springer, Berlin, Heidelberg 2013), https://doi.org/10.1007/978-3-319-01949-9
23.36 R.N. Mysore, A. Pamboris, N. Farrington, N. Huang, P. Miri, S. Radhakrishnan, V. Subramanya, A. Vahdat: PortLand: a scalable fault-tolerant layer 2 data center network fabric, SIGCOMM Comput. Commun. Rev. **39**, 39–50 (2009)
23.37 M. Al-Fares, S. Radhakrishnan, B. Raghavan, N. Huang, A. Vahdat: Hedera: dynamic flow scheduling for data center networks. In: *Proc. 7th USENIX Conf. Netw. Syst. Des. Implement.* (2010)
23.38 C.E. Leiserson: Fat-trees: universal networks for hardware-efficient supercomputing, IEEE Trans. Comput. **34**, 892–901 (1985)
23.39 M. Al-Fares, A. Loukissas, A. Vahdat: A scalable, commodity data center network architecture, SIGCOMM Comput. Commun. Rev. **38**, 63–74 (2008)
23.40 B. Glance, I.P. Kaminow, R.W. Wilson: Applications of the integrated waveguide grating router, J. Lightwave Technol. **12**, 957–962 (1994)
23.41 J.H. Ahn, N. Binkert, A. Davis, M. McLaren, R.S. Schreiber: HyperX: Topology, routing, and packaging of efficient large-scale networks. In: *Proc. Conf. High Perform. Comput. Netw. Storage Anal., Portland* (2009)
23.42 J. Kim, W.J. Dally, D. Abts: Flattened butterfly: a cost-efficient topology for high-radix networks, SIGARCH Comput. Archit. News **35**, 126–137 (2007)
23.43 W. Dally, B. Towles: *Principles and Practices of Interconnection Networks* (Morgan Kaufmann, New York 2003)
23.44 W.J. Dally: Performance analysis of k-ary n-cube interconnection networks, IEEE Trans. Comput. **39**, 775–785 (1990)
23.45 C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, S. Lu: Dcell: a scalable and fault-tolerant network structure for data centers, ACM SIGCOMM Comput. Commun. **38**(4), 75–86 (2008)
23.46 G. Wang, D.G. Andersen, M. Kaminsky, M. Kozuch, T. Ng, K. Papagiannaki, M. Glick, L. Mummert: Your data center is a router: The case for reconfigurable optical circuit switched paths, ACM SIGCOMM Comput. Commun. Rev. SIGCOMM **40**(4), 327–338 (2010)
23.47 R.H. Dennard, F.H. Gaensslen, V.L. Rideout, E. Bassous, A.R. LeBlanc: Design of ion-implanted MOSFET's with very small physical dimensions, IEEE J. Solid-State Circuits **9**(5), 256–268 (1974)
23.48 G.E. Moore: Cramming more components onto integrated circuits, Electronics **38**(8), 114 (1965)
23.49 M. Stucchi, S. Cosemans, J. Van Campenhout, Z. Tőkei, G. Beyer: On-chip optical interconnects versus electrical interconnects for high-performance applications, Microelectron. Eng. **112**, 84–91 (2013)
23.50 H. Cho, P. Kapur, K.C. Saraswat: Power comparison between high-speed electrical and optical interconnects for Interchip communication, J. Lightwave Technol. **22**, 2021 (2004)
23.51 C. Guoqing, C. Hui, M. Haurylau, N. Nelson, D. Albonesi, P.M. Fauchet, E.G. Friedman: Electrical and optical on-chip interconnects in scaled microprocessors, IEEE Int. Symp. Circuit. Syst. **3**, 2514–2517 (2005)
23.52 D.A.B. Miller: Optical interconnects to electronic chips, Appl. Opt. **49**, F59–F70 (2010)
23.53 S. Rakheja, V. Kumar: Comparison of electrical, optical and plasmonic on-chip interconnects based on delay and energy considerations. In: *13th Int. Symp. Qual. Electron. Des. (ISQED)* (2012) pp. 732–739
23.54 T. Ishigure, K. Shitanda, T. Kudo, S. Takayama, T. Mori, K. Moriya, K. Choki, IEEE: Low-loss design and fabrication of multimode polymer optical waveguide circuit with crossings for high-density optical PCB. In: *IEEE 63rd Electron. Compon. Technol. Conf. (ECTC)* (2013) pp. 297–304
23.55 R. Kinoshita, K. Moriya, K. Choki, T. Ishigure: Polymer optical waveguides with GI and W-shaped cores for high-bandwidth-density on-board interconnects, J. Lightwave Technol. **31**, 4004–4015 (2013)
23.56 R. Pitwon, M. Immonen, K. Wang, H. Itoh, T. Shioda, J.H. Wu, L.X. Zhu, H.J. Yan, A. Worrall: International standards for optical circuit board fabrication, assembly and measurement, Opt. Commun. **362**, 22–32 (2016)
23.57 R. Pitwon, K. Wang, M. Immonen, J.H. Wu, L.X. Zhu, H.J. Yan, A. Worrall: International standardisation of optical circuit board measurement and fabrication procedures, Proceedings SPIE (2015), https://doi.org/10.1117/12.2077654
23.58 A.F. Rizky, N. Nishimura, Y. Nekado, T. Uemura, H. Nasu, Ieee: Polymer waveguide-coupled 14-Gb/s x 12-channel parallel-optical modules mounted on optical PCB through Sn-Ag-Cu-solder reflow. In:

IEEE 3rd CPMT Symp. Kyoto (2013), https://doi.org/10.1109/ICSJ.2013.6756090

23.59 K. Soma, T. Ishigure: Fabrication of a graded-index circular-core polymer parallel optical waveguide using a microdispenser for a high-density optical printed circuit board, IEEE J. Sel. Top. Quantum Electron. **19**(2), 3600310 (2013)

23.60 R. Dangel, J. Hofrichter, F. Horst, D. Jubin, A. La Porta, N. Meier, I.M. Soganci, J. Weiss, B.J. Offrein: Polymer waveguides for electro-optical integration in data centers and high-performance computers, Opt. Express **23**, 4736–4750 (2015)

23.61 H. Schröder, L. Brusberg, N. Arndt-Staufenbiel, J. Hofmann, S. Marx: Glass panel processing for electrical and optical packaging. In: *Proc. IEEE 61st Electron. Compon. Technol. Conf. (ECTC)* (2011) pp. 625–633

23.62 H. Schröder, L. Brusberg, N. Arndt-Staufenbiel, K. Richlowski, C. Ranzinger, K.D. Lang: Advanced thin glass based photonic PCB integration. In: *Proc. IEEE 62nd Electron. Compon. Technol. Conf.* (2012) pp. 194–202

23.63 L. Brusberg, H. Schröder, C. Herbst, C. Frey, C. Fiebig, A. Zakharian, S. Kuchinsky, X. Liu, D. Fortusini, A. Evans: High performance ion-exchanged integrated waveguides in thin glass for board-level multimode optical interconnects. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)* (2015), https://doi.org/10.1109/ECOC.2015.7341714

23.64 W. Coomans, L. Gelens, S. Beri, J. Danckaert, G. Van der Sande: Solitary and coupled semiconductor ring lasers as optical spiking neurons, Phys. Rev. E **84**, 036209 (2011)

23.65 S. Horiguchi, T. Ooki: Hierarchical 3D-torus interconnection network. In: *Proc. IEEE Int. Symp. Parallel Archit. Algorithms Netw. (ISPAN 2000)* (IEEE CS, Richardson 2000) pp. 50–56

23.66 D. Abts, M.R. Marty, P.M. Wells, P. Klausler, H. Liu: Energy proportional datacenter networks, ACM SIGARCH Comput. Archit. News **38**, 338–347 (2010)

23.67 J. Kim, W.J. Dally, D. Abts: Flattened butterfly: a cost-efficient topology for high-radix networks, Comput. Archit. News **35**, 126–137 (2007)

23.68 J. Kim, W.J. Dally, S. Scott, D. Abts: Technology-driven, highly-scalable dragonfly topology, ACM SIGARCH Comput. Archit. News **36**(3), 77–78 (2008)

23.69 F. Tang, L. Li, L. Barolli, C. Tang: An efficient sampling and classification approach for flow detection in SDN-based big data centers. In: *IEEE 31st Int. Conf. Adv. Inf. Netw. Appl. (AINA)* (2017) pp. 1106–1115

23.70 Y. Zhang, L. Cui, Y. Zhang: A stable matching based elephant flow scheduling algorithm in data center networks, Comput. Netw. **120**, 186–197 (2017)

23.71 S. Kamei, M. Ishii, M. Itoh, T. Shibata, Y. Inoue, T. Kitagawa: 64×64-channel uniform-loss and cyclic-frequency arrayed-waveguide grating router module, Electron. Lett. **39**, 83–84 (2003)

23.72 Integrated Device Technology: *RXS RapidIO Switches* (IDT, San José 2016), http://www.prnewswire.com/news-releases/idt-launches-next-generation-rapidio-switches-for-5g-mobile-network-development-and-mobile-edge-computing-300221151.html

23.73 Z. Cao, R. Proietti, S. Yoo: HALL: a hierarchical all-to-all optical interconnect architecture. In: *Opt. Interconnects Conf.* (2014) pp. 73–74

23.74 R. Yu, S. Cheung, Y. Li, K. Okamoto, R. Proietti, Y. Yin, S.J. Ben Yoo: A scalable silicon photonic chip-scale optical switch for high performance computing systems, Opt. Express **21**, 32655–32667 (2013)

23.75 G. Yuan, R. Proietti, X. Liu, A. Castro, D. Zang, N. Sun, C. Liu, Z. Cao, S. Yoo: ARON: application-driven reconfigurable optical networking for HPC data centers. In: *Proc. 42nd Eur. Conf. Opt. Commun. (ECOC)* (VDE, Düsseldorf 2016)

23.76 Z. Cao, R. Proietti, M. Clements, S.B. Yoo: Experimental demonstration of flexible bandwidth optical data center core network with all-to-all interconnectivity, J. Lightwave Technol. **33**, 1578–1585 (2015)

23.77 G. Wang, D.G. Andersen, M. Kaminsky, K. Papagiannaki, T.S.E. Ng, M. Kozuch, M. Ryan: c-Through: part-time optics in data centers. In: *Proc. ACM SIGCOMM 2010 Conf., New Delhi* (2010)

23.78 J. Edmonds: Paths, trees, and flowers, Can. J. Math. **17**, 449–467 (1965)

23.79 N. Farrington, G. Porter, S. Radhakrishnan, H.H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, A. Vahdat: Helios: a hybrid electrical/optical switch architecture for modular data centers, ACM SIGCOMM Comput. Commun. Rev. **40**, 339–350 (2010)

23.80 X. Wang, M. Veeraraghavan, Z. Lin, E. Oki: Optical switch in the middle (OSM) architecture for DCNs with Hadoop adaptations. In: *IEEE Int. Conf. Commun. (ICC), Paris* (2017), https://doi.org/10.1109/ICC.2017.7996504

Roberto Proietti

Dept. of Electrical and Computer Engineering
University of California, Davis
Davis, CA, USA
rproietti@ucdavis.edu

Dr Proietti received his MS degree from the University of Pisa, Italy, in 2004 and his PhD from Scuola Superiore Sant'Anna, Pisa, Italy, in 2009. He is currently Associate Project Scientist at the University of California, Davis. His research interests include optical switching technologies and architectures for supercomputing and data centers, coherent transmission systems and elastic optical networking, and radio over fiber.

Pouya Fotouhi

Dept. of Electrical and Computer Engineering
University of California, Davis
Davis, CA, USA
pfotouhi@ucdavis.edu

Pouya Fotouhi earned a BSc in Electrical Engineering from Isfahan University. He started graduate school at the University of Delaware, where he received his MSc in Computer Engineering. He is currently pursuing his PhD in Computer Engineering at the University of California, Davis. Pouya's research activities includes optical interconnects in HPC architectures and their impact on the memory system.

Sebastian Werner

Dept. of Electrical and Computer Engineering
University of California, Davis
Davis, CA, USA
swerner@ucdavis.edu

Dr Sebastian Werner is a research scientist at the Next Generation Networking Systems Laboratory, University of California, Davis. He received his BS and MS degrees in Computer Engineering from Darmstadt University of Technology, Germany, and his PhD in Computer Science from the University of Manchester, UK. His current research interests include computer architecture, memory systems, optical interconnects, and networks-on-chip.

S.J. Ben Yoo

Dept. of Electrical and Computer Engineering
University of California, Davis
Davis, CA, USA
sbyoo@ucdavis.edu

S.J. Ben Yoo is a Distinguished Professor at the University of California at Davis. He obtained his Bachelor's, Master's and PhD degrees from Stanford University (1984, 1986, 1991). His research interests include 2-D/3-D photonic integration for future computing, cognitive networks, communication, imaging, navigation systems, and micro/nanosystems integration. Professor Yoo is a Fellow of IEEE, OSA, and NIAC.

24. System Aspects for Optical Interconnect Transceivers

Brad Booth, David Piehler

Data centers have shown and will continue to show tremendous growth. The data center requirements and the resulting architectures determine the optimal optical interconnect approach. This work outlines hyperscale data center architectures and implications for the optical transceiver technology used therein. The Microsoft hyperscale data center architecture leads to the choice of the four-lane parallel single-mode fiber (PSM4) transceiver as optimal. We review the various optical transmitter technologies, including directly modulated lasers (DMLs), integrated distributed feedback laser/electroabsorption modulators (DFB/EAM), and silicon photonics (SiPho). The increase in optical signaling rates from 25 to 50–100 Gb/s per wavelength is described, along with their inclusion in defined optical specifications from the IEEE Ethernet standardization organization, as well as industry multisource agreements (MSAs). Finally, we show how data center bandwidth requirements will lead to an evolution of optical packaging paradigms from pluggable modules to on-board-optics and copackaged optics.

Data centers come in a variety of shapes and sizes, from those with a few servers in a rack to hundreds of thousands of them in a massive warehouse. Whether the data center is storing corporate email or performing calculations for a new medicine, the network is a critical element for the movement of data, be it a data center in the cloud or in a closet down the hall. People continue to generate data. Machines generate data. The network permits that data to be shared with other people or other machines. It is what gives us the ability to share photographs or videos, permits your favorite app to run, or even have cars communicate with other cars. As the amount of information continues to grow, so does the need for data centers.

24.1 Data Center Requirements

This trend was documented in the IEEE 802.3™ Industry Connections Ethernet Bandwidth Assessment report [24.1]. The report was a snapshot of the industry in 2012, and since that time cloud, the Internet of things (IoT), 5G, artificial intelligence (AI), and machine learning have appeared in the industry to further drive the growth of traffic.

While data centers may use the similar philosophies for creating their networks, there are business-related decisions that influence how they are designed. This design, in turn, impacts the choices made for the interconnect technology. Typically, networks are optimized for their workloads and their future growth, but the total cost of ownership plays a critical role.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_24

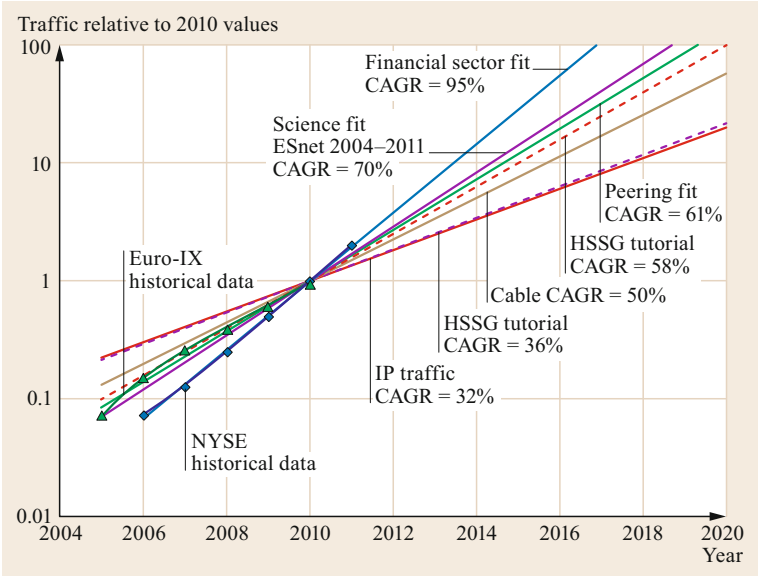


Fig. 24.1 Data center bandwidth growth projections normalized to 2010. (CAGR = compound annual growth rate, HSSG = Higher Speed Study Group) (after [24.1])

24.1.1 Workloads

The variation in workloads is as vast as the difference between posting an update to a social media site versus performing a computational analysis for landing a probe on Mars. Data centers can be designed to support one or two application workloads, or they can be designed to support a vast array of applications. For example, a data center designed for a social media website may only have one or two applications running on it; whereas, a data center designed for business clients may have to support hundreds of applications.

These different workloads place different demands on the data center network architecture. To run a computationally intensive workload like high-performance computing (HPC) or deep neural networks (DNN), latency plays a significant role. The longer the data is in transit over the data center network, the longer the next process in the computation must wait. In the case of uploading a video or photograph, the large amount of data and the speed of the network have a direct and noticeable impact on the person uploading the file; therefore, the speed of the network is key to the time it takes to upload, but latency in the data center network is insignificant in comparison to the time it takes to transfer the data to the data center.

Figure 24.2 shows an image often used by Facebook to describe their network topology.

Figure 24.3 is used by Microsoft to describe their network architecture (DC = data center, RNG = regional network gateway).

The difference in the appearance of these architectures is due the size of the networks, the application workloads supported, and the equipment used within the network. While Microsoft and Facebook both operate large data centers, there is a difference in their customer base, their revenue generation, their data centers in operation, and their data center growth strategy.

24.1.2 Growth

Building and growing large-scale data centers is a complex operation. The networking portion of the data center needs to be considered because it has an impact on the cost and the lifetime. In cloud data centers, the server is connected to the first switch tier via a passive direct-attach copper (DAC) cable or via an active optical cable (AOC). DACs and AOCs are commonly used in areas of the network where there is the greatest volume of connections. In the diagram of the Microsoft architecture (Fig. 24.3), the servers connect to the tier 0 switch via DACs. The servers and tier 0 switch are in a rack. The tier 1 switch is an end-of-row or middle-of-row switch connected to the tier 0 switches with AOCs. Because Microsoft growth and refresh cycles are not based on a single server or a rack of servers, but rather on a row of servers, the network equipment and cables are either new or replaced (Table 24.1). From the tier 1 to the tier 3, that cable infrastructure (fiber optics) can be either part of the building or in conduit.

The type of optical fiber used has a direct impact on the growth of the data center and the technologies that can be considered. Although multimode fiber (MMF) has continued to improve with each successive generation there has been an impact to the reach. Table 24.2 gives an indication of this, with the fastest

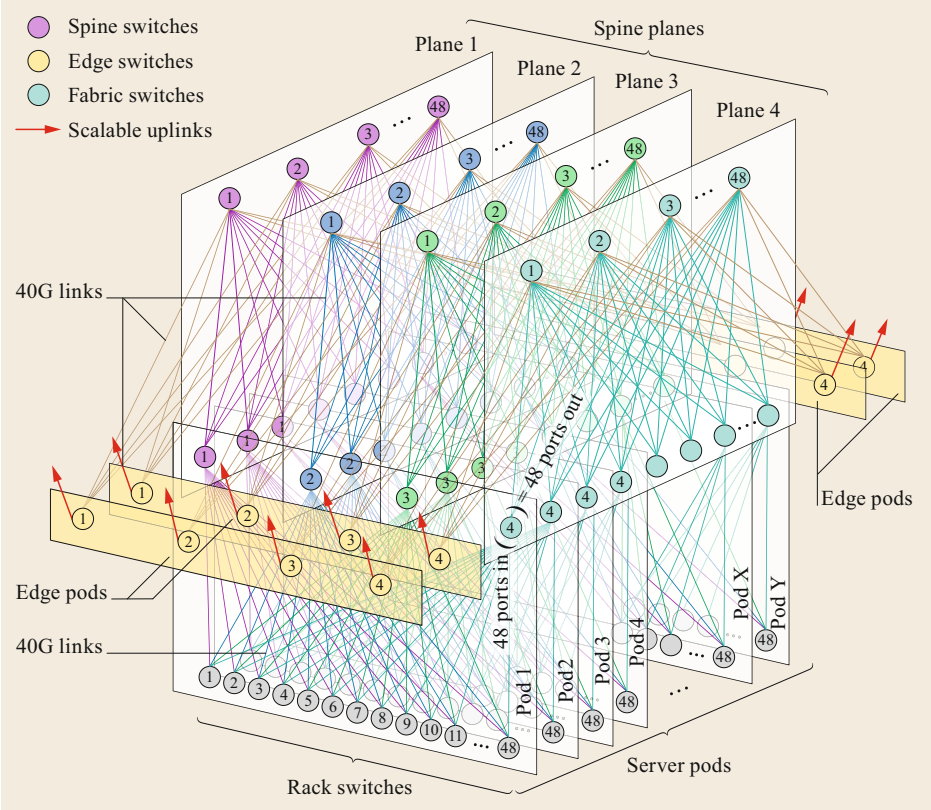


Fig. 24.2 Data center network topology from Facebook (after [24.2])

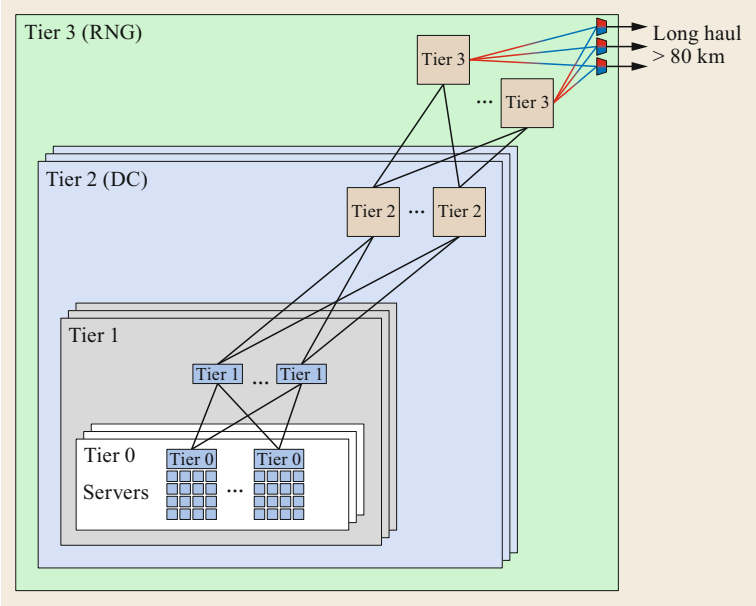


Fig. 24.3 Microsoft data center network architecture. Ethernet switches are labeled tier 0 through tier 3. A whole data center (DC) is represented by the tier 2 level. The tier 3 level is defined by the RNG (regional network gateway)

Table 24.1 Typical connectivity distance in a Microsoft data center

Switch connections	Typical distance (m)	Technology
Server–tier 0	0–3	DAC
Tier 0–tier 1	1–30	AOC
Tier 1–tier 2	Up to 500	SMF
Tier 2–tier 3	Up to 70 000	SMF

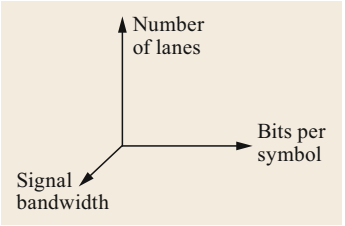


Fig. 24.4 Directions for increasing the data rate of an optical interconnect

MMF links being limited to 150 m or less, while the slowest MMF links can extend to 2 km. For MMF links, InGaAsP-based, directly-modulated, vertical cavity surface emitting lasers (VCSELs) operating at about 850 nm are used as transmitters [24.3]. VCSELs operate on multiple transverse modes, separated by about 0.1 nm. The $\approx 10\,\mu$m emitting diameter of a VCSEL easily couples to the $\approx 50\,\mu$m MMF core, and this is essential to their relative low cost when incorporated in transceivers. The origin of the MMF distance × bandwidth limitations are in modal dispersion and chromatic dispersion. In the past 20 years, significant improvements in the MMF distance × bandwidth product have been enabled by improved fiber design and control of the laser launch conditions, but further major improvements are unlikely. If not for its physical distance × bandwidth limitations, VCSEL-based MMF would remain the low-cost, low-power consumption winner for optical connectivity in the data center. MMF and VCSEL technologies continue to be embedded in relatively short-reach AOCs.

For intensity-modulation direct-detection (IM/DD) transmission systems in single-mode fiber (SMF), the most fundamental limitation to the distance × bandwidth product is chromatic dispersion-induced fading at high modulation frequencies. Operation near ≈ 1310 nm, the minimum dispersion wavelength of standard SMF, maximizes the distance × bandwidth product. For pure intensity modulation (meaning an absence of phase or frequency modulation, i.e., no chirp), the frequency, $f_{3\,\mathrm{dB}}$, at which the fiber transfer function drops by 3 dB, is given by [24.4]

$$f_{3\,\mathrm{dB}} = \frac{1}{\lambda}\sqrt{\frac{c}{3LD}}\,,$$

where L is fiber distance, D is chromatic dispersion, λ is the laser wavelength, and c is the speed of light. For a small chromatic dispersion, 1 ps/(nm km), at a wavelength near 1310 nm and a distance of 10 km, $f_{3\,\mathrm{dB}} = 132$ GHz, which limits the modulation bandwidth. Information theory indicates that the theoretical maximum symbols per unit time transmitted is twice the modulation bandwidth (Nyquist frequency). Other effects such as such as chromatic dispersion-induced signal distortion may further limit modulation bandwidth. SMF, therefore, has been and will continue to be used for tier 1–tier 2 connections in large-scale data centers, and increasingly in smaller-scale data centers as bandwidth needs continue to grow.

One theme of this chapter is the rapidly increasing data-rate requirements of data center optical interconnects (Fig. 24.4). In order to achieve a desired data rate, tradeoffs are made based on technical maturity, cost, and complexity. The three main directions for increasing the overall data rate of an optical link are illustrated below. The simplest, lowest risk approach is always to segment into parallel lanes where the sum of the data rates of the individual lanes equal the net data rate. In general, the lower data rates come with lower technical risk. As optical component costs dominate transceiver costs, the ultimate low-cost solution usually involves minimizing the number of lanes, with higher-speed laser transmitter technology.

SMF has seen significant growth in data centers since it has not yet reached its physical distance × bandwidth product limit, and, therefore, can support higher data rates at the same reach. Data centers have two options for deploying single-mode fiber: duplex or parallel. Duplex (two-fiber) SMF cable is lower cost than parallel single-mode (PSM) but has required the use of wavelength division multiplexing (WDM) to support greater bandwidths per fiber. PSM typically supports a single wavelength on each fiber. For example, to transmit 100 Gb Ethernet over duplex fiber pairs, one could use an optical module that transmits four different 25 Gb/s wavelengths on a single fiber; this is known in the industry as 100G-CWDM4 (Fig. 24.5a). An alternative approach is to use an optical module that uses four different fibers carrying 25 Gb/s wavelengths; this is known in the industry as 100G-PSM4 (Fig. 24.5b). The biggest differences between CWDM4 and PSM4 is that CWDM4 requires each wavelength to be different and a multiplexor to merge the wavelengths, whereas PSM4 can use the same wavelength on four different fibers.

24.1.3 Total Cost of Ownership

Total cost of ownership has two elements: capital expenditure (CapEx) and operational expenditure (OpEx).

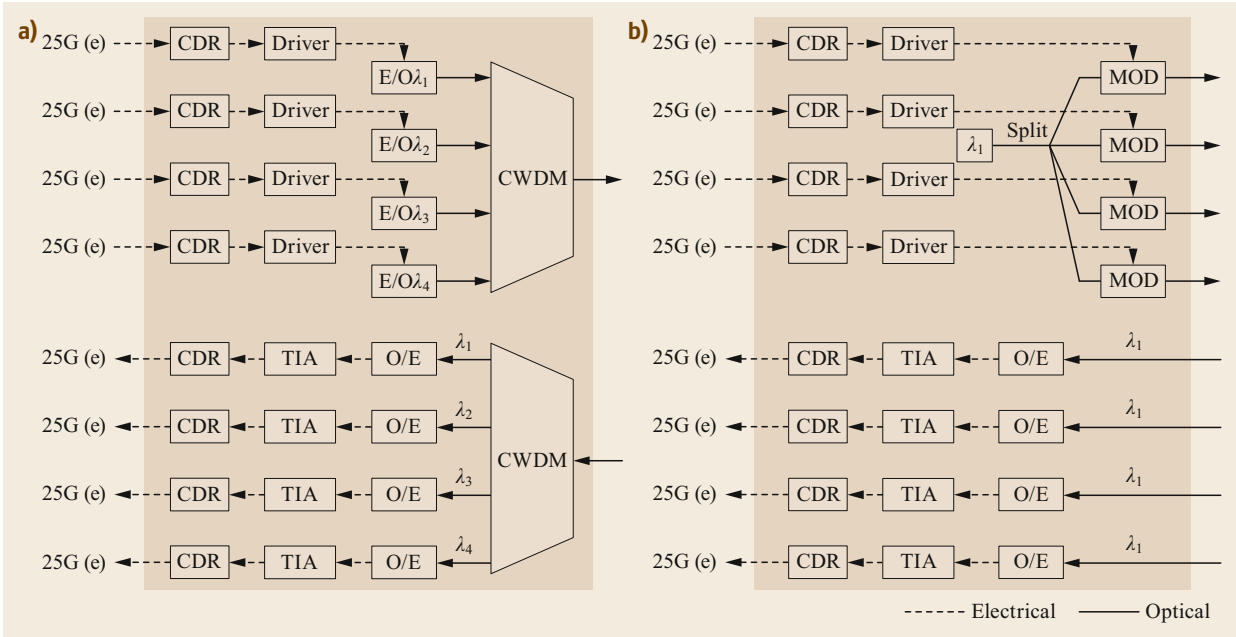


Fig. 24.5 **(a)**70 Schematic diagram of a CWDM4 optical transceiver using two fibers for bi-directional transmission. **(b)** Schematic diagram of a PSM4 optical transceiver using eight fibers for bi-directional transmission. Each transceiver is fed by four independent 25 Gb/s electrical signals for an aggregate data rate of 100 Gb/s. Optical connections are *solid lines*. Electrical connections are *dashed lines*. (CDR = clock and data recovery; TIA = transimpedance amplifier; E/O = electrical to optical signal conversion, typically one of the options outlined in the third section of this chapter; O/E = optical to electric signal conversion, typically a PIN photodiode; CWDM = coarse wavelength division multiplexor; MOD = optical intensity modulator)

It is fairly easy to use CapEx to make a comparison. In the case of a fiber optic network, it is the cost of the equipment, the cost of the module, and the cost of the infrastructure relative to the period of amortization. Switches, routers, and network adapters typically have the same amortization period as the optical module. The fiber infrastructure may have a different amortization period. In the case of duplex SMF versus PSM, the duplex fiber is two strands of fiber and the PSM is typically 12 strands of fiber; therefore, the cost of PSM is up to six times greater than duplex SMF. The MPO connector used for PSM is also costlier than the LC (Lucent connector) duplex connector. This would seem to imply that duplex SMF would be the more cost-effective option, but when the CapEx of the fiber, connectors, and installation is amortized over a longer period than the optical modules, the modules become a bigger influence on the overall CapEx.

Figure 24.6 shows a comparison of the cost of the fiber and modules relative to their period of amortization [24.5]. The cost of PSM fiber is greater than traditional duplex SMF, but PSM4 optics have lower costs than CWDM4 optics. If the cost of the fiber and optical module is combined and amortized over the life of the optical module or the life of the fiber (which is much longer than the life of the module), the cost of PSM fiber and PSM4 optics become more affordable as time progresses.

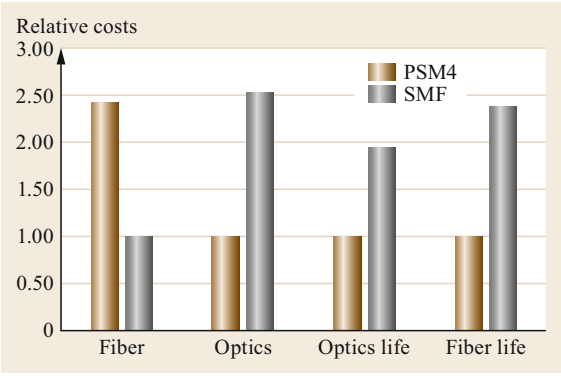


Fig. 24.6 Relative costs for deploying PSM4 versus CWDM4 (SMF) in a Microsoft data center

Unlike CapEx, OpEx is harder to compare. A substantial portion of OpEx is the cost of power. While this is usually the overlying driver in the OpEx equation, the costs of installation and maintenance are also critical elements. Looking at installation and maintenance, one aspect to understand is that many switches installed

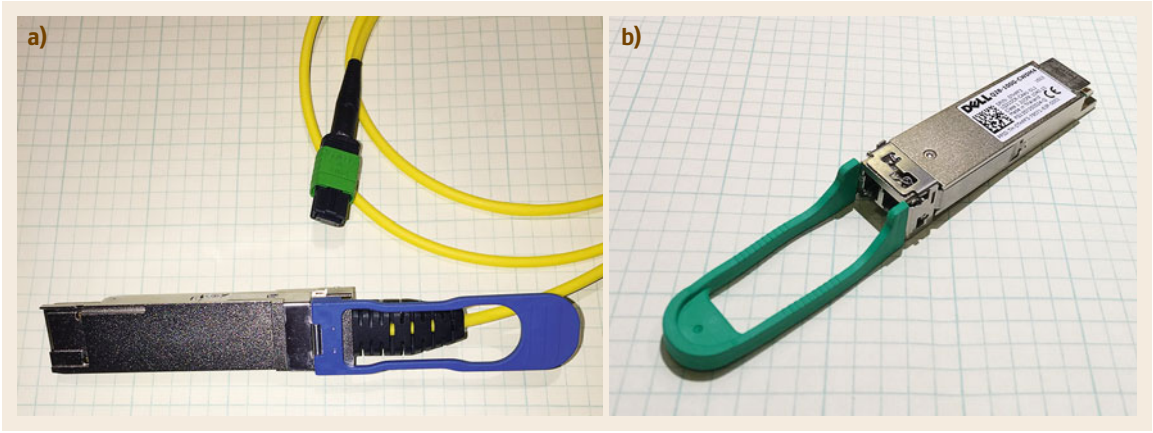


Fig. 24.7a,b QSFP [24.6] (quad small-form pluggable) PSM4 module with fiber pigtail and MPO (multifiber) connector **(a)**; duplex QSFP module with two LC connector receptacles **(b)**

in data centers are typically not populated with optical modules. A technician must unpack and install these modules in the field. Each fiber connection must be cleaned before mating the connectors. When Microsoft analyzed the time involved in cleaning the connectors, they made the choice to use pigtailed optical modules (Fig. 24.7a). These modules, unlike traditional optical modules (Fig. 24.7b), do not have an optical connector on the module. Instead of using fiber-optic patch cords to connect from the module to the patch panel and having to clean two connectors, Microsoft opted to have the modules come with the patch cord permanently attached to the module at the factory. This results in only one connector, instead of two, to be cleaned.

24.2 SMF Transmitter Toolkit

Optical transmitter cost is a major contributor to CapEx. Optical transmitter power consumption is a major contributor to OpEx. The technical approaches available for SMF optical transceivers outlined below are used to trade off performance, cost, and power consumption.

The transmitter toolkit for commercial optical transmitters for SMF is diverse. SMF light sources tend to operate at either 1310 nm for minimum chromatic dispersion or 1550 nm for minimum fiber loss. As transmitter speed increases, chromatic dispersion is a bigger problem than loss, so most light sources operate in the O-band, at around 1310 nm. (Semiconductor lasers operating at 1310 nm also have better electrical efficiency compared to 1550 nm lasers.)

24.2.1 Directly-Modulated DFB Laser

The most cost-effective and lowest-power consumption SMF transmitter solution has most often been the uncooled, directly-modulated distributed feedback (DFB) laser, sometimes simply referred to as a directly-modulated laser (DML). Simple modulation of the laser's bias current generates similarly modulated light intensity (Fig. 24.8). DFB lasers are characterized by their chirp, which is the degree of wavelength (or optical frequency) modulation accompanying intensity modulation. The fiber's chromatic dispersion interacts with laser chirp to distort the modulated waveform as it propagates through the fiber. Operation near 1310 nm minimizes this effect.

When necessary, DFB lasers that can be operated without temperature control over the temperature range of -40 to $+85\,^{\circ}\mathrm{C}$ are available. When operated uncooled, their wavelength varies by about $0.1\,\mathrm{nm}/^{\circ}\mathrm{C}$. In a typical data center environment, an optical transmitter may be required to operate over a temperature range of $+5$ to $+75\,^{\circ}\mathrm{C}$, which leads to a ± 3.5 nm wavelength stability for an uncooled DFB laser.

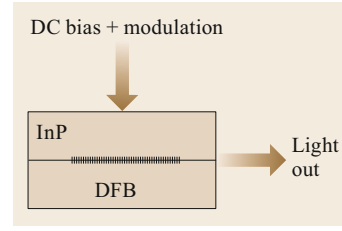


Fig. 24.8 Diagram of a directly-modulated DFB laser

DFB lasers are bandwidth limited by their relaxation oscillations, which lead to an increasingly nonlinear modulation transfer function as frequencies approach the relaxation oscillation frequency [24.7]. The problem is especially acute as the instantaneous laser-drive current approaches the laser threshold current. This has the effect of decreasing the useable extinction ratio when pushing the bandwidth limits. The relaxation oscillation frequency also decreases when the laser operates at the high end of its temperature range. The relaxation oscillation frequency is an intrinsic property of the DFB material and device design.

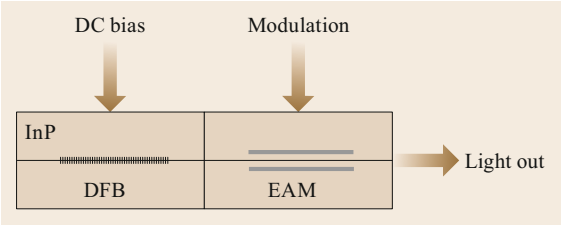


Fig. 24.9 Diagram of an externally-modulated DFB laser. The electro-absorptive modulator (EAM) is external to the DFB laser but fabricated on the same InP substrate

24.2.2 Externally-Modulated DFB Laser

The alternative to direct modulation of DFB laser is to let the DFB continuously generate light and modulate its output with what is essentially a very fast shutter. There are two common ways to create an externally-modulated laser (EML): the electro-absorptive modulator (EAM) and the Mach–Zehnder modulator (MZM). Thus far, high-speed EMLs were commercially available before comparable speed DMLs, primarily due to the longer technology development cycle needed to improve DFB device design for higher-speed operation.

Integrated Electro-Absorptive Modulator/ DFB Lasers

Integrated DFB lasers with EAMs can be fabricated on a common InP substrate (Fig. 24.9). Although the EAM retains residual chirp, it is over an order of magnitude smaller than the DFB DML for equivalent modulation. The EAM bandwidth is limited by resistance × capacitance (RC) time constant and residual inductance [24.8], which can be overcome by appropriate packaging [24.9]. Today, commercial EAMs are available with 3 dB bandwidths up to 50 GHz [24.10]. The EAM transfer function does not have the relaxation oscillation of the DFB DML and is much better behaved at high frequencies. The integrated DFB/EAM modulation transfer function is highly temperature dependent and must often be tightly temperature controlled for optimal operation. In the near future, DFB/EAMs will become available in cooler-less formats [24.11].

Silicon Photonics

A potential low-cost, low-power external modulator is realized through the use of silicon photonics (SiPho).

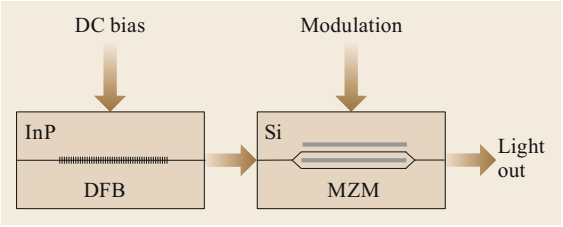


Fig. 24.10 Diagram of an externally-modulated DFB laser. The Mach–Zehnder modulator (MZM) is external to the DFB laser and fabricated in a separate silicon photonic (SiPho) substrate

MZMs can be fabricated on silicon substrates for compact size and high bandwidth. Compared to the EAM, the SiPho MZM is insensitive to changes in temperature [24.12]. The low $V_\pi L$ means that the modulator driver operates with exceptionally low power consumption. One still, however, needs a continuous wave (cw) InP-based laser source or sources coupled to the waveguide (Fig. 24.10). The cost of SiPho benefits from high-volume complementary metal–oxide–semiconductor (CMOS) fabrication; however, packaging remains an issue due to the large mode mismatch between the SiPho waveguide and the cw DFB laser [24.13].

An additional advantage of the SiPho approach is that the WDM multiplexor can be integrated onto the silicon substrate.

Advancements continue to be made with DMLs and EMLs. There is work in the industry to develop EAMs that can be fabricated on a silicon substrate [24.14], and MZMs that are fabricated on an InP substrate [24.15]. The characteristics of these light sources will continue to influence optical transmission specifications.

24.3 Optical Specifications

The IEEE 802.3 Ethernet standards organization has been prevalent in the development of optical (and electrical) data center interconnect specifications, but it is not the only organization to develop optical interconnect specifications. Sometimes, it takes competing forces to move interconnect technology forward to yield new and relevant solutions.

24.3.1 IEEE 802.3 Ethernet

Common interface standards are essential to communication and interoperation between network elements. As the IT industry has grown, different firms have specialized in the design, manufacture, and sale of various network elements. Facilitating this growth, the IEEE Standards Association has been proactive in developing timely and appropriate network interconnect standards under the IEEE 802.3™ Ethernet brand. A new standard must show broad market potential, be compatible with pre-existing Ethernet standards, have a distinct identity, and be technically and economically feasible. In developing a new standard, the relevant IEEE 802.3 task force must make tradeoffs between market needs, technical aggressiveness, and time-to-market.

Table 24.2 lists all the optical transceiver standards approved or in development by IEEE 802.3 [24.16]. Some approved standards make little market impact, either because they are technically obsolete by the time the standard is approved, or the market demand for such an optical transceiver has diminished. The IEEE Ethernet standardization process remains essential, if a transceiver is to be successful in the market. Directly or indirectly, it forms a critical foundation for the industry to build upon.

24.3.2 MSAs and Consortia

Multisource agreements (MSAs) and consortia often fill the gaps when an IEEE 802.3 either cannot be created in a timely fashion or there is no sufficient agreement on the technical solution or market demand. Module form factor and packaging MSAs have been around for some time, as manufacturers grouped together to make sure mechanical and electrical connections between optical modules and their hosts were interoperable and

Table 24.2 IEEE 802.3 Data center Ethernet Optical Standards

Year ratified	Ethernet speed	Designation	Optical data rate	Wavelength (nm)	Fiber count (F)	Maximum reach	
						MMF (km)	SMF (km)
1987	10 MbE	FOIRL	12.5 M/λ	850	2	1	–
1993	10 MbE	10BASE-FB	12.5 M/λ	850	2	2	–
1993	10 MbE	10BASE-FL	12.5 M/λ	850	2	2	–
1995	100 MbE	100BASE-FX	125 M/λ	1310	2	2	–
1998	1 GbE	1000BASE-SX	1.25 G/λ	850	2	0.550	–
1998	1 GbE	1000BASE-LX	1.25 G/λ	1310	2	0.550	2
2002	10 GbE	10GBASE-SR	10 G/λ	850	2	0.400	–
2002	10 GbE	10GBASE-LR	10 G/λ	1310	2	–	10
2002	10 GbE	10GBASE-ER	10 G/λ	1550	2	–	40
2002	10 GbE	10GBASE-LX4	3.125 G/λ	10G-CWDM	2	0.300	10
2005	100 MbE	100BASE-BX10	125 M/λ	1310/1550	1	–	10
2005	100 MbE	100BASE-LX10	125 M/λ	1310	2	–	10
2005	1 GbE	1000BASE-BX10	1.25 G/λ	1310/1550	1	–	10
2005	1 GbE	1000BASE-LX10	1.25 G/λ	1310	2	–	10
2006	10 GbE	10GBASE-LRM	10 G/λ	1310	2	0.220	–
2010	100 GbE	100GBASE-SR10	10 G/λ	850	20	0.150	–
2010	40 GbE	40GBASE-SR4	10 G/λ	850	8	0.150	–
2010	40 GbE	40GBASE-LR4	10 G/λ	CWDM	2	–	10
2010	40 GbE	40GBASE-ER4	10 G/λ	CWDM	2	–	40
2010	100 GbE	100GBASE-LR4	25 G/λ	LAN-WDM	2	–	10
2010	100 GbE	100GBASE-ER4	25 G/λ	LAN-WDM	2	–	40
2011	40 GbE	40GBASE-FR	40 G/λ	1550	2	–	2
2015	100 GbE	100GBASE-SR4	25 G/λ	850	8	0.100	–
2016	25 GbE	25GBASE-SR	25 G/λ	850	2	0.100	–

Table 24.2 (continued)

Year ratified	Ethernet speed	Designation	Optical data rate	Wavelength (nm)	Fiber count (F)	Maximum reach	
						MMF (km)	SMF (km)
2017	25 GbE	25GBASE-LR	25 G/λ	1310	2	–	10
2017	25 GbE	25GBASE-ER	25 G/λ	1310	2	–	40
2017	200 GbE	200GBASE-DR4	50 G/λ	1310	8	–	0.500
2017	200 GbE	200GBASE-FR4	50 G/λ	CWDM	2	–	2
2017	200 GbE	200GBASE-LR4	50 G/λ	LAN-WDM	2	–	10
2017	400 GbE	400GBASE-FR8	50 G/λ	extLAN-WDM	2	–	2
2017	400 GbE	400GBASE-LR8	50 G/λ	extLAN-WDM	2	–	10
2017	400 GbE	400GBASE-SR16	25 G/λ	850	32	0.100	–
2017	400 GbE	400GBASE-DR4	100 G/λ	1310	8	–	0.500
2018	50 GbE	50GBASE-SR	50 G/λ	850	2	0.100	–
2018	100 GbE	100GBASE-SR2	50 G/λ	850	4	0.100	–
2018	200 GbE	200GBASE-SR4	50 G/λ	850	8	0.100	–
2018	50 GbE	50GBASE-FR	50 G/λ	1310	2	–	2
2018	50 GbE	50GBASE-LR	50 G/λ	1310	2	–	10
2018	100 GbE	100GBASE-DR	100 G/λ	1310	2	–	0.500
2020	400 GbE	400BASE-SR8	50 G/λ	850	16	0.100	–
2020	400 GbE	400BASE-SR4.2	50 G/λ	860/910	8	0.150	–

M/λ = Mb/s per wavelength. G/λ = Gb/s per wavelength. 10G-CWDM = (1275, 1300, 1325, 1350 nm); CWDM = coarse wavelength-division multiplexing (grid) = (1271, 1291, 1311, 1331 nm); LAN-WDM = local area network – wavelength-division multiplexing (grid) = (1295.56, 1300.05, 1304.58, 1309.14 nm); ext(ended) LAN-WDM = (1273.54, 1277.89, 1282.26, 1286.66, 1295.56, 1300.05, 1304.58, 1309.14 nm)

interchangeable. MSAs have also been formed to define optical specifications that satisfy a specific market need.

The *Optical Internetworking Forum* (OIF) [24.17] continues to create standards that impact the work of the IEEE. The OIF's common electrical interfaces (CEI) has formed the basis for recent Ethernet high-speed electrical interfaces, and the OIF's standards for optical transport in the wide-area-network (WAN) are increasingly relevant to inter-data center interconnects and may eventually impact intra-data center interconnect.

24.3.3 Proprietary Solutions

IEEE 802.3 defines the logical and physical electronic interface to a transceiver module called an attachment unit interface (AUI). MSAs have defined the mechanical, power, and telemetry interfaces, as well as the form factor for the optical transceiver. It is possible to utilize those specifications to define a proprietary optical specification.

Tables 24.3 and 24.4 list 100 Gb and 400 Gb Ethernet optical transceivers that were defined or are being defined by MSAs, industry consensus, or proprietary organizations.

24.4 25 G/λ, 50 G/λ, and 100 G/λ Optical Technologies

Tables 24.2, 24.3 and 24.4 give IEEE and non-IEEE definitions for various bandwidth optical links. By noting the optical technology used for any overall data rate, one infers how the total bandwidth is partitioned among the optical lanes.

At the time of this writing (2018), 100 Gb Ethernet is the dominant rate for intradata center optical links, and the clear majority of these links are made up of four optical lanes, each modulated with 25 Gb/s nonreturn to zero (NRZ) signaling. Initial implementations used 100GBASE-LR4, which was based on four DFB/EAM lasers operating on the LAN-WDM grid and required active temperature control to stabilize wavelength and to insure proper operation of the DFB/EAM. At the time of 100GBASE-LR4 standardization, DFB/EAMs were the only proven solution with sufficient modulation bandwidth. As cost (OpEx + CapEx) became a dominant issue in deployment, the 100GBASE-LR4 specification needed to be augmented by new lower-cost solutions. One key path to

Table 24.3 Non-IEEE 802.3 100 Gb Ethernet data center optical interconnects

Designation	Organization	Fiber	Wavelength (nm)	Fiber count (F)	Max reach (km)	Remarks
ESR4	Consensus	MMF	850	8	0.300	Extended-range SR4
SWDM4	MSA [24.18]	MMF	SWDM	2	0.100	
BI-DI	Proprietary [24.19]	MMF	860/910	2	0.100	50 G/λ technology
PSM4	MSA [24.20]	SMF	1310	8	0.500	
CWDM4	MSA [24.21]	SMF	CWDM	2	2	
CLR4	MSA [24.22]	SMF	CWDM	2	2	No FEC required
CWDM4/OCP	MSA [24.23]	SMF	CWDM	2	0.500	Temperature range (15–45 °C)
4WDM-10	MSA [24.24]	SMF	CWDM	2	10	
4WDM-20	MSA [24.24]	SMF	LAN-WDM	2	20	
4WDM-40	MSA [24.24]	SMF	LAN-WDM	2	40	
ER4-lite	Consensus	SMF	LAN-WDM	2	≈ 35	
ColorZ	Proprietary [24.25]	SMF	DWDM	2	80	Metro data center interconnect; 50 G/λ technology
100GBASE-FR	MSA [24.26]	SMF	1310	2	2	100 G/λ technology
100GBASE-LR	MSA [24.26]	SMF	1310	2	10	100 G/λ technology

SWDM = short-wavelength-division multiplexing = (850, 880, 910, 940 nm). FEC = forward-error correction

Table 24.4 Non-IEEE 802.3 400 Gb Ethernet data center optical interconnects

Designation	Organization	Fiber	Wavelength (nm)	Fiber count (F)	Max reach (km)	Remarks
400GBASE-FR4	MSA [24.26]	SMF	CWDM	2	2	100 G/λ technology
400GBASE-LR4	MSA [24.26]	SMF	CWDM	2	10	100 G/λ technology
400G ZR	OIF [24.17]	SMF	DWDM	2	120	Metro data center interconnect; 400 G/λ coherent technology

lower cost was to use uncooled DML DFBs, as the industry developed lasers with improved modulation bandwidth [24.27]. DML DFBs intrinsically have lower parts cost than the more complex DFB/EAM, and they use less electrical power through more efficient modulation and the lack of thermo-electric cooling. Over transceiver operating ranges (typically +5 to +75 °C), their decreased wavelength stability required the use of the wider-window CWDM wavelength grid. This led to the CWDM4 MSA, which further lowered costs by reducing the link distance specification from 10 to 2 km. The PSM4 MSA led to even lower optical parts costs by further reducing the link distance specification to 500 m and eliminating the wavelength multiplexor (as well as its associated optical loss). Advantageously, as illustrated in Fig. 24.5b, with the PSM4 SiPho architecture a single laser can feed four MZMs [24.28], further lowering the optical cost. Today, PSM4 and CWDM4 are the dominant 100 Gb Ethernet links used in hyperscale data centers.

The 50 G/λ optics using four-level pulse-amplitude modulation (PAM4) – which has the same Baud rate as 25 G/λ optics except there are two bits per symbol compared to NRZ's one – has more stringent optical requirements (Fig. 24.11). Linear (as opposed to limiting) drivers and transimpedance amplifiers are needed in the transceiver. Compared to NRZ, 4.8 dB ($10\log_{10}(3)$) more signal-to-noise ratio is required to achieve an equivalent bit error ratio (BER), thus, noise sources such as laser RIN (relative intensity noise) and multipath interference (MPI) become increasingly dominant alongside the distance × bandwidth product limit. As with 25 G/λ, most early work with this technology has used externally modulated lasers – mainly DFB/EAMs.

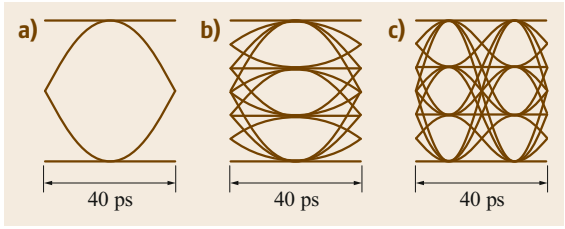


Fig. 24.11a–c Eye diagrams for (a) 25 GBd NRZ signaling (25 Gb/s data rate), (b) 25 GBd PAM4 signaling (50 Gb/s data rate), and (c) 50 GBd PAM4 signaling (100 Gb/s data rate) with identical extinction ratios. Note the eye shrinkage as data rates increase

Recent work with DMLs, however, is promising [24.29, 30].

The 100 G/λ optics beginning to emerge in the industry also use PAM4, but with a symbol rate of 50 GBd, which is twice of that used in 25 G/λ and 50 G/λ technology. Multilevel modulation at these higher data rates are especially sensitive to the frequency response limitations of the optical device and its packaging and to nonlinearities in the transmit and receive electronics and in the optical device, in addition to the above-cited impairments. At these higher data rates, digital signal processing and forward-error correction (FEC) within the transceiver are essential for successful operation [24.31].

As the general trend for reducing optical cost has been to minimize the number of optical lanes by using the highest speed optics, 100 G/λ technology holds the promise of leading to the lowest cost 100 and 400 Gb Ethernet interconnect in the immediate future. Since the IEEE (Table 24.2) defines only 500 m links using 100 G/λ technology (100GBASE-DR and 400GBASE-DR4), a new MSA (100G Lambda MSA) has recently formed to specify longer-range 2 and 10 km 100 and 400 Gb Ethernet links. As of 2020, the IEEE has begun work to standardize these links. We expect that as data center bandwidth needs grow, 100 Gb Ethernet links will transition to 400 Gb links using 100 G/λ optics.

24.5 Optical Packaging

One area that is impacted by the increase in connectivity speeds is the package for the optical module. While traditional client optics use digital interfaces, line-side optics have often used analog interfaces until the power declines enough in the digital signal processor (DSP) chip to permit it to be integrated into the optical module package. Each successive generation of optics has developed either a new form factor or an improvement on an existing form factor.

If an end user wishes to use a past generation of optics in a new switch, then having a form factor that permits migration from one generation to the next has value to that customer. On the other hand, if using optics from different speed generations in the same socket is not a requirement, then the customer may be more likely to select a module optimized for their application.

As the speed of the electrical interfaces between the chip and the optical module progress from today's 25 GBd NRZ (25 Gb/s) to 50 GBd PAM4 (100 Gb/s), there will be additional impact associated with the interface due to power, latency, and materials required. It will become more difficult mitigate those impacts without also analyzing where the optical module is placed. While optics on the faceplate are common today, it is only a question of time until the optics move closer and closer to the chips that are performing the networking.

24.5.1 The Pluggable Module Paradigm

The QSFP (quad small form-factor pluggable) optical transceiver modules illustrated in Fig. 24.7 represent the present state-of-the-art for 100 Gb Ethernet optical packaging. Here, the optics are separate from the networking equipment, which has built-in QSFP receptacles. There are several advantages to this architecture: (1) the end-user can install additional transceivers as bandwidth requirements increase (*pay as you grow*); (2) replacing field-failed modules is straightforward; (3) a mix of modules and cables (for example, short-reach and long-reach) can be used in a single piece of equipment. For a hyperscale data center, however, these advantages may not be great. For example, hyperscale users usually install equipment fully populated with transceivers on day 1. Additionally, there are operational issues when thousands of transceivers must be un-boxed, plugged in, and tested. Receiving networking equipment preprovisioned with transceivers would be optimal.

There are technical reasons why the pluggable module paradigm may fail to scale-up as Ethernet switching speeds continue to increase. Consider the typical form factor for an Ethernet switch – the 1 rack unit (1 RU) *pizza box*. The front panel (48.3 cm wide × 4.45 cm high) of the rack-mounted pizza box can be covered by up to 36 QSFP receptacles with the remaining area dedicated to air vents. Through the back panel power supplies and fans are accessed. The switching application specific integrated circuit (ASIC) communicates to the transceiver through a metallic trace on the printed circuit board (PCB), which is increasingly lossy at high frequencies. In some cases, the length and loss of the trace make a retimer between the ASIC and the transceiver module necessary (Fig. 24.12a). As signaling Baud rates and/or modulation order increase, the high-frequency loss requirements for traces become more demanding. These demands can be met by using higher-power electrical transceivers (also known as SerDes (serializer/deserializer)) on the switch ASIC and/or the retimer, or by using lower-loss materials to bridge the distance. Either way, higher speeds or

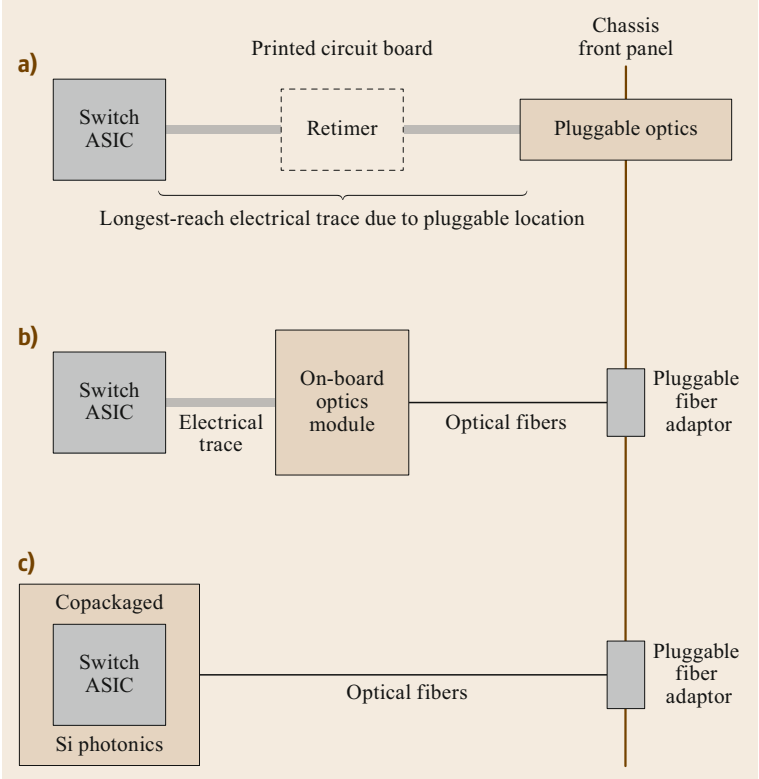


Fig. 24.12a–c Ethernet switching ASIC to chassis front panel electro-optical link for the cases of **(a)** pluggable transceiver module, **(b)** on-board-optical transceiver, and **(c)** optical transceivers copackaged with switching ASIC

modulations will impact the overall power budget and architecture of the switching equipment.

As the data rate of the optical transceiver module increases, so does its power consumption. While a 100 Gb Ethernet QSFP optical module typically dissipates $\leq$ 3.5 W, it is anticipated that a 400 Gb Ethernet optical module will dissipate $\leq$ 12 W independently of the optical module form factor. (Both next-generation, 400 Gb pluggable module form factors, quad small form-factor pluggable double density (QSFP-DD) [24.32] and octal small form pluggable (OSFP) [24.33], can accommodate up to 32 receptacles on a front panel.) While this represents a slight improvement in W/(Gb/s), the total power dissipated at the front panel increases by nearly a factor of 4. Also note that these switch module receptacles, mounted on the front panel, are not in an optimal location for thermal management.

Today, a typical 3.2 Tb/s Ethernet switch in a *pizza box* has 32 100 Gb QSFP ports on its front panel. The next generation of 12.8 Tb/s Ethernet switch will have 32 400 Gb QSFP-DD or OSFP ports on their front panel, as well as twice as many electrical traces within, each facilitating twice the (electrical) data rate (50 Gb/s PAM4) compared to the previous generation. Such switches are being designed as this is written, but they are pushing up against the thermal limitations for a pizza box.

We anticipate that the next generation of Ethernet switch ASICs, operating at 25.6 Tb/s, will utilize 100 Gb/s (PAM4) SerDes, and feed up to 32 800 Gb/s optical modules (using $8\times100\,\mathrm{G}/\lambda$ technology). Fitting into the *pizza box* using the existing pluggable module paradigm will be extremely difficult, if not impossible. A new approach to optical packaging is necessary. In the remainder of this section, we consider two innovative approaches – on-board-optics (Fig. 24.12b) and copackaged optics (Fig. 24.12c). Both architectures improve the power and thermal situations by moving the optics closer to the switching ASIC.

24.5.2 On-Board-Optics

In 2015, the *Consortium for On-Board Optics (COBO)* [24.34] was formed to challenge some of the shortcomings of the pluggable-module paradigm by moving the optical modules from receptacle on the front panel to be mounted closer to the switch ASIC on the PCB. This has several immediate positive effects. First, the electrical traces are shorter, thereby eliminating the need for retimers and permitting the switch

ASIC SerDes to operate at a lower power, since the electrical loss is reduced. Also, the number of modules is no longer limited by the front panel area. The on-board module is connected to a pluggable fiber connector bulk-head adaptor on the front panel through a multiple parallel fiber cable (Fig. 24.12b). This makes available more front panel area for air flow. Finally, the board-mounted optics open the design space for more effective thermal solutions. Each on-board optical module can now have a large and effective finned heat sink, for example. Although on-board optics have been deployed before in proprietary designs for networking equipment, COBO's goal was to create a common specification for industry-wide use.

COBO released version 1.0 of an on-board optical module specification in 2018. The specification addressed the (PCB) host-to-module electrical interface, the host-to-module electro-mechanical connector and form factor interface, as well as thermal and optical connectivity. The electrical interface specification is based on the IEEE Ethernet 400GAUI-8 [24.16] chip-to-module interface (eight lanes of 50 Gb/s PAM4). The electrical specification supports modules using the 400 Gb Ethernet optics defined in Tables 24.2 and 24.4 through an eight-lane interface and a *dual* 400 Gb Ethernet module using a 16-lane interface. The net capacity of 800 (Gb/s)/module allows for fewer modules, saving valuable board space. (Fewer modules also simplify PCB traces and the fiber cables to the front panel.) The specification was created with the next generation of chip-to-module electrical interfaces (100 Gb/s PAM4) presently being standardized in IEEE Ethernet [24.35] and OIF in mind. As a result, COBO testing of the specified on-board optical module indicates that it will support 800 Gb and 1.6 Tb modules when the IEEE and OIF finish their work.

The high-speed electrical connector allows for *snap-in* insertion and removal of the modules. This differs from previous proprietary on-board designs in which the modules are mounted directly onto the PCB. This simplifies the construction of the on-board optics-based switches and makes for straightforward rework (in case the fiber cable is broken, for example). Advantageously for the hyperscale user, Ethernet switches can be delivered fully populated with tested optical modules for quick installation within the data center.

COBO has designated three classes of module that differ based on size and power dissipation capability. The smallest, class A, is designated for VCSEL-based MMF applications. Class B is designated for SMF applications, such as 400GBASE-DR4 or 400GBASE-FR4 (Tables 24.2 and 24.4). Class C is designated for the highest power consumption applications, such as the DSP-intensive coherent optical 400G ZR (Table 24.4) module.

24.5.3 Copackaged Optics

The ultimate in moving optics closer to the switching ASIC is the assembly of both at the chip level in the same package (Fig. 24.12c). This minimizes the length of the electrical signaling path between the switch ASIC and the optics, significantly reducing the power that would be consumed driving signals from the ASIC to the front panel or to an on-board module. In the ideal implementation, the need for signal retiming/CDR functionality in the optical subsystem may be eliminated, which is a further benefit to overall power dissipation.

Let us now consider the *next*-next-generation Ethernet switch ASIC – one with 25.6 Tb/s of total bandwidth. If we assume that we continue to use 100 G/λ optics (and use a 100 Gb/s SerDes for the chip to optics link), we end up with a total of 256 electrical-to-optical and optical-to-electrical conversions near the switch ASIC. SiPho is the most appropriate technology to meet this requirement due to its low power consumption, its amenability to small feature sizes, and its copackaging compatibility.

As this chapter is being written, the key technologies for copackaging are being developed. Several obstacles, both technical and commercial, have yet to be overcome. Although copackaging reduces overall system power consumption, it nevertheless concentrates power dissipation to a relatively small region. Solutions include moving beyond reliance on air-flow cooling to consideration of heat pipes and close-loop liquid cooling. Another challenge is fiber attachment/management. While direct attachment of fibers to the SiPho chips may now be straightforward, management of $(2 \times 256 =)$ 512 fibers hanging off a single package pose significant manufacturing problems. Innovative technologies for fiber attachment and connectorization require development. Finally, a successful commercial model for the convergence of optical and ASIC technologies must emerge. Traditionally, separate and distinct vendors provided these elements. Success may require not only a standardized ASIC – optical interface, but also the emergence of a commercial ecosystem including copackaging specialists to integrate the overall package.

24.6 The Future

As cloud data centers continue to grow, they will continue to place new requirements on optical transceivers. As the ink dries on the specification for 400 Gb Ethernet, there are already discussions on what speed is next: 800 Gb, 1 or 1.6 Tb? Many of the necessary elements to create these high-bandwidth connections are already in existence or in development. As shown in Fig. 24.4, choices can be made to increase the speed of the connection. There are tradeoffs to be made based upon the fiber available and the cost implications of the various technical choices. Looking past 100 G/λ, one may question whether or not intensity-modulated direct detect will continue to be an option [24.36]. The OIF 400G ZR project gives insight that it is possible to use coherent technology for shorter reaches. If that is the case, then 400 G/λ may be a viable technology to develop 1.6 Tb Ethernet products.

24.6.1 Future Market Dynamics

As showcased previously, the nature of the market for data center interconnects has changed rapidly in the past few years. The diverse enterprise data center has seen its market share shrink, while the market growth is now dominated by the cloud data centers. The hallmark of the enterprise market is the diversity in data center sizes and designs requiring a wide variety of connectivity solutions. The cloud data centers are notable for their exceptional scale and uniformity.

As cloud data centers grow to take on tasks like IoT, DNN, and AI, the requirements are going to have a direct impact on the optical technology of the future. With the ability to bring significant market volume to the adoption of innovative technologies and new topologies, there is a growing need to understand how the technical choices impact the adoption and deployement of optical transceivers in the data centers of the future. Power, latency, and cost will become critical factors in optical transceiver technology.

References

24.1 IEEE 802.3 Ethernet Working Group: IEEE industry connections ethernet bandwidth assessment, http://www.ieee802.org/3/ad_hoc/bwa/BWA_Report.pdf (2012)

24.2 Facebook: Introducing data center fabric, the next-generation Facebook data center network, https://engineering.fb.com/production-engineering/introducing-data-center-fabric-the-next-generation-facebook-data-center-network/ (2014)

24.3 K.P. Jackson, C.L. Schow: VCSEL-based transceivers for data communications. In: *VCSELs – Fundamentals, Technology and Applications of Vertical-Cavity Surface-Emitting Lasers*, ed. by R. Michalzik (Springer, Berlin, Heidelberg 2013)

24.4 L.A. Neto, D. Erasme, N. Genay, P. Chanclou, Q. Deniel, F. Traore, T. Anfray, R. Hmadou, C. Aupetit-Berthelemot: Simple estimation of fiber dispersion and laser chirp parameters using the downhill simplex fitting algorithm, J. Lightwave Technol. **31**, 334–342 (2013)

24.5 B. Booth: Thoughts on big ticket items, http://www.ieee802.org/3/bs/public/15_03/booth_3bs_01a_0315.pdf (2015)

24.6 SFF-8665: Specification for QSFP+ 28 Gb/s 4X pluggable transceiver solution (QSFP28) Rev 1.9, https://www.snia.org/sff/specifications (June 29, 2015)

24.7 G. Morthier, P. Vankwikelberge: *Handbook of Distributed Feedback Laser Diodes* (Artech House, London 2013)

24.8 G.L. Li, C.K. Sun, S.A. Pappert, W.X. Chen, P.K.L. Yu: Ultrahigh-speed traveling wave electroabsorption modulator – design and analysis, IEEE Trans. Microw. Theory Tech. **47**, 1177–1183 (1999)

24.9 S. Kanazawa, H. Yamazaki, Y. Nakanishi, Y. Ueda, W. Kobayashi, Y. Muramoto, H. Ishii, H. Sanjoh: 214-Gb/s 4-PAM operation of flip-chip interconnection EADFB laser module, J. Lightwave Technol. **35**, 418–422 (2017)

24.10 NeoPhotonics Corporation: Laser 28G and 56G EML High Performance, https://www.neophotonics.com/products/laser-28g-56g-eml-high-performance/

24.11 T. Fujisawa, S. Kanazawa, W. Kobayashi, K. Takahata, A. Ohki, R. Iga, H. Ishii: 50 Gbit/s uncooled operation (5–85 °C) of 1.3 μm electroabsorption modulator integrated with DFB laser, Electron. Lett. **49**, 204–205 (2013)

24.12 T.L. Chen, C. Doerr, R. Aroca, S.Y. Park, J.C. Geyer, T. Nielsen, C. Rasmussen, B. Mikkelsen: Silicon photonics for coherent transmission. In: *Opt. Fiber Commun. Conf., OSA Technical Digest (online)* (Optical Society of America, Washington, DC 2016), Paper Th1B.1

24.13 N. Pavarelli, J.S. Lee, M. Rensing, C. Scarcella, S. Zhou, P. Ossieur, P.A. O'Brien: Optical and electronic packaging processes for silicon photonic systems, J. Lightwave Technol. **33**, 991–997 (2015)

24.14 J. Liu, M. Beals, A. Pomerene, S. Bernardis, R. Sun, J. Cheng, J. Michel: Waveguide-integrated, ul-

tralow-energy GeSi electro-absorption modulators, Nat. Photonics **2**, 433–437 (2008)
24.15 S. Lange, S. Wolf, J. Lutz, L. Altenhain, R. Schmid, R. Kaiser, M. Schell, C. Koos, S. Randel: 100 GBd intensity modulation and direct detection with an InP-based monolithic DFB laser Mach–Zehnder modulator, J. Lightwave Technol. **36**, 97–102 (2018)
24.16 IEEE: *802.3-2018 – IEEE Standard for Ethernet* (IEEE, New York 2018)
24.17 The Optical Internet Working Forum: http://www.oiforum.com
24.18 SWDM Alliance: http://www.swdm.org/msa/
24.19 Broadcom, Inc.: Product Brief, https://docs.broadcom.com/doc/PB_AFBR-89BDDZ_2018-01-19
24.20 100G PSM4 MSA: http://www.psm4.org
24.21 Interface standard 100 Gb/s lowcost 100 G optical interface: http://www.cwdm4-msa.org
24.22 A. Bechtolsheim, M. Paniccia: 100G CLR4 Industry Alliance, https://www.intel.com/content/dam/www/public/us/en/documents/presentation/clr4-press-deck.pdf (2014)
24.23 Open Compute Project: Facebook – CWDM4-OCP, https://www.opencompute.org/wiki/Networking/SpecsAndDesigns#Facebook_-_CWDM4-OCP
24.24 Industry Consortium 4 WDM: www.4wdm-msa.org
24.25 Inphi Corporation: ColorZ specifications, https://www.inphi.com
24.26 100GLambda MSA: https://www.100glambda.com
24.27 T. Yamamoto: High-speed directly modulated lasers. In: *Opt. Fiber Commun. Conf., OSA Technical Digest* (Optical Society of America, Washington, DC 2012), Paper OTh3F.5
24.28 P. De Dobbelaere, G. Armijo, J. Balardeta, B. Chase, Y. Chi, A. Dahl, Y. De Koninck: Silicon-photonics-based optical transceivers for high-speed interconnect applications. In: *SPIE OPTO* (2016), Paper 977503-977503
24.29 Y. Matsui, T. Pham, T. Sudo, G. Carey, B. Young, C. Roxlo: 112-Gb/s WDM link using two directly modulated Al-MQW BH DFB lasers at 56 Gb/s. In: *Opt. Fiber Commun. Conf. Post Deadline Papers, OSA Technical Digest (online)* (Optical Society of America, Washington, DC 2015), Paper Th5B.6
24.30 P.P. Baveja, M. Li, D. Wang, C. Hsieh, H. Zhang, N. Ma, Y. Wang, J. Lii, E. Liang, C. Wang, M. Ho, J. Zheng: 56 Gb/s PAM-4 directly modulated laser for 200G/400G data-center optical links. In: *Opt. Fiber Commun. Conf., OSA Technical Digest (online)* (Optical Society of America, Washington, DC 2017), Paper Th4G.6
24.31 V. Bhatt: Deploying DSP in optical transceiver modules. In: *Opt. Fiber Commun. Conf., OSA Technical Digest (online)* (Optical Society of America, Washington, DC 2015), Paper W4H.4
24.32 The QSFP-DD MSA: https://www.qsfp-dd.com
24.33 OSFP MSA: https://osfpmsa.org
24.34 Consortium for Onboard Optics: https://onboardoptics.org
24.35 IEEE P802.3ck 100 Gb/s, 200 Gb/s, and 400 Gb/s Electrical Interfaces Task Force: http://www.ieee802.org/3/ck/
24.36 D.V. Plant, M. Morsy-Osman, M. Chagnon: Optical communication systems for datacenter networks. In: *Opt. Fiber Commun. Conf., OSA Technical Digest (online)* (Optical Society of America, Washington, DC 2017), Paper W3B.1

Brad Booth

Microsoft
Snohomish, WA, USA
brbooth@microsoft.com

Brad Booth is a distinguished leader in Ethernet technology development and standardization. Currently heading up the Ethernet Technology Consortium and the Consortium for On-Board Optics, he is a Network Hardware Manager at Microsoft, leading the development of hyperscale networking products for Microsoft's cloud data centers. He is the founder and past Chairman of the Ethernet Alliance. Brad was previously a Distinguished Engineer in the Office of the CTO at Dell Networking.

David Piehler

Dell Technologies
Santa Clara, CA, USA
david.piehler@dell.com

David Piehler is a Distinguished Engineer at Dell Technologies' Networking business unit, where his work focuses on high-speed inter and intradata center optical transmission. Previous positions include Chief Scientist at NeoPhotonics, Vice President, Broadband Access Networks R+D at Harmonic, Inc., and Entrepreneur-in-Residence at the Mayfield Fund. He received his PhD in Physics from the University of California, Berkeley for experimental work in nonlinear optics.

25. Optical Switching for Data Center Networks

Nick Parsons, Nicola Calabretta

Cloud computing, the Internet of Things, and Big Data applications are imposing stringent requirements on communications within warehouse-scale data centers (DC) in terms of high bandwidth, low latency, and massive interconnectivity. Traditional DC networks based on electronic switching use hierarchical tree-structured topologies that introduce communication bottlenecks and require high energy consumption. Thus, to enable scalable growth both in the number of connected endpoints and in the exchanged traffic volume, novel architectural and technological innovations have to be investigated.

Optical switching technologies are attractive due to their transparency to data rate and data format, and enable energy-efficient network architectures that eliminate layers of power-consuming optoelectronic transceivers. In particular, new architectures that exploit optical circuit switching (OCS), optical packet switching (OPS), and optical burst switching (OBS) technologies have been widely investigated recently for intra-DC networks.

This chapter reports on the technologies used to implement OCS, OPS, and OBS nodes, together with recently investigated and demonstrated optical data center network (DCN) architectures.

The modern data center is a facility designed to accommodate large numbers of computing elements that are connected together to perform distributed data processing, storage, and distribution tasks. The scale and number of data centers have grown rapidly over the past decade, driven by the accelerating demand for cloud-based services.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_25

25.1 Data Center Network Architecture: Requirements and Challenges

Data centers require reliable and resilient power distribution, cooling, and communication networks. Considerable advances have been made in data center infrastructure engineering and management, aimed at bringing down the capital, operating, and energy costs of such facilities. However, server utilization remains low because of imperfect management and distribution of resources, together with data latency caused by oversubscription in the data center network (DCN) that interconnects clusters of compute, memory, and storage devices. Some cloud service providers have gone so far as to state that the network is throttling their data center.

Consequently, there is considerable scope for developing novel devices and architectures to enhance DCN efficiency. In this chapter, we examine the potential for switching directly at the optical layer in data center networks and discuss recent advances in the state of the art.

25.1.1 Data Center Network Evolution

As Internet traffic continues to grow exponentially, the limitations of existing data center network (DCN) architectures to scale have become increasingly apparent. Bandwidth requirements inside data center networks are typically 3–4 times greater than the external Internet demand and doubles at least every 12–15 months as the rate of data generation accelerates, with hundreds of new devices being connected every second across the globe. Data center facilities are under continuous pressure to grow infrastructure capacity while also improving quality of service, reducing costs, and importantly, minimizing their carbon footprint. Recent publications have documented how data center networks have evolved rapidly in response to this demand [25.1, 2], using uniform commodity hardware and a fault-tolerant multilayer network with a *Clos*-structured [25.3] nonblocking architecture, as shown in Fig. 25.1.

Point-to-point optical fiber links at 40 Gb/s or more are used to interconnect network nodes where routing is performed electrically by layer 3 packet switches, resulting in multiple optical–electrical–optical (OEO) conversions between endpoints. Because fiber runs in large data center networks may span up to 2 km or more, standard single-mode fiber is generally used in preference to multimode fiber as link rates increase to 100 Gb/s and beyond.

25.1.2 Requirements for Optical Switching Technologies Used in Data Centers

Data centers consist of a multitude of servers as computing nodes and storage subsystems, interconnected by appropriate networking hardware and accompanied by highly engineered power and cooling systems [25.4]. The DCN is designed to support the large data workloads exchanged between the parallel server machines. Traditional DCNs use a multitier architecture, with tens of servers housed in individual racks, which are themselves grouped into clusters. Top-of-rack (ToR) switches interconnect the servers via copper or optical links, while interrack communication is handled by layers of electronic switches. Ideally, the DCN should provide a full bisection bandwidth, thus the oversubscription ratio is 1:1, indicating high server utilization and computation efficiency. However, due to the superlinear costs associated with scaling the bandwidth and port density of electronic switches, such a design would be prohibitively expensive for a large-scale DCN. In practice, DCs tend to enforce an oversubscription of 1:4 to 1:10 [25.5]. More bandwidth is available for intrarack than interrack communication, and similar ratios are to be found at higher switching layers.

A set of stringent requirements are imposed on DCNs, a few key points of which are listed in the following:

- *Capacity*: An increasing fraction of data centers are migrating to warehouse scale. Although substantial traffic will continue to cross between users and data centers, the vast majority of data communication takes place within the data center [25.6]. Recent studies have shown a continuous increase of interrack traffic with a clear majority (> 50%) of traffic being intracluster [25.7]. Higher-bandwidth interconnects in combination with high-capacity switching elements are required, especially for interrack and intercluster communications, to avoid congestion drops caused by the inherent burstiness of flows in the intentionally oversubscribed network [25.1].
- *Latency*: Packet latency is defined as the time it takes for a data packet to traverse the network from the sender to receiver node (end-to-end latency), which includes both the propagation and switch latency. Within closed environments such as DCs, the latency is dominated by the contributions from the buffering, routing algorithm, and arbitration in the

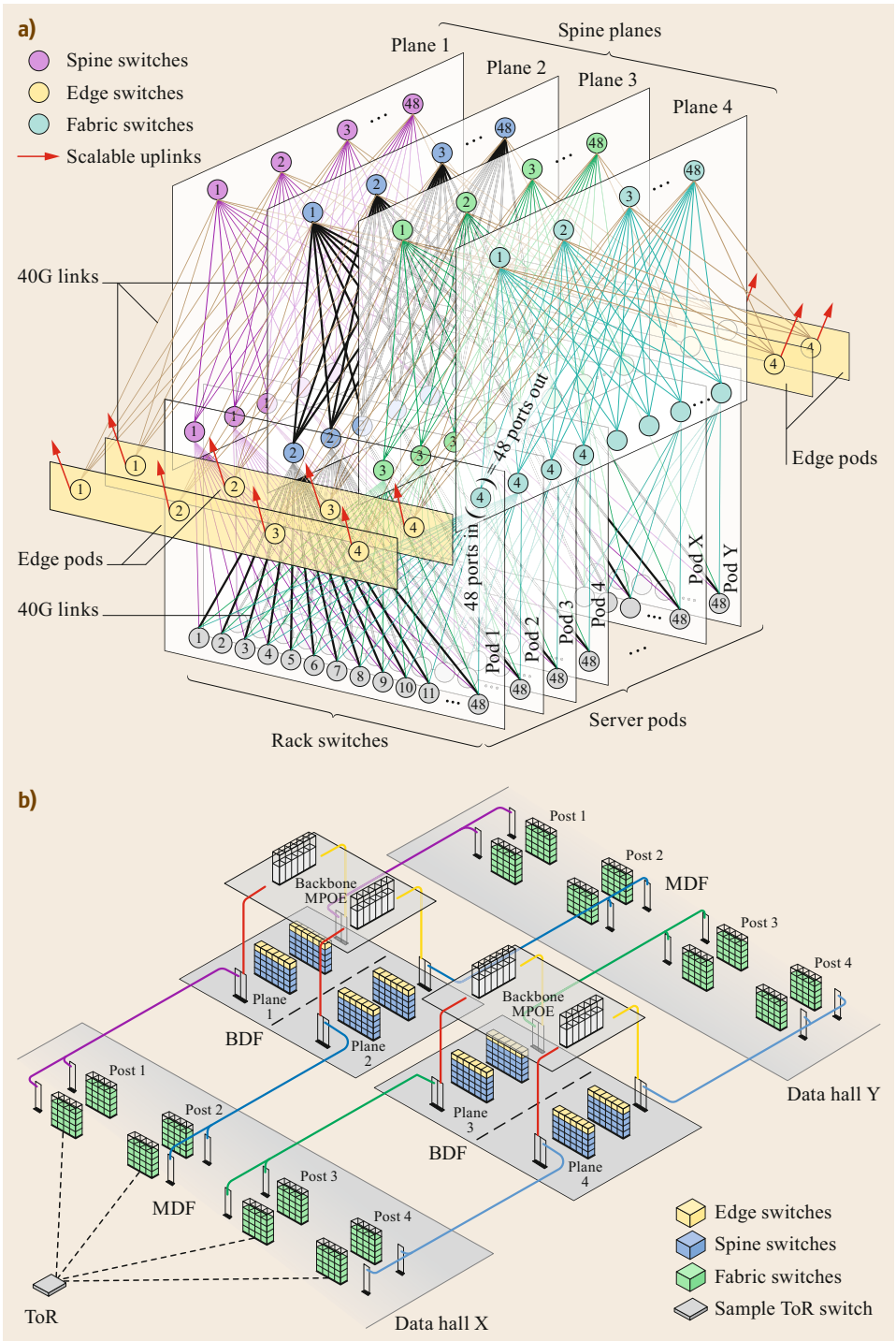


Fig. 25.1a,b Architecture and physical arrangement of scale-out data center networks. (After [25.2])

switching elements. Low latency is a crucial performance requirement, especially for mission-critical and latency-sensitive applications, where microseconds matter (e.g., financial networking).

- *Interconnectivity*: The servers in data centers have tens to hundreds of concurrent flows on average [25.7]. Considering the small fraction of intrarack traffic, almost all flows will traverse an uplink at the ToR switch as interrack communication. Therefore, the degree of interconnectivity supported by the switching network should be large enough to accommodate the number of concurrent flows. Moreover, considering that most flows are short and tend to be internally bursty, fast and dynamic reconfiguration of such interconnectivity (e.g., statistical multiplexing) is also needed to guarantee efficient bandwidth utilization and timely service delivery.
- *Scalability*: The network architecture should enable scaling to large numbers of nodes to address future capacity needs in a cost-efficient manner. Extension of an existing network in terms of both node count and bandwidth in an incremental fashion is preferable, i.e., without having to replace a disproportionate amount of the installed hardware.
- *Flexibility*: Data centers are expected to adopt technologies that allow them to flexibly manage the service delivery and adapt to changing needs. To this end, the resources (such as computing, storage, and network) are pooled and dynamically optimized by the control plane through software configuration. In addition, open standards, open protocols, and open-source developments are increasingly used to facilitate and speed up the deployment, operation, and management of the application- and service-based environment.
- *Power/cost efficiency*: A data center represents a large capital investment to which the DCN makes a significant contribution [25.8]. Besides the costs of installing computing hardware and software, running a large-scale data center focuses attention on power consumption. Power efficiency is a key target for reducing energy-related costs while increasing network capacity. In this sense, significant efforts have been made towards the employment of optical technology and resource virtualization, both of which can lead to enhancements in power and cost efficiency [25.9].

As can be seen from these requirements, high-capacity switching networks with low switching latency and fine switching granularity (e.g., deploying statistical multiplexing) are necessary to effectively improve the bandwidth efficiency and handle DC traffic flows. The large number of concurrent flows makes large interconnectivity as well as fast reconfiguration a necessity for the switches, in which case a combination of circuit and fast optical switching approaches may be employed. Pairwise interconnection of optical circuits with reconfiguration times of tens of milliseconds has been exploited for applications with well-scheduled and long-lived tasks, while fast (packet or burst) optical switches have potential to be used for dynamic and small flows.

With the increasing number of server nodes and rapid upgrade in input/output (I/O) bandwidth, the above-mentioned requirements would be quite challenging for current DCNs, in terms of both switching node and network architectures. First, it is difficult for electronic switches to satisfy future bandwidth needs. The increasing parallelism in microprocessors has enabled continued advancements in computational density. Despite the continuous efforts from commercial silicon providers towards the development of application-specific integrated circuits (ASICs), the implementation of high-bandwidth electronic switch nodes is limited (to multi-Tb/s) by the switch ASIC I/O bandwidth due to the scaling limitations of the ball grid array (BGA) package [25.10]. Higher bandwidth is achievable by stacking several ASICs in a multitier structure, but at the expense of increased latency and higher cost. Another factor limiting the scaling of electronic switches is power consumption. As an electronic switch has to store and transmit each bit of information, it dissipates energy with each bit transition, resulting in power consumption at least proportional to the bit rate of the information it carries. In addition, the OEO conversions and format-dependent interfaces also need to be included as a front-end, greatly deteriorating the power- and cost-efficiency performance.

Interconnecting thousands of ToRs, each with a large amount of aggregated traffic, places an enormous pressure on the multitier tree-like topology employed by current DCNs. Due to the limited performance in terms of bandwidth and port density of conventional electronic switches, networks are commonly configured with oversubscription. Consequently, data-intensive computations become bottlenecked, especially regarding communication between servers residing in different racks or clusters. The multiple layers of switches also result in high latency, mainly caused by the queueing delay of buffer processing when a packet traverses the whole DCN to reach its destination. Therefore, to effectively address the bandwidth, latency, scalability, and power requirements imposed by next-generation DCNs, optical switching technologies and network architectures are of paramount significance. With the prevalence of high-capacity optical interconnects, optically switched DCNs have been proposed as a solution to overcome the potential scaling issues of electronic

switching and traditional tree-like topologies [25.11, 12]. A switching network that handles data traffic in the optical domain can avoid power-consuming OEO conversions, along with their associated data latency. Such a network also eliminates the dedicated interface required for modulation-dependent processes, achieving better efficiency and less complexity. Moreover, benefiting from their optical transparency, the operation of optical switching (including its power consumption) is independent of the bit rate of the information carried. Scalability to higher bandwidth and employment of wavelength-division multiplexing (WDM) technology can be seamlessly supported, enabling superior performance in terms of power per unit bandwidth.

Various optical switching techniques have been investigated for DC applications, among which OPS, OBS, and OCS are the most prominent. With respect to the requirements for DCNs, optical switching technologies and the potential of photonic integration can support high capacity and power/cost-efficient scaling. Software-defined networking (SDN) is also seeing penetration into newly proposed optical DCNs, to facilitate flexible provisioning and performance enhancement.

Automation of network operations is a key aim to reduce operating costs and increase service velocity for both telecom and data center network operators. SDN paradigms promise open, vendor-independent control planes that enable rapid innovation in service creation, resource optimization, and monitoring. However, the fiber layer still relies predominantly on manual intervention for moves, adds, and changes, requiring days to weeks to implement.

OCS networks employ mature and commercially available switching technologies with reconfiguration times on the order of tens of milliseconds and are best suited to applications with well-scheduled and long-lived tasks. OPS and OBS networks use fast optical switches that allow rapid on-demand resource utilization with highly flexible connectivity enabled by statistical multiplexing and are appealing switching schemes for DCNs, albeit currently at the research stage. In the following sections, the main classes of technologies employed to realize OCS, OPS, and OBS are discussed.

25.1.3 Optical Circuit Switching Technologies

Optical circuit switches (OCS) are a class of photonic devices that enable routing of data traffic directly at the fiber layer, without requiring OEO conversion. They are sometimes referred to as all-optical (OOO) switches, to differentiate them from OEO cross-connects where switching is performed in the electrical domain. For widespread OOO deployment in data center networks, several technology challenges need to be addressed:

- *Scalability*: Users see an ultimate demand to scale to several thousand ports, so that all nodes in a warehouse-scale data center can be connected without blocking. To avoid the creation of single points of failure when using one large switching matrix, scalable multiswitch architectures built from nonblocking elements of 100–500 port pairs are likely to have more attractive reliability, serviceability, and first-in costs.
- *Loss*: To enable use of commodity optical transceivers, OOO elements must minimize their demand on link power and optical signal-to-noise ratio (OSNR) budgets. If average losses are kept at around 1 dB as the fabric size increases, the impact on link loss is minimal, allowing consideration of multistage architectures.
- *Speed*: Ideal switching times down to the microsecond scale are potentially required so that paths can be set up and torn down rapidly as traffic demand changes. This is difficult for established micromechanical technologies, where the fastest switch times are currently on the order of 10 ms. Hybrid optical–electrical architectures are likely to emerge where short-term changes in demand are accommodated in the electrical domain, while OOO switches are reconfigured to align with longer-term, more persistent traffic flows.
- *Cost*: Probably the greatest barrier to entry historically has been the per-port cost of OOO switches. OOO economics are now competitive even compared with 10-GbE OEO and become more attractive as data rates increase.

Technologies for all-optical switching have matured considerably over the past decade and are well established in markets adjacent to data communications. They can be grouped broadly into four classes:

- *Robot patch panels* automate the manual cross-connection of fibers inside an enclosure [25.13]. While their optical and reliability performance is similar to that of a patch cord, each connection takes many seconds to complete, so the technology is more suited to applications where rapid and repeated reconfiguration is not required.
- *Planar lightwave circuits* (PLCs) use arrays of elemental 2×2, 1×2, or on/off waveguide device interactions (Fig. 25.2) to build larger optical switch matrices. A wide variety of device interactions and material systems can be used to create a switching

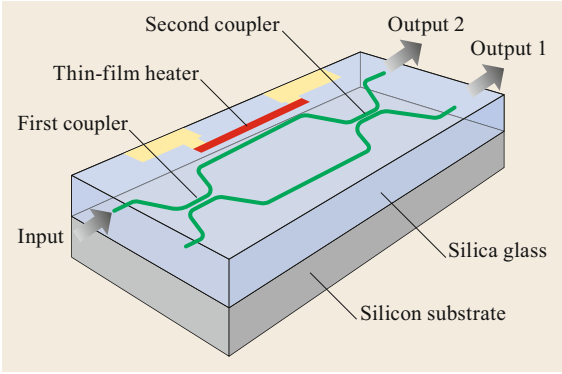


Fig. 25.2 Schematic of a 2×2 Mach–Zehnder interferometric PLC switch element in silica on silicon. (After [25.14])

element, but the maximum matrix size is limited by the accumulated fiber-to-fiber optical loss, polarization dependence, and crosstalk from the switch elements and waveguide crossovers necessitated by the inherent planar geometry. Typical $N \times N$ devices use wide-sense nonblocking or dilated Beneš topologies. Diversity schemes have been proposed to reduce polarization dependence, at the expense of chip complexity and waveguide crossovers.

- *3-D MEMS* devices make connections in free space by reflecting collimated light beams off pairs of steerable silicon micromirrors (Fig. 25.3) and are able to switch in around 20 ms but have relatively high optical loss. These devices generally require light to be present on the fiber to complete connections. Although early research demonstrated the possibility of switching fabrics of up to 1296×1296 [25.15] and, more recently, 512×512 [25.16], commercial products have been restricted to 320×320 matrix size by yield limitations on both the monolithic micromirrors and the fiber collimator arrays.
- *Direct beam-steering switches* point pairs of collimated fibers directly at each other using piezoelectric actuators to combine good optical performance with millisecond switching speeds and high radix connectivity. Very low backreflections minimize impairments to high-speed signals. The integrated position control of each port is independent of the light level and enables switching of dark fibers for path preprovisioning. Direct beam steering is currently the most promising approach to create high-radix optical circuit switches, with nonblocking cross-connects using DirectLight technology available commercially at matrix sizes up to 384×384 fiber ports and median connection loss of 1.5 dB [25.17].

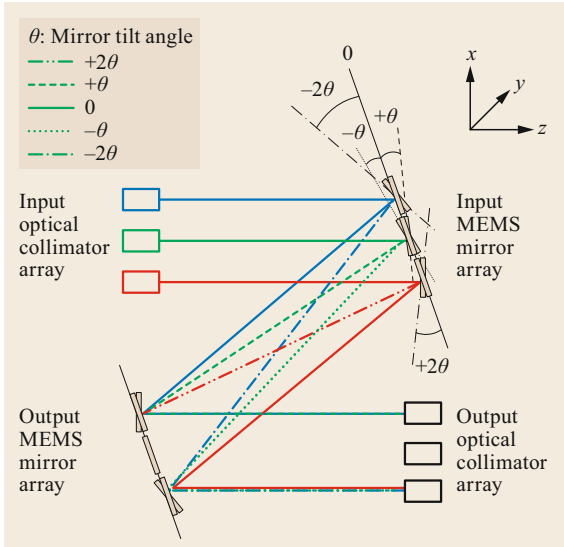


Fig. 25.3 Cross-sectional view of a conventional 3-D MEMS optical switch. (After [25.17])

The patented DirectLight dark fiber beam-steering optical switch technology provides nonblocking connectivity between 2-D arrays of fiber-coupled lenses in free space by using piezoelectric actuation in combination with integrated position sensor feedback to directly align the collimated optics on each path, thus avoiding the performance impairments associated with conventional microelectromechanical system (MEMS) micromirrors. Switching occurs entirely independently of the power level, color, or direction of the light on the path. The principle of operation is shown in Fig. 25.4, where opposing 2-D arrays of fiber-pigtailed collimators are individually steered by piezoelectric actuators via a low-stress flexure pivot. Voltages applied to the actuators independently control the collimator orientation in two angular dimensions. The pointing angles of the actuators are continuously monitored by high-accuracy absolute position sensors.

The maximum number of ports addressable in each dimension of the 2-D array varies approximately with the product of the actuator angular range and the separation of the arrays, and inversely with the actuator spacing. However, because the collimators in a DirectLight actuator rotate about the fiber termination and have a finite axial length, the minimum actuator spacing is also a function of the angular range. Consequently, the fill factor of the array is relatively low, leading to very good crosstalk isolation between adjacent ports. Interleaving of actuator arrays can also be used to double the apparent port density. The ingress and egress fiber arrays are built up in rows using modular slices containing up to 16 fiber ports. This allows the configu-

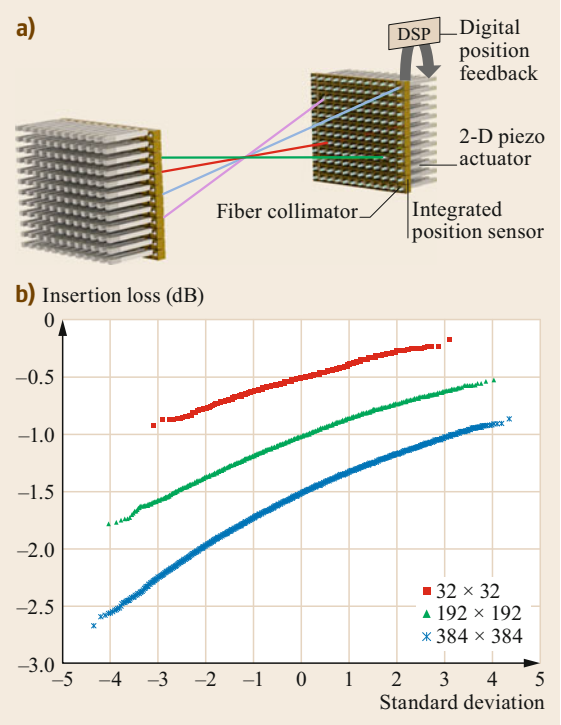


Fig. 25.4 **(a)** DirectLight beam-steering optical switch technology; **(b)** normal scores of measured optical loss distribution for all N^2 connections in 32×32, 192×192, and 384×384 switch modules. (Adapted from [25.18], © H+S Polatis)

ration of a wide range of switch matrix sizes to address multiple requirements while maintaining a consistent cost per port.

Table 25.1 compares the relative attributes of commercialized examples of each technology.

Emerging Optical Switching Technologies

Considerable research efforts continue to be applied towards realizing compact, high-radix optical circuit switches that can be closely integrated with switch and signal processing electronics using planar lightwave circuits based on silicon photonics. The challenges of minimizing on- and off-chip loss as well as OSNR degradation still remain, however. A promising approach reported recently [25.22] combines silicon photonics and 1-D MEMS vertical adiabatic couplers as shown in Fig. 25.5, to create a 50×50 switch matrix with switching time below 1 μs and chip dimensions smaller than 1 cm^2. Although each connected path needs to traverse many waveguide crossovers, only one vertical coupler is active in each path, reducing the loss variation across the matrix.

The use of space- and wavelength-division multiplexing to increase the capacity and efficiency of optically switched data center networks is also a topic of current academic research. Wavelength-switched networks are discussed in the next section and show promise for fast reconfiguration and for multicast (one-to-many) connectivity. SDM approaches are also receiving attention because of the relatively high cost of (electrically and/or optically) multiplexing data streams at 100 Gb/s and above. In the fiber-rich environment within a data center, parallel optics approaches such as 100GBASE PSM4 (four fibers per direction at 28 Gb/s) are gaining support for short-reach transport at the lowest cost per bit.

In the (very) long term, there may be a case for deploying more exotic multicore fibers in data center networks if the challenge of direct coupling to such fibers can be solved economically, using for example vertical-cavity surface-emitting laser (VCSEL) and detector arrays closely integrated with silicon processors. The feasibility of matrix switching of four core fibers as shown in Fig. 25.6a has recently been demonstrated [25.23] using DirectLight technology and microlens array telecentric optics (Fig. 25.6b) to perform core-to-core imaging between potentially up to 96 multicore fibers (MCFs) with demonstrated loss and crosstalk of 1–2 dB and 34 dB, respectively.

Table 25.1 Performance of commercially available optical circuit switch technologies

Technology	Direct beam steering	3-D MEMS	Robot patch panel	Planar lightwave circuit (PLZT)
$N \times N$ matrix size, max	384×384	320×320	256×256	4×4
Wavelength range (nm)	1260–1675	1260–1630	1260–1625	1550
Optical loss, typical/max (dB)	1.5/2.5	1.8/3.5	–/1.0	7.0/–
Wavelength-dependent loss (dB)	0.3	1.0	–	–
Polarization-dependent loss (dB)	0.1	0.3	–	1.0 typ
Optical return loss (dB)	55	35	50	–
Dynamic crosstalk (dB)	–50	–38	–	30 typ
Optical input power min/max (dBm)	Dark/+27	–20/+5	–	–
Switching speed (ms)	25	25	25 000	0.00001 typ
Vendor	Polatis [25.18]	Calient [25.19]	Wave2Wave [25.20]	Epiphotonics [25.21]

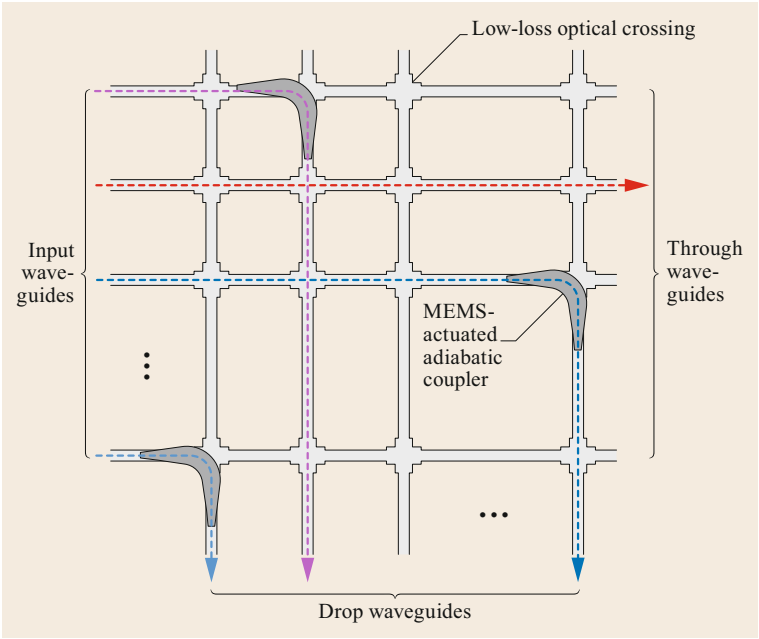


Fig. 25.5 Matrix architecture of silicon photonic switch with MEMS adiabatic coupler. (After [25.19])

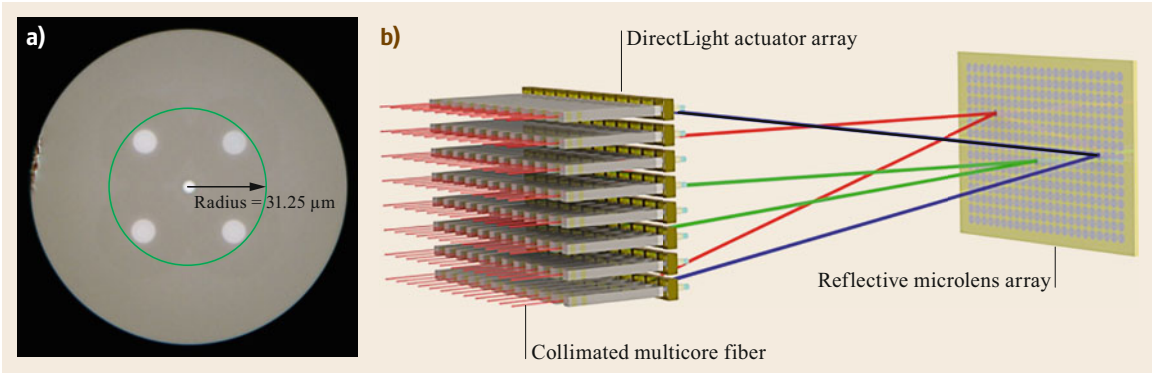


Fig. 25.6a,b Multicore fiber optical circuit switch: **(a)** four-core fiber cross-section with central alignment core; **(b)** free-space optical beam-steering arrangement (adapted from [25.18], © H+S Polatis)

25.1.4 Optical Packet and Burst Switching Technologies

Table 25.2 summarizes some examples of different classes of optical switching technologies for OPS/OBS nodes, where comparisons of different attributes in terms of switching performance are also reported. A broad range of technologies has been developed for OPS and OBS systems.

Fast switching technologies generally use either interferometric or gating switch elements, holding the potential for photonic integration to further scale capacity. Reasonably high-radix connectivity can be enabled by cascading 1×2 or 2×2 switching elements such as 2×2 Mach–Zehnder interferometers (MZIs) and microring resonators (MRRs). MZIs with electrooptic switching offer faster reconfiguration than thermooptic tuning, but extra optical amplification is normally needed due to their relatively high insertion loss, thereby compromising the scalability due to OSNR degradation. Another category of fast (nanosecond) optical switches is implemented by passive arrayed waveguide grating routers (AWGRs) together with tunable lasers (TLs) or tunable wavelength converters (TWCs). The interconnection scale and performance are largely dependent on the capability of the TLs and TWCs. Note that wavelength-selective switches (WSSs), MRRs, and AWGRs are all wavelength dependent. For the broadcast-and-select

Table 25.2 Optical switching technologies for implementing OPS and OBS nodes

	Switching time	Transparency	Scale	Loss	Application	Reference
2-D optical MEMS	≈ 50 µs	Good	50 × 50	Fair	OBS	[25.24]
LCoS WSS	≈ 10s ms	Good	1 × 40	Fair	OBS	[25.25]
Mach–Zehnder (thermooptic)	≈ 10s µs	Good	32 × 32	High	OBS	[25.26]
Microring resonator (thermooptic)	≈ 10s µs	Fair	8 × 8	Fair	OBS	[25.27]
PLZT MZI	≈ 10 ns	Good	8 × 8	High	OPS, OBS	[25.21]
InP MZI	2.5 ns	Good	8 × 8	High	OPS, OBS	[25.28]
$LiNbO_3$ MZI	≈ 1 ns	Fair	32 × 32	High	OPS, OBS	[25.29]
MZI+EAM	< 10 ns	Good	8 × 8	Fair	OPS, OBS	[25.30]
TWC+AWGR	≈ 10s ns	Poor	10 s × 10 s	Fair	OPS, OBS	[25.31]
TL+AWGR	≈ 10s ns	Good	100 s × 100 s	Low	OPS, OBS	[25.32]
EAM B&S	≈ 1 ns	Good	8 × 8	High	OPS, OBS	[25.33]
SOA B&S	≈ 1 ns	Good	64 × 64	Fair	OPS, OBS	[25.34]
SOA multistage	< 10 ns	Good	16 × 16	Fair	OPS, OBS	[25.35]
Semiconductor optical phase array	≈ 20 ns	Good	64 × 64	Fair	OPS, OBS	[25.36]

(B&S) architecture, the semiconductor optical amplifier (SOA) and electroabsorption modulator (EAM) are commonly used as gating elements. The broadcast stage introduces high splitting losses, in which case the SOA can provide loss compensation, which is essential to realize large connectivity. In the practical implementation of an OPS/OBS network, the techniques listed here—individually or in combination—can be further included in a network to provide basic switching units [25.37, 38].

25.2 Data Center Network Architectures Based on Optical Circuit Switching

High-speed optical fiber links are used extensively to make connections between data center servers and higher-level top-of-rack/end-of-row switches, but switching and routing are still performed mainly in the electrical domain, requiring power-hungry OEO conversions. However, at locations where large information pipes are connected without requiring packet-level grooming, OEO routers can be replaced with lower-cost, energy-efficient, all-optical (OOO) matrix switches.

Recently, leading data centers in the USA and Europe have begun deploying OOO switches for managed cross-connect services at line rates of 10 Gb/s and higher in large co-located Internet peering rooms. Not only are OOO fabrics physically smaller and offer around 100 times lower power consumption than typical OEO switches, but also fiber-layer switching becomes relatively more cost- and power-efficient as connection speeds climb to 100 Gb/s and beyond.

All-optical switching provides additional performance benefits in comparison with alternatives, including:

- *Format independence*: In contrast to OEO routers, OOO switches are transparent to the data rate, modulation scheme, color, and direction of the light in the fiber, thereby future-proofing the investment in infrastructure connectivity as transponder technology continues to advance.
- *Dynamic provisioning*: Compared with a fiber patch panel, OOO switches are cabled once at installation and enable accurate tracking of additions, moves, and changes with instant remote provisioning. The risks of disturbing premium connections and breaching service-level agreements through errors in manual intervention are eliminated.
- *Speed-of-light latency*: Data delays through an OEO switch are typically in the microsecond range, leading to slower data transfer across clusters and reduced compute efficiency. By contrast, OOO switches add just a few nanoseconds of propagation delay, with no timing skew or jitter across the whole network.
- *Monitoring and protection*: Optical switching facilitates real-time quality-of-service monitoring, real-time intrusion detection, and in-circuit testing by scanning high-value test assets across multiple channel taps. Use of integrated optical power meters also enables autonomous optical-layer protection switching to enhance service availability.

Conventional network equipment is managed and controlled by proprietary, vendor-specific software. This limits architecture and service innovation, since the introduction of new capabilities becomes dependent on equipment vendor development cycles, meaning that it can take months to design, test, and field new services. The addition of features, or integration of new technologies, likewise requires proprietary system upgrades that require coordination with each individual equipment supplier. The European project COSIGN [25.39] is studying the benefits of integrating optical packet- and circuit-switched architectures under software-defined network control and management planes (Fig. 25.7).

The power of the software-defined network paradigm lies in the use of open interface standards that expose the network element functionality that enables operators to take back control of their network to automate basic operations and facilitate rapid service delivery and innovation. New features can be prototyped in hours or days rather than weeks or months, which greatly enhances competitiveness and avoids vendor lock-in.

Furthermore, the open, programmable interfaces of SDN controllers greatly simplify the adoption of innovative technologies such as all-optical switching into the existing network infrastructure. Moreover, the ability of the SDN control plane to dynamically monitor network performance and react in real time to changing traffic loads allows operators to take greater advantage of the inherent benefits of a virtualized fiber infrastructure.

25.2.1 Automating Data Center Network Operations

Energy-efficient optical circuit switching is a key technology in scale-out and multitenant data centers to facilitate virtualization and orchestration of optical resources. Key pain points for data center operators today focus on the time and effort required (measured in days to weeks) to provision new client services both within and between data centers and to recover from network or equipment faults. A fully programmable optical network improves data center service velocity, reliability, and efficiency by enabling: rapid provisioning of virtualized fibers on demand; optimization of resources in containerized and disaggregated architectures; autonomous optical layer protection for critical connections; and dynamic, low-latency, router bypass for offloading elephant flows to reduce congestion and

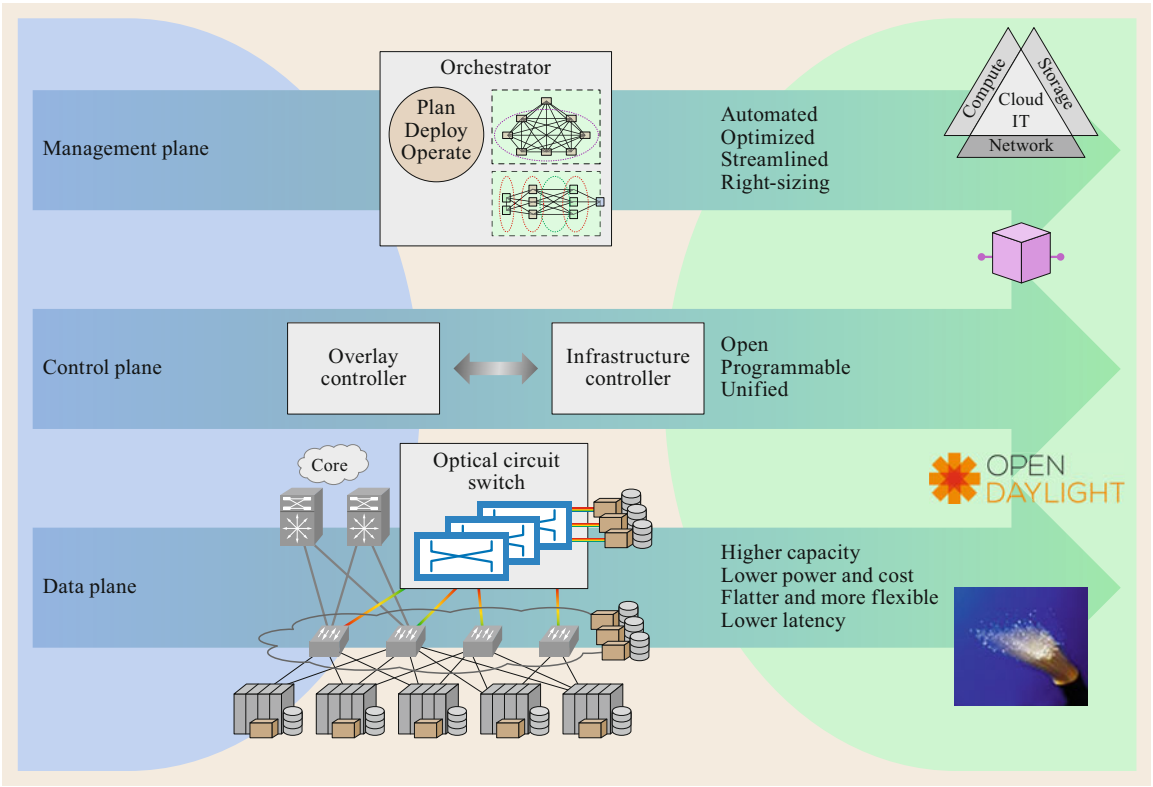


Fig. 25.7 Benefits of optical switching in a software-defined data center network

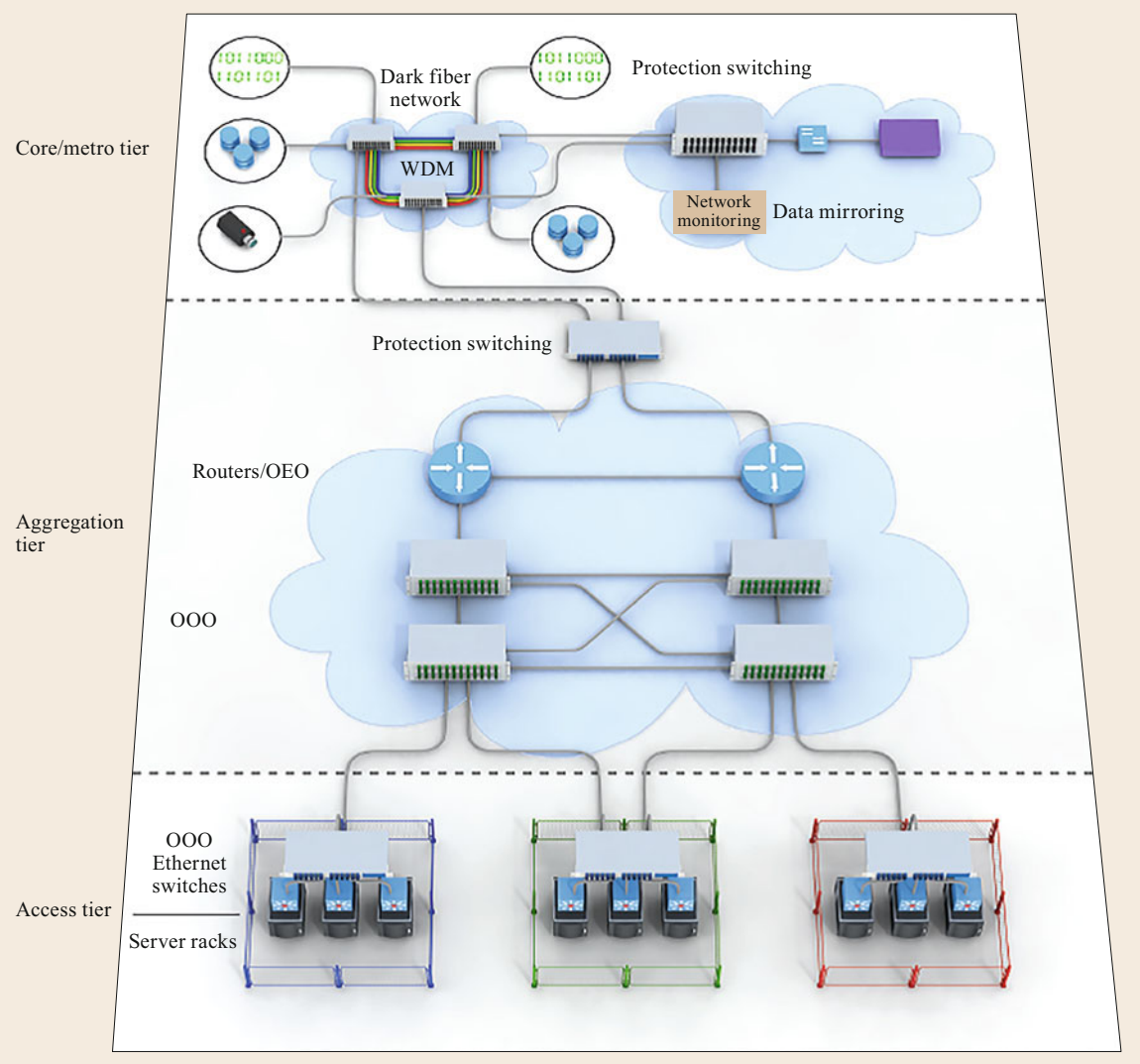


Fig. 25.8 Opportunities for all-optical switching in data center networks

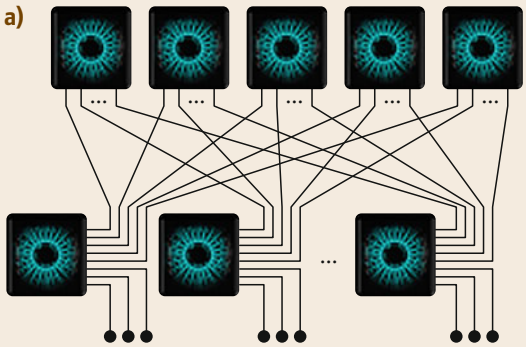
improve server utilization. Figure 25.8 illustrates areas where optical switching can be used to advantage in the access, core, and aggregation tiers of a data center network.

Adoption of optical circuit switching brings the fiber layer under software control to allow rapid provisioning, protection, and reconfiguration of network resources on demand, which creates additional benefits to operators, including:

- Eliminating manual patch errors and the potential for service interruption
- Maintaining the current state of fiber-layer connectivity in a software database
- Creating optical demarcation points in multitenant/multi-service-provider environments
- Facilitating bridge-and-roll during equipment commissioning, upgrade, and replacement
- Providing physical isolation between virtualized network slices for enhanced security
- Enabling aggregation of optical taps for network monitoring.

As the number of managed fiber-layer connections in large data center networks continues to grow well beyond 10 000 endpoints, multistage nonblocking optical switch fabrics are needed to provide dynamic configurability with high availability and maintainability, using

a)



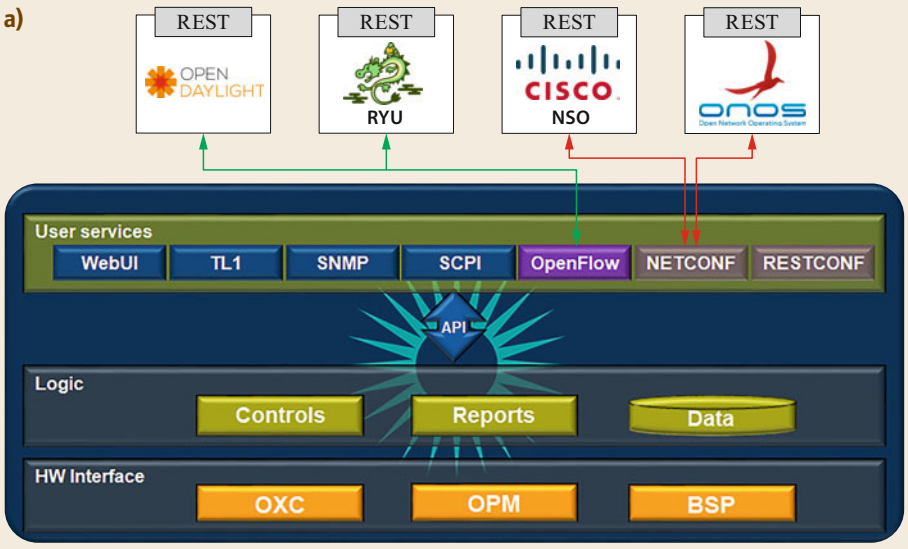
b)

Two-level Clos fabric size (endpoints)

Max module ports	96	192	384	768
Common module size	3072	12 288	49 152	196 608
Maximum edge module	4608	18 432	73 728	294 912
Typical optical loss (dB)	4.0	4.0	5.5	–

Fig. 25.9 (a) Two-level Clos switching fabric; (b) Achievable number of connected endpoints as a function of Direct-Light matrix size

a)



b)

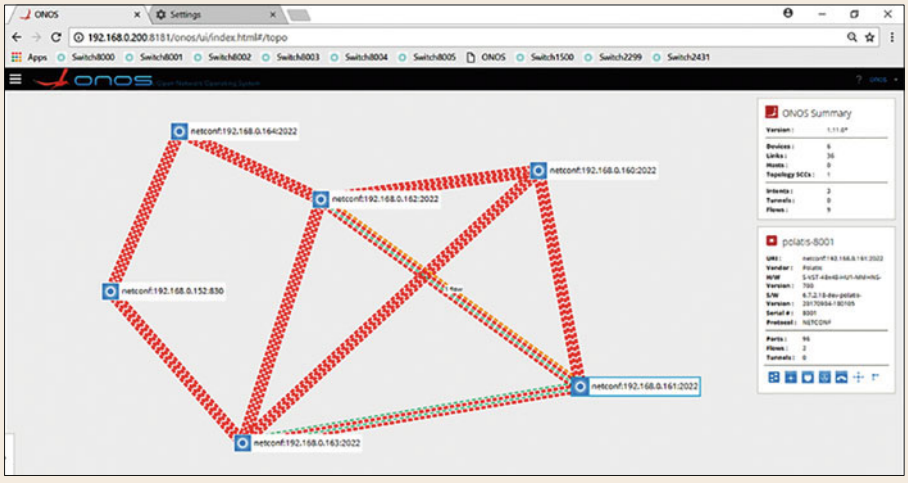


Fig. 25.10 (a) Embedded software architecture in a Polatis optical circuit switch, exposing conventional and SDN user interfaces; (b) example screenshot from an ONOS SDN controller with six OCS elements and an intent-based connection path. (Adapted from [25.18], © H+S Polatis)

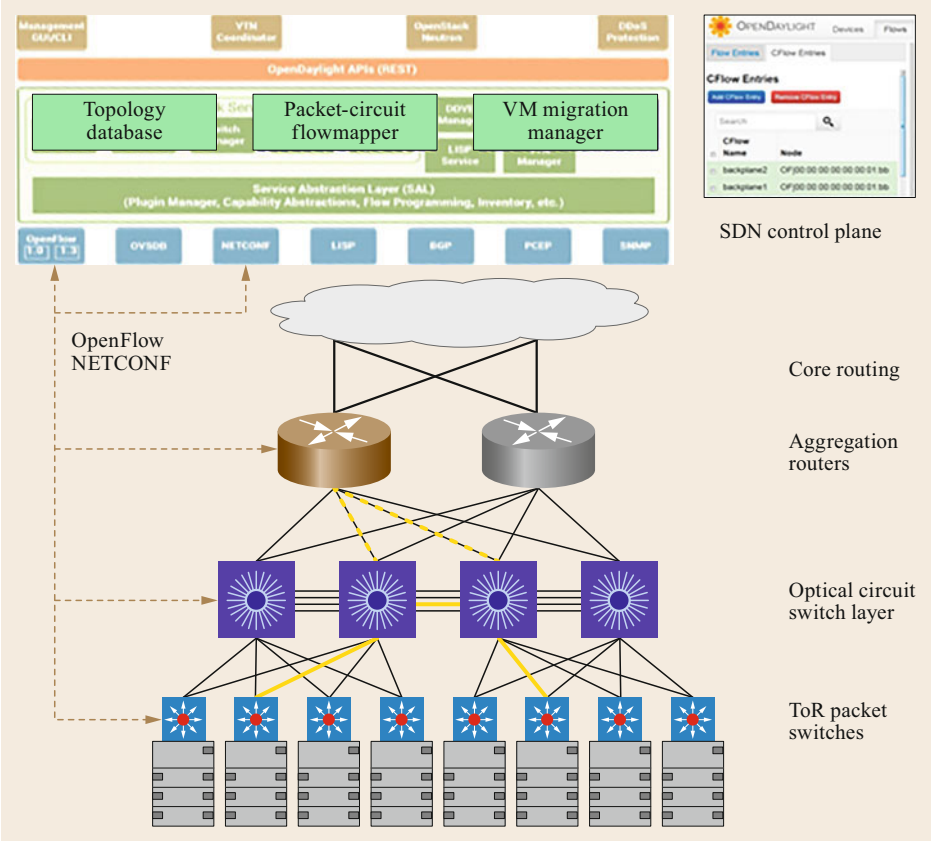


Fig. 25.11 Data center operators can use SDN-enabled all-optical switching to optimize network resources

open, programmable interfaces for network management and automation. In addition to matrix size, the key design considerations for multistage fabrics include the total optical loss, reconfiguration speed, reliability, and cost.

The low-loss 384×384 OCS discussed in Sect. 25.1.3 allows the concatenation of switch elements to realize strict nonblocking dynamic fiber cross-connects that can provide hitless scaling to over 49 000 endpoints in a uniform two-level Clos architecture, with a typical loss of less than 6 dB (Fig. 25.9). This loss is well within the available power budgets of low-cost transceivers.

A key requirement for the adoption of OCS technologies in scale-out DCNs is that they are simple to control, monitor, and upgrade via existing SDN control planes. Consequently, OCS network elements need to expose interfaces for protocols based on open standards. Figure 25.10a shows an example of an embedded software architecture for a Polatis OCS, which supports, alongside conventional interfaces such as Transaction Language 1 (TL1) and Simple Network Management Protocol (SNMP), the Open Networking Foundation's OpenFlow 1.4 protocol [25.40] to enable close integration with packet-based L2/L3 OEO switches, together with the Internet Engineering Task Force (IETF)-ratified NETCONF [25.41] and RESTCONF [25.42] protocols which are more generally applicable to transport networks. NETCONF and RESTCONF use the YANG data modeling language [25.43] to describe explicitly the features of a network element that can be abstracted by SDN controllers. Figure 25.10b presents an example screenshot from the ONOS [25.44] SDN controller graphical user interface, showing six interconnected OCS elements where a routing intent has been enacted using the southbound NETCONF interface.

One challenge with real-world deployments of optical circuit switches is that, in contrast to packet-switched networks, discovery of the interconnection topology between OCS elements is not automatic, since optical switches by their nature do not interact directly with the data plane. Consequently, manual or out-of-band methods of peer port discovery (such as connection point identification) need to be employed when commissioning and validating such networks.

25.2.2 Hybrid Packet/Optical Switching for Efficient DC Router Bypass

Data center aggregation is generally performed by stacks of packet routers and switches that connect server farms and provide access to the core/metro networks. Routers also host many other functions including connectivity, discovery, firewall, load balancing, and intrusion detection. Current data center aggregation architectures are optimized for handling the short, bursty traffic patterns that have dominated in the past. However, they are neither efficient nor cost effective for handling the larger persistent data flows between server clusters that are increasingly common with virtual cloud-based services. These persistent data flows are now common in routine operations such as virtual machine migration, load balancing, and data storage backup. Some early studies showed that these so-called elephant flows dominated cluster-to-cluster traffic in Microsoft data centers, with over 90% of network capacity being occupied with flows that last more than 10 s [25.46, 47]. However, more recent traffic engineering studies in a Facebook data center [25.7] indicated that the picture was less clear—possible because of the extensive use of load balancers—but that elephant flows could only really be identified in clusters running Hadoop algorithms. Offloading elephant flows to dynamically provisioned optical circuits has been demonstrated by a number of research groups [25.39, 48, 49] and not only provides a low-latency, high-speed path between endpoints but also simultaneously relieves congestion at the packet layer.

Figure 25.11 shows an example of how SDN-enabled all-optical switching can optimize resources in a multivendor data center network. Here we show a hybrid packet/optical data center aggregation layer operating under an SDN control plane using OpenFlow and NETCONF protocols to manage the optical circuit switches and other network elements. This arrangement is just one representation of a configuration-on-demand architecture that optimizes the utilization of data center resources. In this scenario, the topology and properties of the optical switch layer are published to the SDN controller and orchestration layers via a plug-in residing in the service abstraction layer. In this arrangement, persistent east–west data flows can be identified by the packet-circuit flowmapper in the SDN controller using data analytics calculated from monitoring flows between network endpoints. Once identified, the controller can dynamically reconfigure the network using OpenFlow and NETCONF to create a low-latency optical circuit for these large data flows through the optical switch layer. Optical switching is ideally suited to managing persistent data flows that do not need any subwavelength grooming. Offloading these elephant flows to the optical circuit switch layer relieves congestion and reduces latency through L2/L3 packet-switched routers, thereby substantially improving the efficiency and throughput of the network infrastructure.

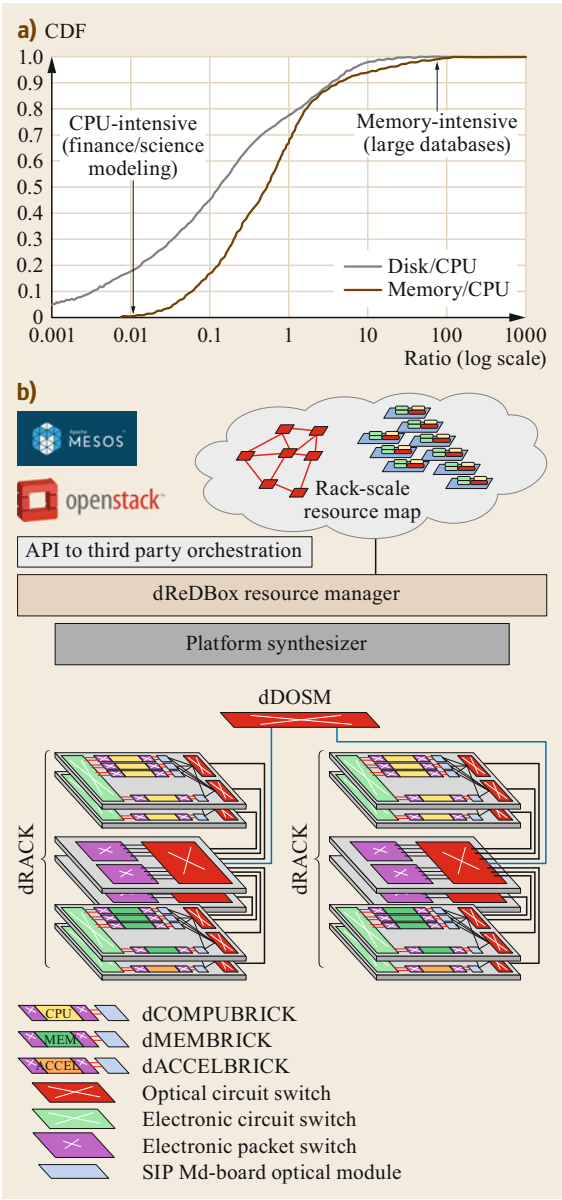


Fig. 25.12 **(a)** Distribution of storage and memory to compute requirements by task in Google data centers (after [25.45]); **(b)** a function-centric disaggregated DC architecture with network-attached memory and dynamic optical connectivity ◄

25.2.3 Disaggregated Data Center Architectures

Conventional server-centric data centers typically use a fixed ratio between compute, memory, and storage resources, based on average or typical data center load requirements. However, the actual resource demands vary widely depending on the task, as shown in Fig. 25.12a for a Google data center [25.45]. Consequently, much of the available resource in a data center is fragmented and left idle because of the wide spread of demand. If the necessary resources could be disaggregated within a pod, cluster, or even a server and dynamically reconfigured on demand, then considerable efficiency benefits could be realized. The Horizon2020 project dReDBox [25.50] is developing new disaggregated data center architectures, with low-latency connectivity between central processing units (CPUs) and network-attached memory using function-programmable hardware and optical circuit switching at the server, rack, and cluster level, as shown in Fig. 25.12b.

SDN orchestration enables optimal mapping of tasks to resources as well as efficient zero-copy migration of virtual machines (VMs). The low loss of the DirectLight OCS elements allows multiple switches to be traversed within the available power budget of silicon photonics midboard transceivers, enabling efficient usage of all resources within a physical boundary defined by the maximum latency tolerable between compute and memory. Minimizing end-to-end latency is a key challenge, since the main contributor to latency in these short links is the signal processing delay in the transceiver, rather than the length of optical fiber used. Sub-microsecond round-trip latency has been demonstrated on such links by ensuring that sufficient optical power is available at the receiver to avoid the use of forward error correction.

25.3 Data Center Network Architectures Based on Optical Packet/Burst Switching

The main motivation to investigate OPS and OBS technologies and design novel DCN architectures by exploiting such technologies is the potential to achieve subwavelength granularity by exploiting statistical multiplexing of bursty flows, similar to electronic switching. Typical DCN architectures that employ OPS/OBS nodes consist of a set of electronic edge nodes, such as ToR switches that interconnect 40–80 servers or blades, each with embedded multiple microservers, interconnected by optical switches. Electrical data flows or packets from the client network with similar attributes are aggregated at the edge nodes in optical packets, bursts, or flows. The optical packets/flows at the edge node outputs are transparently forwarded by the optical switches without OEO conversion to the destination node. After arriving at the destination edge node, the optical packets/flows are disassembled and forwarded by the electronic edge node to the client network. The switching operation of the packet/burst (usually referred to as the payload) is determined by a packet header/burst control header (BCH), which is typically optically encoded in the same wavelength as the payload or carried by a separate wavelength which undergoes OE conversion and electronic processing at the optical switch node [25.51]. The main differences between OPS and OBS are as follows:

- In DCN architectures based on OPS technologies, the typical packet durations are quite short, on the order of a few hundreds of nanoseconds to few microseconds range. The packet header is transmitted in the same channel as the payload and either overlaps the payload in time or sits ahead of it. Typically, there is no advance reservation for the connection, and the statistical multiplexing of the channel in time and bandwidth can be utilized in the most flexible way. These features of OPS are very attractive and suitable for data center applications which require transmission of small datasets in an on-demand manner.
- DCN architectures based on OBS technologies employ more extensive burst aggregation on the order of tens to thousands of microseconds. The BCH is created and sent towards the destination in a separate channel prior to payload transmission. The BCH informs each node of the arrival of the data burst and drives the allocation of an optical end-to-end connection path. OBS enables subwavelength granularity by reserving the bandwidth only for the duration of the actual data transfer.

Note that the reconfiguration time of the optical switch including the control operation should be much shorter than the duration of the packet/burst, to ensure the low-latency delivery at a fast arrival rate as well as optimized bandwidth utilization. Practical realization of OPS/OBS relies heavily on the implementation of control techniques and the scheme adopted for contention resolution [25.52].

25.3.1 Technical Challenges in OPS/OBS Data Center Networks

Despite the advantages of increased capacity and power efficiency brought by optical transparency, practical use of OPS/OBS faces several challenges which need to be carefully considered for DC networking applications:

- *No practical optical memory available*: Contention resolution is one of the most critical functionalities that need to be addressed for an OPS/OBS node. However, no effective optical memory is available today. Some approaches have been proposed to solve contention by exploiting one or multiple domains:
  - *Time domain*: The contending packet/burst can be stored in an electronic buffer or delayed by a variable fiber delay line (FDL).
  - *Wavelength domain*: Exploiting a number of parallel wavelength converters, packet/burst contention can be avoided by forwarding the packets to alternative wavelength channels. Alternatively, wavelength-tunable transmitters can be employed to avoid contention by sending packets to an available free wavelength.
  - *Space domain*: Space switches are employed to solve contention of data packets by forwarding data to another output port (deflection routing).

  Techniques based on FDLs, wavelength conversion, and deflection routing suffer from the main drawback of a significant increase in system complexity for routing control and packet synchronization. Moreover, degradation of the signal quality, i.e., OSNR degradation, channel cross-talk, and signal distortion, results in limited and fixed buffering time. A more practical solution is to exploit the electronic buffer at the edge nodes and implement an efficient optical flow control. To minimize latency, the optical switch should be as close as possible to the edge nodes and fast decision-making is required. This is feasible in a DC environment with interconnects of limited range up to a few hundred meters.
- *Fast reconfiguration and control mechanism*: To fully benefit from the flexibility enabled by statistical multiplexing, fast reconfiguration of the optical switch is a key feature. Although OBS is less time demanding, slower OBS can cause inefficiency and unpredictability, especially under high network loads. Therefore, optical fabrics with fast switching times together with fast control mechanisms are desired. For DCN applications, the implementation of control techniques should allow an increase of the network scale and optical switch port count, and more importantly, occupy as least resources as possible.
- *Scalability*: Depending on the design and technology employed in optical switches, signal impairment and distortion are observed due to noise and optical nonlinearities. Consequently, optical switches are realized with limited port count. Scaling network interconnectivity while maintaining performance would require switches to have as large a port count as possible and to be intelligently connected. A flat topology also brings the benefits of simplified control and buffering, which may be problematic for fast optical switches. On the other hand, optical transparency and WDM technology would benefit the DCN in the context of scaling up the bandwidth density. Further improvements could be made by means of photonic integration, which greatly reduces the optical switch footprint and power consumption.
- *Burst-mode receivers*: Currently deployed electronic switch protocols are based on Ethernet, therefore synchronization is obtained between two connected ports by filling idle periods with repetitive patterns to provide continuous high-quality clock recovery. However, in the optical DCN, data are continuous but bursty. Therefore, the clock and phase of the data signal differ per link. Thus, for practical implementation, burst-mode receivers are required to synchronize the phase and clock of the received data packets. Moreover, although a DCN is a closed environment with more controlled optical power variation, the receiver should still be capable of handling packets with different lengths and optical power levels. Those functions contribute an overhead to the overall operational time of burst-mode receivers, which should be minimized to achieve higher throughput and lower latency. This is especially important in an intra-DC scenario where many applications produce short traffic flows.

OPS and OBS technologies providing high bandwidth and power efficiency have been adopted in recent research on optical DCNs [25.53, 54]. The next sections provide an overview of and general insight into recently proposed optical DCN architectures based on OPS/OBS, classified into different categories according to the switching technologies used. Although obviously incomplete due to the many publications in this field [25.55–63] and more examples, we provide here a brief description of representative architectures for different technologies.

25.3.2 Optical DCN Architectures Based on OBS

OBS with Fast (Commercial) Optical Switches

A DCN consisting of ToR switches at the edge and an array of fast (commercial) optical switches at the core to perform optical burst forwarding on preconfigured light paths has been proposed [25.64]. It has separate data and control planes, as shown in Fig. 25.13. Two-way reservation OBS is implemented, facilitated by the single-hop topology with the configuration of only one switch per request. It achieves zero burst loss with slight degradation of the latency owning to the limited round-trip time in the DC environment. The centralized control plane is responsible for the routing, scheduling, and switch configuration. It processes the control packets from the ToRs sent through a dedicated optical transceiver, finds an appropriate path to the destination ToR, and configures optical switches as allocated in the control packets. Since the fast optical switch connects to every ToR, scalability is challenging in terms of achievable port count for large number of ToRs. The resultant complexity in the control plane may be another bottleneck in scaling up the network.

OBS Ring Topology

OBS technology is utilized in [25.65] to improve the interpod communications in DCNs. The network architecture and the pods are depicted in Fig. 25.14. The pods are connected through multiple optical burst rings.

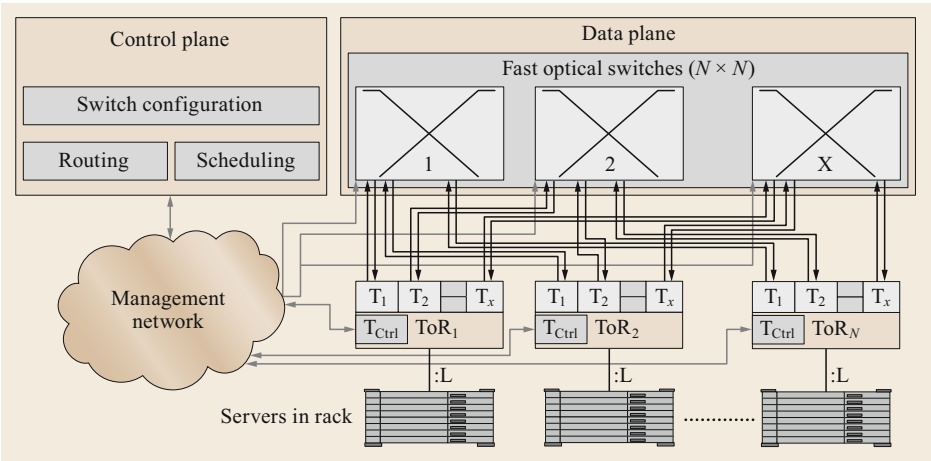


Fig. 25.13 DCN based on OBS with fast (commercial) optical switches

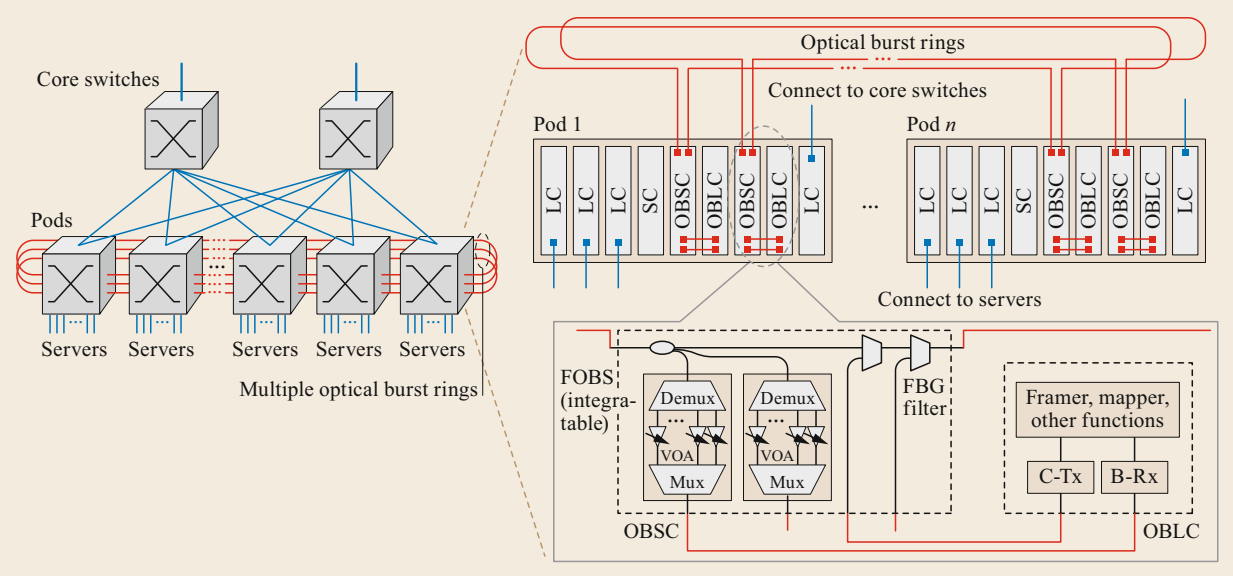


Fig. 25.14 Multiple optical burst rings and internal architecture of the pod. FBG: fiber Bragg grating, FOBS: fast optical burst selector

Bursty and fast-changing interpod traffic is handled by the core switches, while the relatively stationary traffic is handled via the optical burst rings. Some line cards (LCs) are configured for connecting the servers, while others are used to access the core switches. The switch cards (SCs) aggregate the traffic and, together with the control unit, make decisions to forward the traffic to the LCs connecting to the core switches or to an optical burst line card (OBLC) which sends the traffic in the form of a burst to the optical rings. The optical burst switch cards (OBSCs) perform optical burst add/drop to/from the optical burst rings, as shown in Fig. 25.14. The advantages of this architecture are the high interpod transmission bandwidth and high-radix interconnectivity (> 1000 pods). Much shorter connection reconfiguration time is offered compared with OCS-based solutions, achieving better bandwidth utilization.

Hybrid Optical Switching (HOS)

An optical switched interconnect based on hybrid optical switching (HOS) was proposed and investigated in [25.65]. HOS integrates optical circuit, burst, and packet switching within the same network, so that different DC applications are mapped to the most suitable optical switching mechanisms. As shown in Fig. 25.15, the HOS is arranged in a traditional fat-tree three-tier topology, where the aggregation switches and the core switches are replaced by the HOS edge and core node, respectively. The HOS edge nodes are electronic switches which perform traffic classification and aggregation. The core node has parallel optical switches composed of switching elements (SEs). A slow optical switch based on 3-D MEMS handles circuits and long bursts, and a fast SOA-based optical switch with a three-stage Clos network deals with packets and short bursts. The HOS control plane manages the scheduling and transmission of the optical circuits, bursts, and packets. Wavelength converters (WCs) are used to solve the possible contentions. Numerical studies show low loss rates and low delays, although the practical implementation of a large scale network remains challenging.

HOSA

The hybrid optical switch architecture (HOSA) shown in Fig. 25.16 is another DCN architecture that employs both fast and slow optical switches [25.54]. In contrast to the previous work that uses only fast optical switches [25.66], slow MEMS optical switches are added to exploit the benefits of both types of fabric. Traffic assembly/disassembly and classification is implemented in the newly designed ToR switch. The control plane still uses a centralized controller which receives connection requests and configures the data plane through a management network. The array of fast optical switches operates in an OBS manner, forwarding the data burst along a predefined connection path. The evaluation results show low-latency and high-throughput performance with low power consumption, assuming that slow/fast optical switches with large port counts are deployed in a single-stage network.

25.3.3 Optical DCN Architectures Based on OPS/OCS

Torus-Topology DCN

Figure 25.17 shows the torus DCN [25.67] based on codeployment of OPS and OCS. The architecture

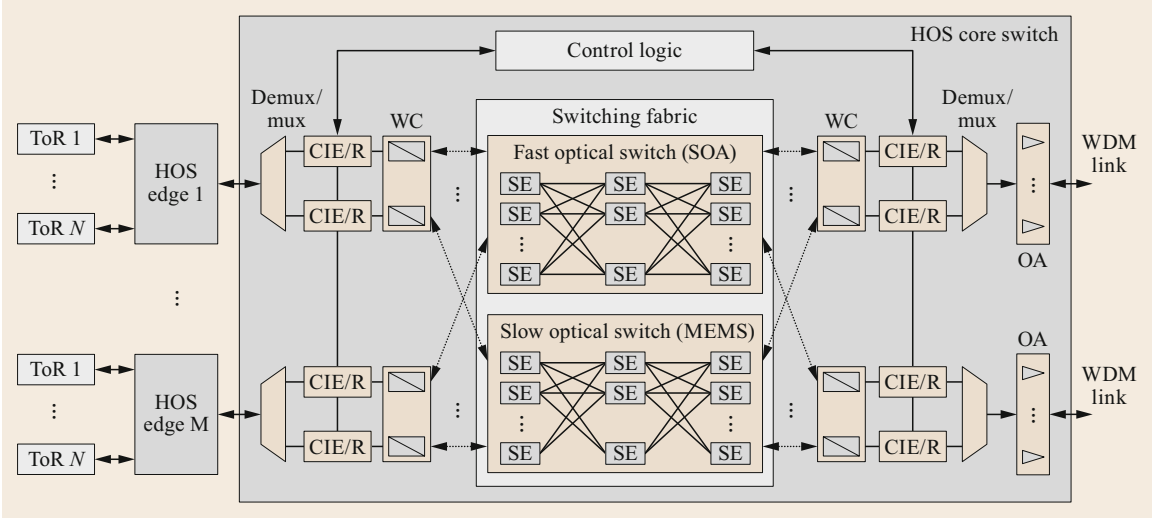


Fig. 25.15 HOS interconnection network

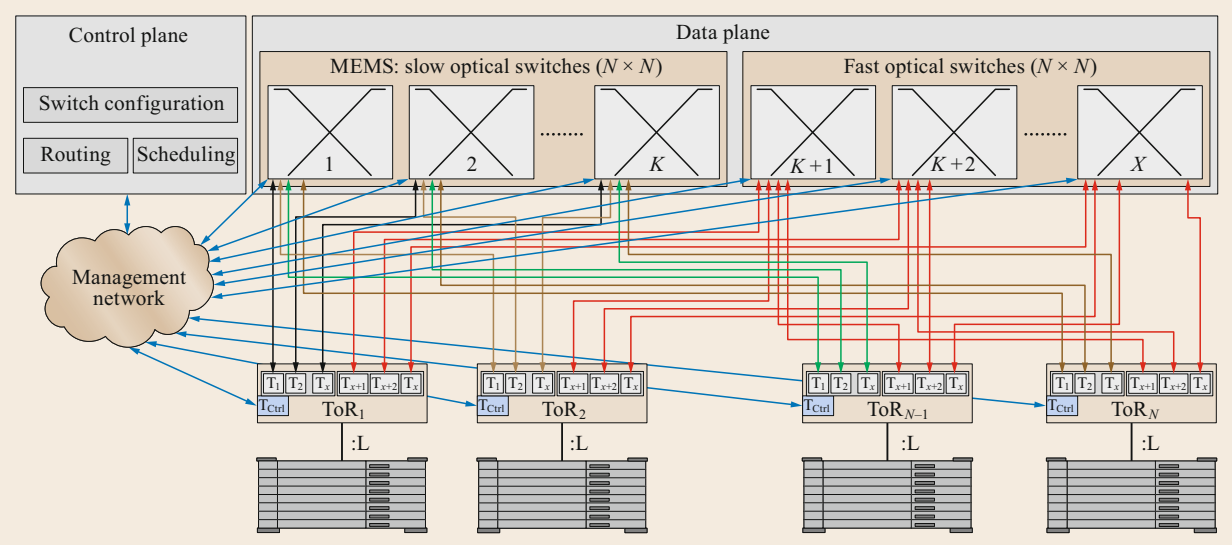


Fig. 25.16 HOSA DCN architecture

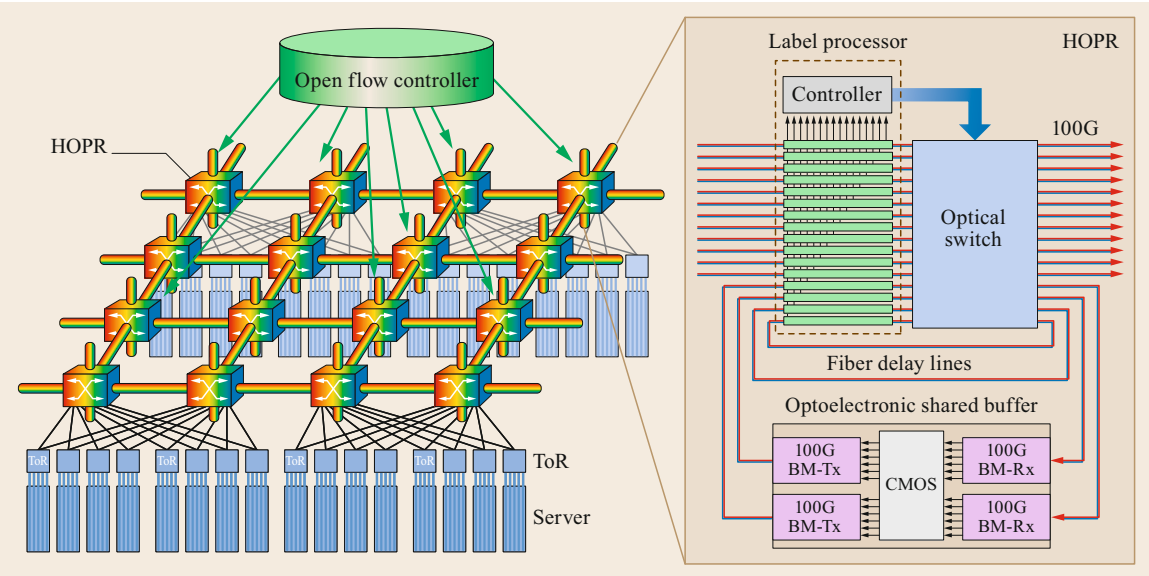


Fig. 25.17 Torus DCN employing hybrid optoelectronic routers (HOPRs)

features a flat topology where each hybrid optoelectronic router (HOPR), interconnecting a group of ToR switches, is connected to the neighboring HOPRs. The server traffic is converted into optical packets and fed into the corresponding HOPR with an attached fixed-length optical label. HOPR uses a fast optical fabric (EAM-based broadcast-and-select structure) which supports both packet operation and circuit operation (express path). Packet contention, which happens when a link is desired by more than one packet or is reserved by an express path, is solved by different schemes (i.e., deflection routing, FDLs, and optoelectronic shared buffers). The enabling technologies for implementing an HOPR have been detailed, aiming at high energy efficiency and low latency in the 100-ns regime. For efficient transfer of high-volume traffic, flow management has been implemented with OpenFlow-controlled express paths. Despite the multihop transmission needed for interconnecting the ToRs, the torus topology offers the advantages of superior scalability and robust connectivity.

Lightness

A flat DCN architecture integrating both OPS and OCS switching technologies to deal with heterogeneous application requirements has been investigated in the Lightness project [25.68]. The hybrid network interface card (NIC) located in each server supports the switching of traffic to either OPS or OCS, resulting in efficient utilization of the network bandwidth. As illustrated in Fig. 25.18, the SOA-based OPS, which employs a broadcast-and-select architecture, is plugged into the architecture-on-demand (AoD) backplane as a switching module to handle short-lived data packets. The AoD itself is a large-port-count fiber switch which can be configured to support OCS function for long-lived data flows. The network can be scaled by interconnecting multiple intracluster AoDs with an intercluster AoD. Another innovation made by Lightness is the fully programmable data plane enabled by the unified SDN control plane. It is worth noting that the switching operation of the OPS is controlled by the local switch controller based on the in-band optical labels, which is decoupled from the SDN-based control (e.g., lookup table update and statistical monitoring). Similar schemes can be found in Archon [25.69] and burst-over-circuit architectures [25.70], where the OPS is replaced by a PLZT-based optical switch and the AWGR with a TWC, respectively.

25.3.4 Optical DCN Architectures Based on OPS

IRIS: Photonic Terabit Router

The IRIS project has developed a photonic packet router that scales to hundreds of Tb/s capacity [25.71]. As shown in Fig. 25.19, the router employs a load-balanced multistage architecture. Each node (e.g., ToR

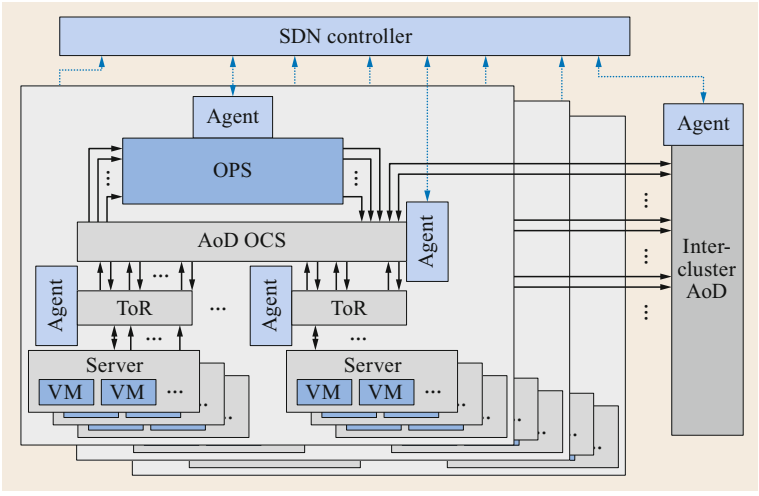


Fig. 25.18 The Lightness DCN architecture

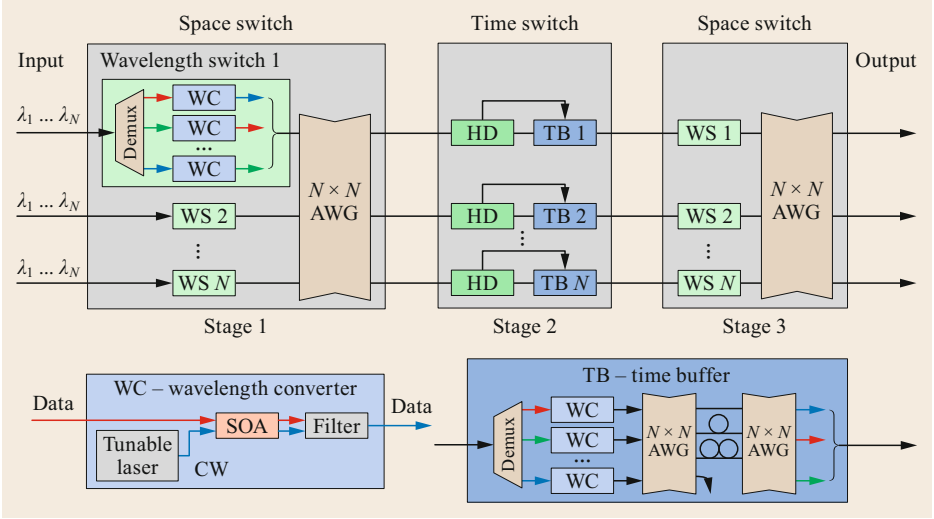


Fig. 25.19 IRIS photonic terabit router

switch) is connected to an input port of the first stage using N WDM wavelength channels, each carrying synchronous fixed-length data packets. The wavelength switch is based on an array of all-optical SOA-based wavelength converters to set the wavelength routing. The second stage is the time switch section, which contains N time buffers based on shared fiber delay lines. The wavelength is configured so that a packet entering a port of the time buffer always exits on the corresponding output port. The third stage then forwards the packet to the desired destination. Due to the periodic operation of the third space switch, the scheduling is local and deterministic for each time buffer, which greatly reduces the complexity of the control plane. The IRIS project has demonstrated the operation of a partially populated router with integrated photonic circuits and developed an interoperability card that can connect electronic routers with 10-Gb Ethernet interfaces to the IRIS router. The use of 40 Gb/s data packets and 80×80 AWGs allows this architecture to scale to 256 Tb/s capacity.

Petabit Optical Switch

The petabit optical switch is based on tunable lasers (TLs), tunable wavelength converters (TWCs), and AWGRs, as shown schematically in Fig. 25.20 [25.72]. The ToR switches are connected to the optical switch, which is a three-stage Clos network including input modules (IMs), central modules (CMs), and output modules (OMs). Each module uses an AWGR as the core. The SOA-based TWCs as well as the TLs in the line cards are controlled by the scheduler. A prominent feature of the switch is that packets are buffered only at the line cards, while the IMs, CMs, and OMs do not require buffers. This helps to reduce implementation complexity and achieve low latency. The performance of the petabit optical switch was evaluated via simulations, revealing high throughput because of the efficient scheduling algorithm.

Hi-LION

The large-scale interconnect optical network Hi-LION was proposed in [25.73]. It exploits fast tunable lasers and high-radix AWGRs in a hierarchy to achieve very large-scale and low-latency interconnection of computing nodes. The architecture of the full system and an example of a six-node rack are depicted in Fig. 25.21. The essence is to rely on the unique wavelength routing property assisted by electrical switching embedded in the node to provide all-to-all flat interconnectivity at each level of the hierarchy (node-to-node and rack-to-rack). As shown in Fig. 25.21, local and global AWGRs are used to handle intra- and interrack communication, respectively. Single-hop routing in the optical domain also avoids the utilization of optical buffers. However, a maximum hop count of up to seven can be required, including intrarack forwarding. Compared with previous AWGR-based solutions such as DOS (LIONS) [25.74] and TONAK LION [25.75], interconnectivity of more than 120 000 nodes can potentially be achieved.

Osmosis Optical Packet Switch

The Osmosis project targets accelerating the state of the art in optical switching technology for use in supercomputers [25.76]. The architecture of the single-stage 64-port optical packet switch implemented is illustrated in Fig. 25.22. It is based on a broadcast-and-select architecture, and the switching modules consist of a fiber and a wavelength-selection stage, both built with SOAs as the gating elements. The switching of the synchronous fixed-length optical packets is controlled via a separate central scheduler.

Performance studies on the Osmosis demonstrator confirmed its high-capacity and low-latency switching capabilities. A two-level fat-tree topology could potentially be built, further scaling the network to 2048 nodes.

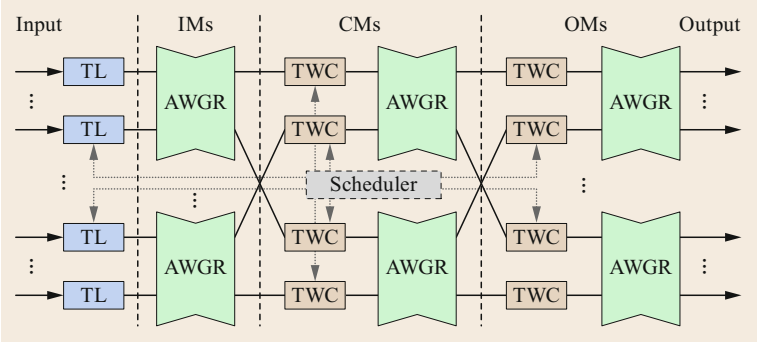


Fig. 25.20 The petabit optical switch architecture

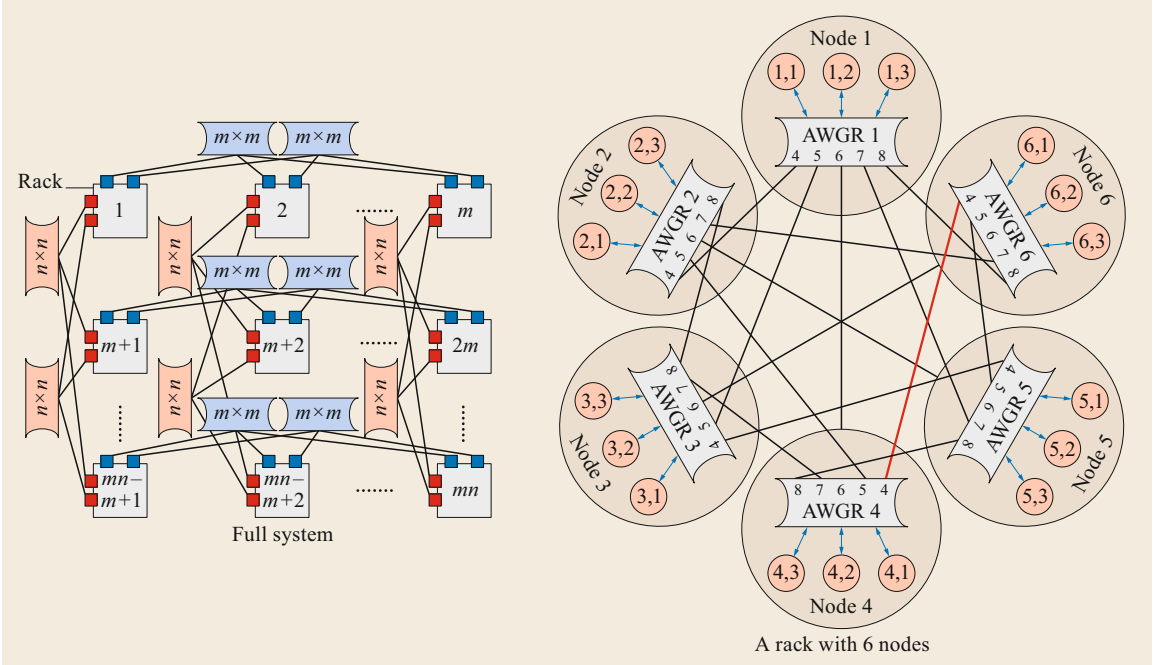


Fig. 25.21 Hi-LION full system with inter/intrarack AWGR communication

Data Vortex

The Data Vortex is a distributed interconnection network which is entirely composed of 2×2 switching nodes arranged in concentric cylinders [25.77]. As illustrated in Fig. 25.23, the Data Vortex topology integrates internalized virtual buffering with banyan-style bitwise routing, specifically designed for implementation with fiber-optic components. The 2×2 node uses an SOA as the switching element. The broadband operation of SOAs allows for successful routing of multichannel WDM packets. Packet contentions are resolved through deflection routing. The hierarchical multistage structure is easily scalable to larger network sizes. However, the practical scale is limited by the increased and nondeterministic latency, as well as deteriorated signal quality.

PetaStar

PetaStar (a petabit photonic packet switch architecture) [25.78] exploits the space, time, wavelength, and subcarrier domains. The architecture is shown in Fig. 25.24 and consists of a three-stage Clos-network

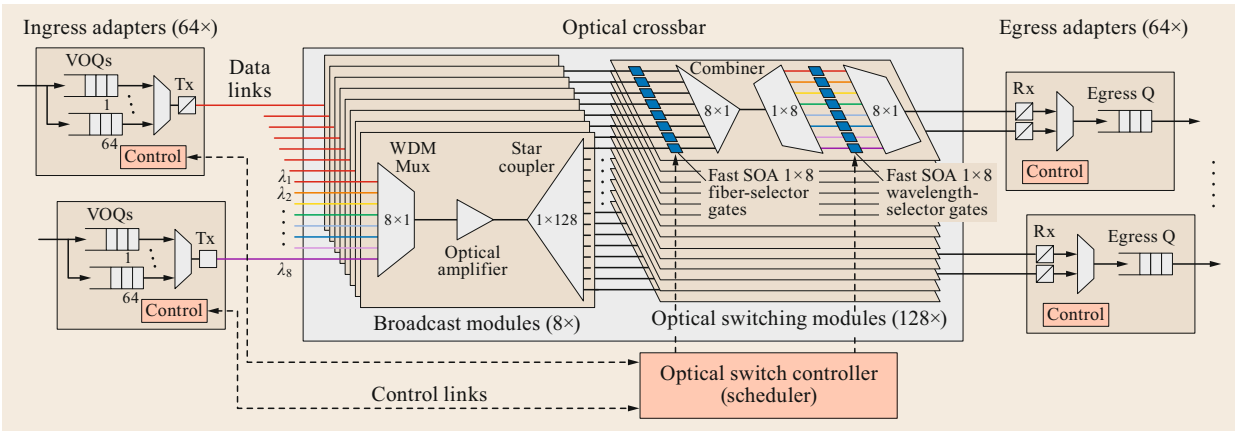


Fig. 25.22 Single-stage Osmosis switching system. VOQ: virtual optical queue

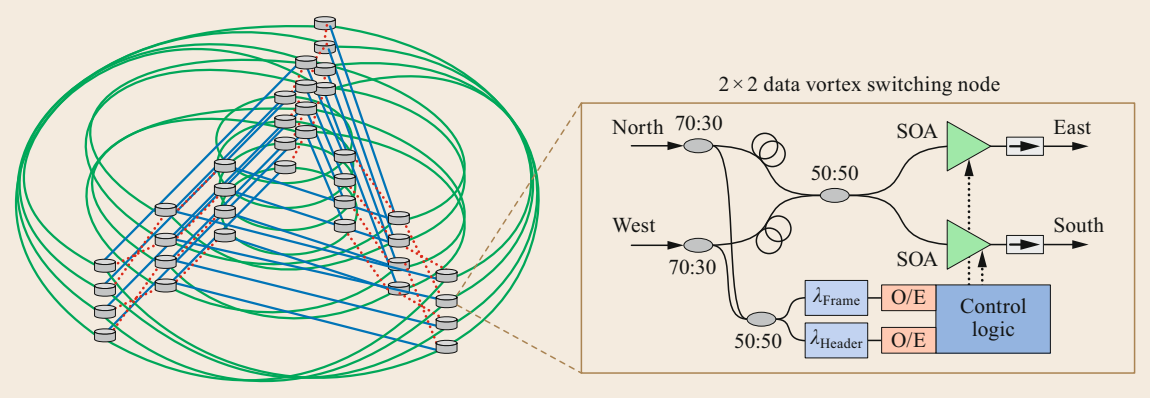


Fig. 25.23 The Data Vortex topology and distributed 2×2 nodes

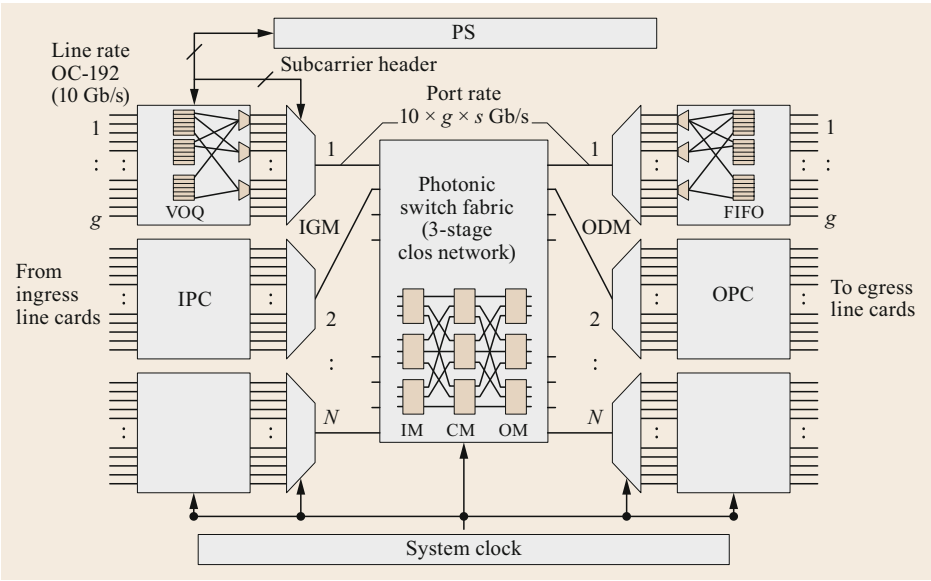


Fig. 25.24 The PetaStar topology

photonic switch fabric that allows the provision of scalable large-dimension switch interconnections. There is no optical buffer, and the buffer is implemented electronically at the input and output port controllers. On the other side, the photonic switch fabric transparently forwards the optical signals to the proper output. Optical time-division multiplexing technology further scales the port speed beyond electronic speeds up to 160 Gb/s to minimize the fiber connections. A novel concurrent matching algorithm called c-MAC is employed for packet contention. It is highly distributed such that input–output matching and routing-path finding are performed concurrently by scheduling modules. It has been discussed in [25.78] how the PetaStar architecture is capable of interconnecting 6400 nodes and that a total capacity of 1024 Pb/s can be achieved at a throughput close to 100% under various traffic conditions.

OPSquare

OPSquare was recently proposed [25.80] based on distributed fast WDM optical switches, which allow for flexible switching capability in both the wavelength and time domains. The architecture shown in Fig. 25.25 is based on two parallel switching networks to properly handle intra- and intercluster communication. It consists of N clusters, each grouping M racks. Each rack contains K servers interconnected via an electronic ToR switch. Two bidirectional WDM optical links are equipped at the ToR to access the parallel intra- and intercluster switching networks. The N $M \times M$ intracluster optical switches (ISes) and M $N \times N$ intercluster optical switches (ESs) are dedicated for intra- and intercluster communication, respectively. The i-th ES interconnects the i-th ToR of each cluster, with $i = 1, 2, \ldots, N$. The number of interconnected ToRs (and servers) scales

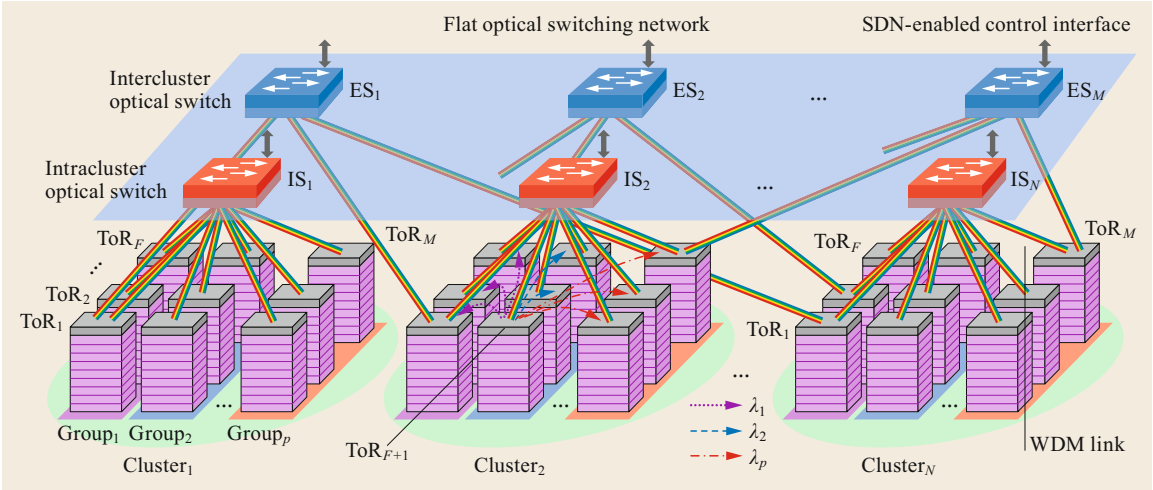


Fig. 25.25 OPSquare DCN architecture built on fast optical switches

as $N \times M$, so that by using a moderate port count of 32×32 ISs and ESs, up to 1024 ToRs (40,960 servers in case of 40 servers per rack) can be interconnected. The interface for the intra/intercluster communication consists of p WDM transceivers with dedicated electronic buffers to interconnect the ToR to the IS optical switch through the optical flow-controlled link [25.81], while q WDM transceivers interconnect the ToR to the ES optical switch.

The main benefits of this architecture are the low latency and bufferless operation due to the single-hop optical interconnection and the fast optical flow-control mechanism between the ToRs and optical switches, and the large path diversity, which improves the resilience of the network. Exploiting the transmitter wavelength assignment for the grouped top-of-rack (ToR) switches, large interconnectivity can be achieved by utilizing optical switches with a moderate port count, which leads to a significant improvement of the scalability and feasibility of the network.

A schematic of a fast optical switch node acting as an IS/ES is shown in Fig. 25.26. The optical switching is realized by an SOA-based broadcast-and-select architecture. The fast optical switch node has a modular structure, and each module consists of N units, each handling the WDM traffic from one of the M ToRs in a single group. The SOA has nanosecond switching speed and can provide optical amplification to compensate the losses caused by the broadcasting stage.

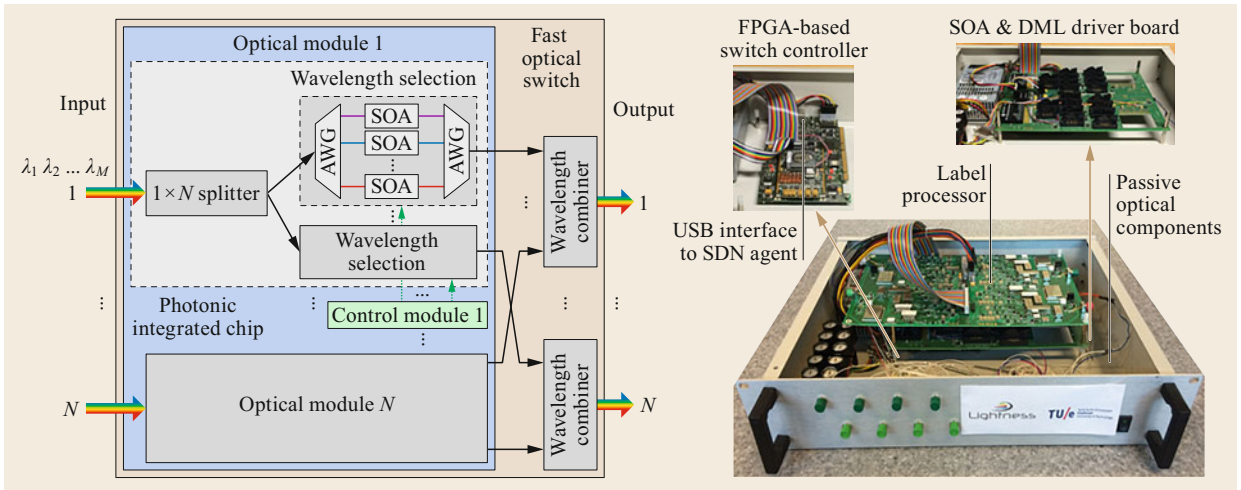


Fig. 25.26 Schematic of fast optical switch node and flow-controlled optical switch node prototype. (Adapted from [25.79], © TU Eindhoven)

Contention is solved through the optical flow control, according to which the ToR releases the packets stored in the buffers (for ACK) or triggers retransmission (for NACK). Different classes of priority can be applied to guarantee traffic with more stringent QoS requirements. The priority is defined by provisioning the lookup table in the switch controller through the SDN control interface [25.82]. In addition, the SDN control plane can create and manage multiple virtual networks over the same infrastructure by configuring the lookup table, and apply further optimization through the developed monitoring functions. The developed 4×4 optical switch prototype integrating an FPGA-based switch controller (with an interface to the SDN agent), label processor, SOA drivers, and passive optical components (circulators, couplers, etc.) is also shown in Fig. 25.26. Photonic integrated circuits (PICs) can reduce the footprint and power consumption. In view of this, a 4×4 WDM fast optical switch PIC has been designed and fabricated exploiting the modular architecture, as shown in Fig. 25.27. The modular photonic chip shown in Fig. 25.27 has 4×16 ports (the combiners shown in the schematic on the left side of Fig. 25.27 were not integrated into this photonic circuit due to lack of space) and integrates four optical modules. As reported in [25.55, 79], the compensation of the losses offered by the SOAs allowing for large dynamic range, together with the low cross-talk and the wavelength, time, and nanosecond switch operation, indicate the potential for scalability to higher data rates and port counts of the optical switch PIC and potential enhancement of the OPSquare DCN performance.

25.3.5 Comparison Between DCN Architectures Based on OPS and OBS Switching Technologies

Optical DCN architectures based on OPS and OBS have been presented in this chapter, and their different characteristics in terms of scalability, flexibility, and power/cost efficiency are summarized in Table 25.3. As seen from this table, for contention resolution, most of the schemes use practical electronic buffers (EBs) placed at the edge, either waiting for the scheduler command or retransmitting the packet/burst in case of contention. The efficiency of scheduling, the configuration time of the switch, and the round-trip time play an important role in reducing the processing latency and the size of the costly buffer. It is difficult to scale architectures with a single switching element to large numbers of interconnections. In this respect, multistage and parallel topologies have been adopted in many solutions. The fast reconfiguration of optical switches used for OPS and OBS has allowed for flexible interconnectivity, which is a desired feature for DC applications. Relatively lower power/cost efficiency is the price to pay compared with OCS technology, mainly due to the active components and the loss experienced in the switch fabrics. Performance improvement is expected with the maturing of fast optical switching technologies in combination with photonic integration.

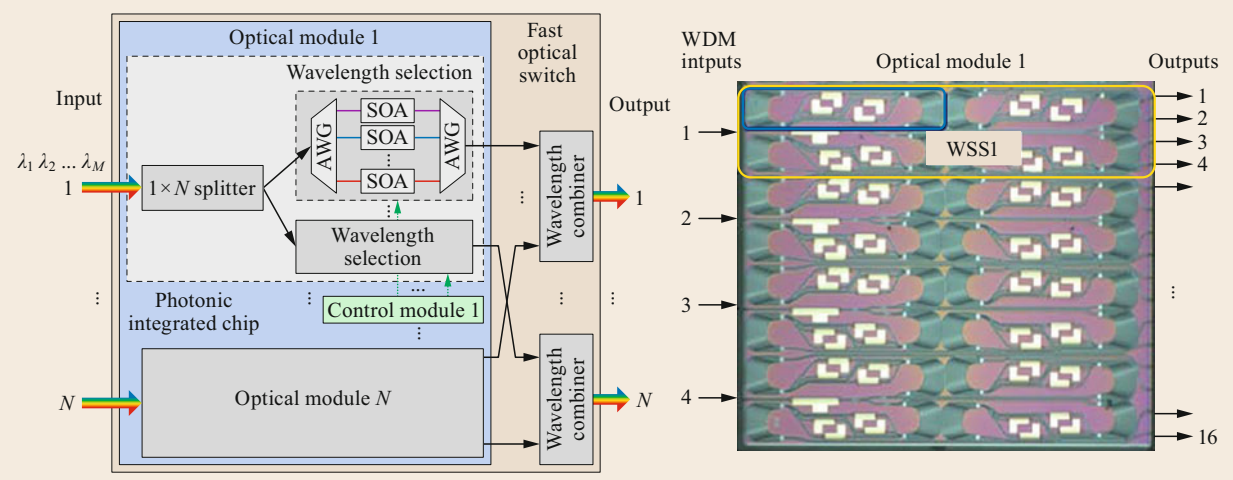


Fig. 25.27 Schematic of the fabricated 4×4 fast optical switch PIC. (Adapted from [25.79], © TU Eindhoven)

Table 25.3 Optical DCN architectures based on OPS and OBS technologies (TL: tunable laser; WC: wavelength converter; SpS: space switches; WTS: wavelength, time, space switch)

	Technology	Switching technology	Contention resolution	Scalability	Flexibility	Power/cost efficiency
IRIS	OPS	WC	ODL	Fair	Good	Fair
Petabit	OPS	TL+WC	EB at edge	Fair	Good	Fair
DOS (LIONS)	OPS	TL	EB	Poor	Good	Fair
TONAK LION	OPS	TL	EB at edge	Poor	Good	Fair
Hi-LION	OPS	TL	EB at edge	Good	Fair	Fair
Osmosis	OPS	SpS	EB at edge	Fair	Good	Fair
Data Vortex	OPS	SpS	Deflection	Fair	Good	Fair
OPSquare	OPS	WTS	EB at edge	Good	Good	Fair
OBS with fast optical switch	OBS	SpS	EB at edge	Poor	Fair	Good
Optical burst ring	OBS+EPS	TL	EB at edge	Fair	Fair	Good
HOS	OPS+OBS+OCS	WC+SpS	EB at edge	Fair	Good	Fair
HOSA	OBS+OCS	SpS	EB at edge	Fair	Fair	Good
Torus	OPS+OCS	SpS	Deflection+ODL+EB	Good	Good	Fair
Lightness	OPS+OCS	WTS	EB at edge	Good	Fair	Fair
Archon	OPS+OCS	SpS	EB at edge	Good	Fair	Fair
Burst-over-circuit	OBS+OCS	SpS	EB at edge	Poor	Fair	Good

25.4 Perspective on Optical Switching Technologies in Future DCs

Data center networks are designed to support the large data workloads exchanged between the parallel server machines. As discussed before, the amount of data exchanged within the DC (east–west traffic) is already growing, and it will continue to increase in the future. To cope with this traffic growth, future DC networks will have stringent requirements in terms of capacity, latency, scalability, flexibility, and power and cost efficiency. In particular, considering the increasing number of server nodes and data rates (and multilevel data format upgrades in the near future to sustain 400 Gb/s or even 1 Tb/s), the above-mentioned requirements will become quite challenging for current DCNs in terms of both switching node and network architectures. Indeed, the implementation of high-bandwidth electronic switch nodes is limited (to multi-Tb/s) by the switch ASIC I/O bandwidth due to the scaling limitations of the ball grid array (BGA) package [25.10]. Higher bandwidth is achievable by stacking several ASICs in a multitier structure, but at the expense of increased latency and higher cost. Another factor limiting electronic switch scaling is the power consumption of the OEO conversions and format-dependent interfaces that need to be further included as the front end, greatly deteriorating the energy- and cost-efficiency performance.

Switching data traffic in the optical domain can avoid power-consuming OEO conversions, along with the data latency associated with multiplexing and (de)serialization. Moreover, optical switching also eliminates the format- and modulation-dependent interfaces, achieving better efficiency and less complexity. Benefiting from their optical transparency, optical switching operation (including power consumption) is independent of the bit rate of the information carried. Scalability to higher bandwidth and employment of WDM technology can be supported seamlessly, enabling superior performance in terms of power per unit bandwidth. In this context, we have discussed various optical switching technologies for DC applications, which can be grouped into two classes: slow, high-radix optical switching fabrics (from tens of microseconds up to milliseconds) useful for optical circuit switching, and fast (few tens of nanoseconds up to one microsecond) optical switching fabrics for optical packet/burst switching. What is the perspective for deploying these optical switching technologies in future DC networks?

25.4.1 Optical Circuit Switching in DCNs

Optical circuit switch technology is now relatively mature and commercially available from several vendors to be employed in DCNs. The millisecond-scale reconfiguration times of OCS make them best suited to applications with well-scheduled and long-lived tasks, such as resource provisioning and optimization, monitoring and testing of network performance, and protection of equipment or fiber paths. An attractive application that has recently been demonstrated is the use of OCS for

dynamic interconnection of disaggregated server resources for rack-scale computing.

While OCS technology has been adopted in numerous DCNs for these purposes, there are still several challenges that need to be addressed for its wider deployment in future data center networks. Hitherto, the main barriers to entry have been the novelty and cost of the technology for use in DCNs. But the cost of current optical switches is now approaching that of electronic switches with 10-GbE interfaces, and the technology is becoming even more competitive with higher-data-rate ports ($> 100\,\mathrm{Gb/s}$), where the low latency of OCS becomes more and more attractive.

High-radix optical circuit switches have already been deployed with up to 384×384 nonblocking ports, and this size is expected to scale further in next-generation optical switch products. DC providers see an ultimate demand for networks to scale up to many thousands of ports, so that all nodes in a warehouse-scale data center can be connected without blocking. A single switch matrix with several thousand ports still presents a considerable technical challenge. Therefore, it is more likely that such networks will make use of scalable multiswitch architectures built from nonblocking elements of 100–500 port pairs and loss of around 1 dB. These switch sizes are likely to offer more attractive system reliability, serviceability, and first-in costs.

25.4.2 Fast Optical Switch Fabrics for Optical Packet/Burst Switching-Based DCNs

The advantage of fast optical switching to implement DC networks is the possibility to exploit full statistical multiplexing, be it in the wavelength, time, or space domain, in the same way as electronic switches. However, despite the considerable research works so far presented, this technology is still not mature and there are no commercially available switches with sufficiently large port count to be meaningful for DC networks. There are still several technological challenges that need to be addressed to realize commercial fast optical switch fabrics. First of all, even if a large-port-count switch fabric can be realized in InP or SiPh technologies, packaging such a switch with hundreds (or thousands) of electrical contacts to control the device is difficult. Many international and national pilot projects on packaging such devices have been initiated. It is therefore envisioned that, within 5–10 years, packaging solutions based on optical interposers, spot-size converters, and advanced flip-chip optoelectronics packaging will become available. The second important challenge is how to control the fabric in a sufficiently fast way to realize fully the benefit of statistical multiplexing. Novel protocols, fast scheduling, and fast optical flow control are under research in many universities worldwide. The control problem becomes more evident as future DCs scale to large numbers of servers. Therefore, it is reasonable that, once the packaged fast optical switches are commercially available, the fast switches will be employed in architectures that allow highly decentralized and parallel control with fast schedulers. Possible applications would be in intra- and interblades to interconnect $> 100\,\mathrm{Gb/s}$ microservers, and in disaggregated architectures interconnecting ultrahigh-bandwidth computing and memory nodes.

25.5 Conclusion and Discussion

The never-ending growth in demand for high bandwidth in data centers is accelerating the deployment of more powerful servers and more advanced optical interconnects. To accommodate the increasing volume of traffic with low communication latency and high power efficiency, technological and architectural innovations of DCNs are required. Optical circuit switching technologies are now starting to be deployed in leading-edge data center networks. Combining OCS and OPS/OBS based on fast optical switches could be an attractive solution as the technologies develop, by providing efficient statistical multiplexing and transparent high-capacity operation while eliminating the opaque front-ends created by OEO conversions. However, the lack of optical memory, the limited scalability due to the relatively low port count of fast OPS/OBS optical switches, and the inefficiency and lack of scalability of a centralized scheduler and control system capable of rapidly configuring the overall optical data plane are some of the practical hurdles to overcome to fully exploit the potential of optical switching technologies in DCNs. Solving those problems will require complete solutions from the DCN architecture down to the device technology. Promising results have been shown in recent investigations to solve those issues and pave the way towards mainstream deployment of optical switching technologies in next-generation data center networks.

References

25.1 A. Singh, J. Ong, A. Agarwal, G. Anderson, A. Armistead, R. Bannon, S. Boving, G. Desai, B. Felderman, P. Germano, A. Kanagala, J. Provost, J. Simmons, E. Tanda, J. Wanderer, U. Hölzle, S. Stuart, A. Vahdat: Jupiter rising: a decade of Clos topologies and centralized control in Google's datacenter network, ACM SIGCOMM Comput. Commun. Rev. **45**(4), 183–197 (2015)

25.2 A. Andreyev: Facebook's data center fabric, Networking @scale https://youtu.be/kcI3fGEait0 (Feb 2015)

25.3 C. Clos: A study of non-blocking switching networks, Bell Syst. Techn. J. **32**(2), 402 (1953)

25.4 L.A. Barroso, J. Clidaras, U. Hölzle: The datacenter as a computer: an introduction to the design of warehouse-scale machines, Synth. Lect. Comput. Archit. **8**(3), 1–154 (2013)

25.5 A. Greenberg, J.R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D.A. Maltz, P. Patel, S. Sengupta: VL2: A scalable and flexible data center network, ACM SIGCOMM Comput. Commun. Rev. **39**(4), 51–62 (2009)

25.6 Cisco: Cisco global cloud index: forecast and methodology 2014–2019. In: *Cisco Knowledge Network (CKN) Session* (2015)

25.7 A. Roy, H. Zeng, J. Bagga, G. Porter, A.C. Snoeren: Inside the social network's (datacenter) network. In: *SIGCOMM'15, London* (2015) p. 125

25.8 A. Greenburg, J. Hamilton, D.A. Maltz, P. Patel: The cost of a cloud: research problems in data center networks, ACM SIGCOMM Comput. Commun. Rev. **391**, 68–73 (2009)

25.9 A. Hammadi, L. Mhamdi: A survey on architectures and energy efficiency, Data Center Netw. Comput. Commun. **40**, 1–21 (2014)

25.10 A. Ghiasi: Large data centers interconnect bottlenecks, Opt. Express **23**(3), 2085–2090 (2015)

25.11 C. Kachris, I. Tomkos: A roadmap on optical interconnects in data centre networks. In: *Int. Conf. Transpar. Opt. Netw.* (2015), https://doi.org/10.1109/ICTON.2015.7193535

25.12 K. Aziz, M. Fayyaz: Optical interconnects for data center networks. In: *Handbook on Data Centers*, ed. by S.U. Khan, A.Y. Zomaya (Springer, Berlin, Heidelberg 2015) pp. 449–483

25.13 GigaCom Benelux, Averbode: Automating fiberspace, http://www.automatingfiberspace.com/

25.14 K. Watanabe, R. Kasahara, Y. Hashizume: Extremely-low-power-consumption thermo-optic switch with silicon-silica hybrid structure, NTT Tech. Rev. **8**(2), 1–5 (2010)

25.15 R. Ryf, J. Kim, J.P. Hickey, A. Gnauck, D. Carr, F. Pardo, C. Bolle, R. Frahm, N. Basavanhally, C. Yoh, D. Ramsey, R. Boie, R. George, J. Kraus, C. Lichtenwalner, R. Papazian, J. Gates, H.R. Shea, A. Gasparyan, V. Muratov, J.E. Griffith, J.A. Prybyla, S. Goyal, C.D. White, M.T. Lin, R. Ruel, C. Nijander, S. Arney, D.T. Neilson, D.J. Bishop: 1296-Port MEMS transparent optical crossconnect with 2.07 petabit/s switch capacity. In: *Opt. Fiber Commun. Conf. (OFC)* (2001), https://doi.org/10.1364/OFC.2001.PD28, Postdeadline paper: PD28

25.16 Y. Kawajiri, N. Nemoto, K. Hadama, Y. Ishii, M. Makihara, J. Yamaguchi, T. Yamamoto: 512 × 512 port 3D MEMS optical switch module with toroidal concave mirror, NTT Tech. Rev. **10**(11), 1–7 (2012)

25.17 N. Parsons, R. Jensen, A. Hughes: High radix all-optical switches for software-defined datacentre networks. In: *Eur. Conf. Opt. Commun. (ECOC)* (2016), Paper W.2.F.1 (Invited)

25.18 Polatis: http://www.polatis.com

25.19 Calient: http://www.calient.net

25.20 Wave2Wave: https://www.wave-2-wave.com

25.21 Epiphotonics: http://www.epiphotonics.com

25.22 T. Seok, N. Quack, S. Han, M. Wu: 50×50 digital silicon photonic switches with MEMS-actuated adiabatic couplers. In: *Opt. Fiber Commun. Conf.* (2015), https://doi.org/10.1364/OFC.2015.M2B.4, Paper M2B.4

25.23 H. Mulvad, A. Parker, B. King, D. Smith, M. Kovacs, S. Jain, J. Hayes, M. Petrovich, D. Richardson, N. Parsons: Beam-steering all-optical switch for multi-core fibers. In: *Opt. Fiber Commun. Conf.* (2017), https://doi.org/10.1364/OFC.2017.Tu2C.4, Paper Tu2C.4

25.24 S. Han, T.J. Seok, K. Yu, N. Quack, R.S. Muller, M.C. Wu: 50×50 Polarization-insensitive silicon photonic MEMS switches: design and experiment. In: *42nd Eur. Conf. Opt. Commun.* (2016), Paper Th.3.A.5

25.25 M. Iwama, M. Takahashi, M. Kimura, Y. Uchida, J. Hasesawa, R. Kawahara, N. Kagi: LCOS-based flexible grid 1×40 wavelength selective switch using planar lightwave circuit as spot size converter. In: *Opt. Fiber Commun. Conf.* (2015), https://doi.org/10.1364/OFC.2015.Tu3A.8, Paper Tu3A.8

25.26 K. Tanizawa, K. Suzuki, M. Toyama, M. Ohtsuka, N. Yokoyama, K. Matsumaro, M. Seki, K. Koshino, T. Sugaya, S. Suda, G. Cong, T. Kimura, K. Ikeda, S. Namiki, H. Kawashima: Ultra-compact 32 × 32 strictly-non-blocking Si-wire optical switch with fan-out LGA interposer, Opt. Express **23**, 17599–17606 (2015)

25.27 F. Testa, C.J. Oton, C. Kopp: Design and implementation of an integrated reconfigurable silicon photonics switch matrix in IRIS project, J. Sel. Top. Quantum Electron. **22**(6), 155–168 (2016)

25.28 H. Kouketsu, S. Kawasaki, N. Koyama, A. Takei, T. Taniguchi, Y. Matsushima, K. Utaka: High-speed and compact non-blocking 8×8 InAlGaAs/InAlAs Mach–Zehnder-type optical switch fabric. In: *Opt. Fiber Commun. Conf.* (2015), https://doi.org/10.1364/OFC.2014.M2K.3, Paper M2K.3

25.29 H. Okayama, M. Kawahara: Prototype 32×32 optical switch matrix, Electron. Lett. **30**(14), 1128–1129 (1994)

25.30 Y. Muranaka, T. Segawa, R. Takahashi: Integrated fat-tree optical switch with cascaded MZIs and

EAM-gate array. In: *21st OptoElectron. Commun. Conf.* (2016), Paper WF3-2
25.31 Y. Yin, R. Proietti, X. Ye, C.J. Nitta, V. Akella, S.J.B. Yoo: LIONS: An AWGR-based low-latency optical switch for high-performance computing and data centers, IEEE J. Sel. Top. Quantum Electron. **19**(2), 3600409 (2012)
25.32 Z. Cao, R. Proietti, S.J.B. Yoo: Hi-LION: Hierarchical large-scale interconnection optical network with AWGRs, J. Opt. Commun. Netw. **7**(1), A97–A105 (2015)
25.33 T. Segawa, M. Nada, M. Nakamura, Y. Suzaki, R. Takahashi: An 8×8 broadcast-and-select optical switch based on monolithically integrated EAM-gate array. In: *Eur. Conf. Opt. Commun. (ECOC)* (2013), Paper TuT4.2
25.34 R.P. Luijten, R. Grzybowski: The OSMOSIS optical packet switch for supercomputers. In: *Opt. Fiber Commun. Conf.* (2009), https://doi.org/10.1364/OFC.2009.OTuF3, Paper OTuF.3
25.35 H. Wang, A. Wonfor, K.A. Williams, R.V. Penty, I.H. White: Demonstration of a lossless monolithic 16×16 QW SOA switch. In: *35th Eur. Conf. Opt. Commun. (ECOC)* (2009), Supplement
25.36 T. Tanemura, I. Soganci, T. Oyama, T. Ohyama, S. Mino, K. Williams, N. Calabretta, H.J.S. Dorren, Y. Nakano: Large-capacity compact optical buffer based on InP integrated phased-array switch and coiled fiber delay lines, IEEE/OSA J. Lightwave Technol. **29**(4), 396–402 (2011)
25.37 R. Stabile, A. Albores-Mejia, A. Rohit, K.A. Williams: Integrated optical switch matrices for packet data networks, Microsyst. Nanoeng. **2**, 15042 (2016)
25.38 M. Glick, M. Dales, D. McAuley, T. Lin, K. Williams, R. Penty, I. White: SWIFT: a testbed with optically switched data paths for computing applications. In: *Proc. 7th Int. Conf. Transpar. Opt. Netw.* (2005), https://doi.org/10.1109/ICTON.2005.1506091
25.39 K. Barabash, Y. Ben-Itzhak, A. Levin, G. Landi, M. Biancani, L. Dittmann, J. Aznar: COSIGN: combining optics and SDN in next generation data centre network. In: *8th ACM Int. Syst. Storage Conf. (SYSTOR 2015), Haifa* (2015)
25.40 Open Networking Foundation: OpenFlow v1.4, https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.4.0.pdf
25.41 R. Enns, M. Bjorklund, J. Schoenwaelder, A. Bierman: Network Configuration Protocol (NETCONF), IETF RFC 6241 (2011)
25.42 A. Bierman, M. Bjorklund, K. Watsen: RESTCONF Protocol, IETF RFC 8040 (2017)
25.43 M. Bjorklund: YANG—a Data Modeling Language for the Network Configuration Protocol (NETCONF), IETF RFC 6020 (2010)
25.44 Open Network Operating System: https://onosproject.org/
25.45 S. Han, N. Egiy, A. Panda, S. Ratnasamy, G. Shiy, S. Shenker: Network support for resource disaggregation in next-generation datacenters. In: *Proc. 12th ACM Workshop Hot Top. Netw.* (2013), https://doi.org/10.1145/2535771.2535778
25.46 S. Kandula, S. Sengupta, A. Greenberg, P. Patel: The nature of datacenter traffic: measurements & analysis. In: *Proc. IMC* (2009), https://doi.org/10.1145/1644893.1644918
25.47 T. Benson, A. Akella, D.A. Matlz: Network traffic characteristics of data centers in the wild. In: *Proc. 10th ACM SIGCOMM Conf. Internet Meas.* (2010), https://doi.org/10.1145/1879141.1879175
25.48 N. Farrington, G. Porter, S. Radhakrishnan, H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, A. Vahdat: Helios: a hybrid electrical/optical switch architecture for modular data centers, in ACM SIGCOMM Comput. Commun. Rev. **40**(4), 32 (2010)
25.49 M. Channegowda, T. Vlachogiannis, R. Nejabat, D. Simeonidou: Optical flyways for handling elephant flows to improve big data performance in SDN-enabled datacenters. In: *Proc. Opt. Fiber Commun. Conf.* (2016), Paper W3F.2
25.50 G. Zervas, F. Jiang, Q. Chen, V. Mishra, H. Yuan, K. Katrinis, D. Syrivelis, A. Reale, D. Pnevmatikatos, M. Enrico, N. Parsons: Disaggregated compute, memory and network systems: a new era for optical data centre architectures. In: *Proc. Opt. Fiber Commun. Conf. (OFC)* (2017), Paper W3D.4
25.51 L. St Ville, A. O'Donnell, N. Parsons, I. Burnett: Fast packet switching in an optical time-multiplexed space switch. In: *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Vol. 1 (1991) pp. 173–176
25.52 S.J. Ben Yoo: Optical packet and burst switching technologies for the future photonic internet, J. Lightwave Technol. **24**(12), 4468–4492 (2006)
25.53 C. Kachris, I. Tomkos: A survey on optical interconnects for data centers, IEEE Commun. Surv. Tutor. **14**(4), 1021–1036 (2012)
25.54 C. Kachris, K. Bergman, I. Tomkos (Eds.): *Optical Interconnects for Future Data Center Networks* (Springer, New York 2013)
25.55 N. Calabretta, W. Miao, K. Mekonnen, K. Prifti, K. Williams: Monolithically integrated WDM cross-connect switch for high-performance optical data center networks. In: *Opt. Fiber Commun. Conf. (OFC)* (2017), Paper Tu3F.1
25.56 A. Singla, C.-Y. Hong, L. Popa, P.B. Godfrey: Jellyfish: networking data centers, randomly. In: *9th USENIX Symp. Netw. Syst. Design Implement.* (2012) pp. 225–238
25.57 D. Zhang, J. Wu, H. Guo, R. Hui: An optical switching architecture for intra data center interconnections with ultra-high scalability. In: *Opt. Interconnects Conf.* (2014), https://doi.org/10.1109/OIC.2014.6886080
25.58 G.M. Lee, B. Wydrowski, M. Zukerman, J.K. Choi, C.H. Foh: Performance evaluation of an optical hybrid switching system. GLOBECOM '03. In: *IEEE Global Telecommun. Conf.*, Vol. 5 (2003) pp. 2508–2512, https://doi.org/10.1109/GLOCOM.2003.1258689
25.59 A. Singla, A. Singh, K. Ramachandran, L. Xu, Y. Zhang: Proteus: a topology malleable data center network. In: *Hotnets-IX Proc. 9th ACM SIGCOMM Workshop* (2010), https://doi.org/10.1145/1868447.1868455

25.60 K. Sato, H. Hasegawa, T. Niwa, T. Watanabe: A large-scale wavelength routing optical switch for data center networks, Commun. Mag. IEEE **51**(9), 46–52 (2013), https://doi.org/10.1109/MCOM.2013.6588649

25.61 M. Fiorani, S. Aleksic, M. Casoni, L. Wosinska, J. Chen: Energy-efficient elastic optical interconnect architecture for data centers, IEEE Commun. Lett. **18**(9), 1531–1534 (2014)

25.62 M.C. Yuang, P.-L. Tien, H.-Y. Chen, W.-Z. Ruan, T.-K. Hsu, S. Zhong, J. Zhu, Y. Chen, J. Chen: OPMDC: architecture design and implementation of a new optical pyramid data center network, J. Lightwave Technol. **33**(10), 2019–2031 (2015)

25.63 A. Pal, K. Kant: RODA: A reconfigurable optical data center network architecture. In: *IEEE 40th Conf. Local Comput. Netw. (LCN)* (2015), https://doi.org/10.1109/LCN.2015.7366371

25.64 M. Imran, M. Collier, P. Landais, K. Katrinis: Software-defined optical burst switching for HPC and cloud computing data centers, J. Opt. Commun. Netw. **8**(8), 610–620 (2016)

25.65 C.Y. Li, N. Deng, M. Li, Q. Xue, P.K.A. Wai: Performance analysis and experimental demonstration of a novel network architecture using optical burst rings for interpod communications in data centers, IEEE J. Sel. Top. Quantum Electron. **19**(2), 3700508 (2013)

25.66 M. Imran, M. Collier, P. Landais, K. Katrinis: HOSA: hybrid optical switch architecture for data center networks. In: *Proc. 12th ACM Int. Conf. Comput. Front.* (2015), https://doi.org/10.1145/2742854.2742877

25.67 K.-I. Kitayama, Y.-C. Huang, Y. Yoshida, R. Takahashi, T. Segawa, S. Ibrahim, T. Nakahara, Y. Suzaki, M. Hayashitani, Y. Hasegawa, Y. Mizukoshi, A. Hiramatsu: Torus-topology data center network based on optical packet/agile circuit switching with intelligent flow management, J. Lightwave Technol. **33**(5), 1063–1071 (2015)

25.68 S. Peng, D. Simeonidou, G. Zervas, R. Nejabati, Y. Yan, Y. Shu, S. Spadaro, J. Perelló, F. Agraz, D. Careglio, H. Dorren, W. Miao, N. Calabretta, G. Bernini, N. Ciulli, J.C. Sancho, S. Iordache, Y. Becerra, M. Farreras, M. Biancani, A. Predieri, R. Proietti, Z. Cao, L. Liu, S.J.B. Yoo: A novel SDN enabled hybrid optical packet/circuit switched data centre network: The LIGHTNESS approach. In: *Eur. Conf. Netw. Commun. (EuCNC)* (2014), https://doi.org/10.1109/EuCNC.2014.6882622

25.69 S. Yan, E. Hugues-Salas, V.J.F. Rancaño, Y. Shu, G.M. Saridis, B.R. Rofoee, Y. Yan, A. Peters, S. Jain, T. May-Smith, P. Petropoulos, D.J. Richardson, G. Zervas, D. Simeonidou: Archon: A function programmable optical interconnect architecture for transparent intra and inter data center SDM/TDM/WDM networking, J. Lightwave Technol. **33**(8), 1586–1595 (2015)

25.70 Q. Huang, Y. Yeo, L. Zhou: Optical burst-over-circuit switching for multi-granularity traffic in data centers. In: *Opt. Fiber Commun. Conf* (2013), Paper OW3H.5

25.71 J. Gripp, J.E. Simsarian, J.D. LeGrange, P. Bernasconi, D.T. Neilson: Photonic terabit routers: the IRIS project. In: *Opt. Fiber Commun. Conf. (OFC)* (2012), https://doi.org/10.1364/OFC.2010.OThP3

25.72 K. Xi, Y.-H. Kao, H.J. Chao: A Petabit bufferless optical switch for data center networks. In: *Optical Interconnects for Future Data Center Networks*, ed. by C. Kachris, K. Bergman, I. Tomkos (Springer, Berlin, Heidelberg 2013) pp. 135–154

25.73 Z. Cao, R. Proietti, S.J.B. Yoo: Hi-LION: Hierarchical large-scale interconnection optical network with AWGRs, J. Opt. Commun. Netw. **7**(1), A97–A105 (2015)

25.74 Y. Yin, R. Proietti, X. Ye, C.J. Nitta, V. Akella, S.J.B. Yoo: LIONS: An AWGR-based low-latency optical switch for high-performance computing and data centers, IEEE J. Sel. Top. Quantum Electron. **19**(2), 3600409 (2012)

25.75 R. Proietti, Y. Yawei, Y. Runxiang, C.J. Nitta, V. Akella, C. Mineo, S.J.B. Yoo: Scalable optical interconnect architecture using AWGR-based TONAK LION switch with limited number of wavelengths, J. Lightwave Technol. **31**, 4087–4097 (2013)

25.76 R. Luijten, C. Minkenberg, R. Hemenway, M. Sauer, R. Grzybowski: Viable opto-electronic HPC interconnect fabrics. In: *Proc. 2005 ACM/IEEE Conf. Supercomput.* (2005), https://doi.org/10.1109/SC.2005.78

25.77 O. Liboiron-Ladouceur, A. Shacham, B.A. Small, B.G. Lee, H. Wang, C.P. Lai, A. Biberman, K. Bergman: The data vortex optical packet switched interconnection network, J. Lightwave Technol. **26**(13), 1777–1789 (2008)

25.78 H.J. Chao, K.-L. Deng, Z. Jing: PetaStar: a petabit photonic packet switch, IEEE J. Sel. Areas Commun. **21**(7), 1096–1112 (2003)

25.79 N. Calabretta, K. Williams, H. Dorren: Monolithically integrated WDM cross-connect switch for nanoseconds wavelength, space, and time switching. In: *Eur. Conf. Opt. Commun. (ECOC)* (2015), https://doi.org/10.1109/ECOC.2015.7341615

25.80 W. Miao, F. Yan, N. Calabretta: Towards petabit/s all-optical flat data center networks based on WDM optical cross-connect switches with flow control, J. Lightwave Technol. **34**(17), 4066–4075 (2016)

25.81 W. Miao, S. Di Lucente, J. Luo, H. Dorren, N. Calabretta: Low latency and efficient optical flow control for intra data center networks, Opt. Express **22**(1), 427–434 (2014)

25.82 W. Miao, F. Agraz, S. Peng, S. Spadaro, G. Bernini, J. Perello, G. Zervas, R. Nejabati, N. Ciulli, D. Simeonidou, H. Dorren, N. Calabretta: SDN-enabled OPS with QoS guarantee for reconfigurable virtual data center networks, IEEE/OSA J. Opt. Commun. Netw. **7**(7), 634–643 (2015)

Nick Parsons

HUBER+SUHNER Polatis
Cambridge, UK
nick.parsons@hubersuhner.com

Nick Parsons serves as SVP Engineering and Technology at HUBER+SUHNER Polatis, the leading manufacturer of all-optical switches. Previously, he held senior positions at the avionics divisions of BAE Systems and GEC-Marconi, focusing on optical communications and electrooptic sensors. Nick is a member of the IET and OSA and has coauthored over 50 papers and 20 patents in photonic systems and devices.

Nicola Calabretta

Dept. of Electrical Engineering
Technical University of Eindhove
Eindhoven, The Netherlands
n.calabretta@tue.nl

Nicola Calabretta is Associate Professor at the Technical University of Eindhoven. He has worked as a researcher with Sant'Anna University Pisa and Technical University of Denmark. He has coauthored over 320 journal articles and conference papers, and holds three patents. He has participated in several EU and national projects. His research interests include smart optical telecom and datatacom networks, and photonic integrated circuits.

Optical A

Part D

Part D Optical Access and Wireless Networks

Ed. by Massimo Tornatore

Part D of this Handbook covers optical communication and networking technologies in the access network segment (focusing on passive optical networks and mobile traffic backhauling) and recent trends in optical wireless communication (visible light communication and free-space optical links).

The access segment of a telecommunications network is the segment that connects subscribers to their immediate service provider, in contrast to the core network (covered in Part 2 of this Handbook), which connects service providers to one another. Access networks can offer wired connectivity (e.g., using twisted pairs, coaxial cables, or fibers) to fixed subscribers or wireless connectivity (e.g., using cellular base stations) to mobile subscribers. Part D first discusses fiber-based wired technologies for fixed subscriber access, but later chapters show that fiber-based connectivity is also the main infrastructure for aggregating and transporting mobile traffic.

Until the early 1990s, fixed access was dominated by copper-based solutions such as twisted pairs for traditional telephone networks (plain old telephone system, POTS) or coaxial cables for cable TV services. With the increasing adoption of broadband speeds, growing bandwidth demands called for fiber-to-the-home (FTTH) deployments, which started in the middle of the 1990s. Since then, optical access technologies have represented a growing share of fixed access lines.

The most important type of optical access network is the passive optical network (PON). PONs can deliver high data rates to network subscribers through a cost-effective tree-like fiber deployment that allows fiber capacity to be shared among a large number of subscribers. Several generations of PONs have already been standardized and deployed, but PONs are still evolving and adapting to new telecom service requirements. Chapters 26–30 provide the reader with a wealth of knowledge regarding this evolution.

Chapter 26 provides a gentle introduction to optical access technologies. It reviews the evolution of fixed access networks from the first deployment of PONs to the recent standardization of NG-PON2 and the latest developments aimed at preparing the fixed access segment to serve mobile traffic.

Chapter 27 delves into the details of the first generation of PON technology based on time-division multiplexing (TDM). After introducing the overall system architecture, the chapter addresses more advanced topics about the physical layer (such as burst mode transmission and analog video distribution) and the network layer (such as access control and bandwidth allocation and protection).

Chapter 28 extends the PON overview to other emerging multiplexing technologies. TDM is now being replaced with or joined by other multiplexing techniques (e.g., wavelength-division multiplexing) to cope with increasing subscriber bandwidth demands. The chapter starts by describing the second generation of PONs (NG-PON2), which extend first-generation PONs by adding WDM, and then moves on to describe experimental technologies (the third generation of PONs).

Other advanced technical proposals to evolve PONs (which are comprehensively described in Chapters 29 and 30) entail not only upgrades to transmission and multiplexing but also more disruptive evolutions of the classical tree-like PON architecture.

Chapter 30 starts from the assumption that a modern optical transmission network facilitates much longer transmission distances and represents an extended version of the traditional PON—a long-reach (LR) PON. LR-PONs are used in both the access and metro network segments, as they offer advantages in terms of infrastructure sharing, node consolidation, and network-layering simplification. Chapter 30 recalls the history and the design and operation principles of LR-PONs, and it reviews recent experimental demonstrations.

Similarly, Chapter 29 shows how next-generation access/metro networks can benefit from the improved networking capabilities of advanced PON systems, where the optical layer can be reconfigured and dynamically managed while maintaining the low-cost business cases that PONs were originally devised for. The solutions described in this chapter are almost entirely beyond what PON systems can offer today. However, most of them are currently being investigated in research or are already in development.

Aside from providing higher capacities for fixed subscribers, optical networks are also expected to be the main technology used to transport ('backhaul') mobile traffic, supporting growing data rates and denser radio access networks. Chapters 31 and 32 present two alternative solutions for the effective transport of mobile traffic. More precisely, Chapter 31 focuses on the transport of fronthaul traffic through digital radio over fiber (D-RoF). The term *fronthaul* refers to a new network segment that has recently been introduced in mobile networks to accommodate the high bandwidth generated when the two units of a radio base station (the remote radio head and the baseband unit) are split and located remotely. This new network segment has specific needs in terms of latency, jitter, and bandwidth that affect the specification of the fronthaul interface. Chapter 31 discusses the different switching and multiplexing options for D-RoF-based fronthaul.

Although D-RoF currently dominates the fronthaul segment, increasing bandwidth requirements of digital interfaces mean that, as capacity grows, ana-

log techniques may also become favorable. Therefore, in Chapter 32, the requirements and different optical configurations of analog RoF transport systems are reviewed.

An even more drastic increase in capacity and an enhanced ability of the network to adapt to different application needs are expected in the incoming 5G mobile communications. In Chapter 33, to highlight how optics can be used to support 5G communications, technologies used in fiber-wireless radio access networks (RANs) are reviewed, including mobile fronthaul evolution, all-spectrum fiber-wireless access technologies, and optical signal processing techniques.

Part D of this Handbook also covers two forms of optical transmission that are accomplished directly in the wireless domain. Chapter 34 describes the basic principles of and current trends in free-space optical communication (FSOC), a technology that can be used to increase satellite bandwidths while reducing the size, weight, and power of the system and taking advantage of a license-free spectrum. Chapter 35 discusses visible light communication (VLC), an optical transmission technique that transmits data by modulating the intensity of the light source. This chapter describes VLC devices (such as light sources and receivers) and technologies (such as the main modulation techniques), as well as current applications of VLC (such as indoor lighting, wireless local area networks (LANs), and underwater transmission).

Finally, Chapter 36 reviews fiber-based communication and sensing systems for avionics applications. The fly-by-light (FBL) approach to aircraft control is described. This is a representative and successful application of optical technology to avionics due to its light weight, compact size, high bandwidth, and immunity to interference. Various types of fiber-optic sensors for temperature, strain, pressure, vibration, and acoustic emissions are discussed, meaning that this emerging field of optical communications is well covered in this Handbook.

26. Introduction to Optical Access Networks

Björn Skubic, Lena Wosinska

Fixed-access networks have had a tremendous impact on society over the last few decades enabling residential broadband services and being a driver for the digitalization of society. With increasing broadband speeds, optical access technologies are playing an increasingly important role for fixed access. Growing capacity demand is driving deeper fiber penetration and fiber-to-the-home (FTH) deployments. An important category of optical access systems is passive optical networks (PONs). PON systems are designed to meet the requirements of access networks, supporting cost effective deployment and high-end user peak rates. Several generations of PONs have been specified both in ITU-T and IEEE. Deployed systems have predominantly been based on time division multiplexing (TDM)-PON. Recent standardization in ITU-T have specified next-generation (NG)-PON2, which is the first multi-wavelength access standard. Beyond higher capacity residential access, optical access is also expected to play an increasingly important role in providing transport services for mobile networks, supporting growing data rates and denser radio access networks. This introductory chapter on optical-access reviews the evolution of fixed-access network architecture and presents a technology overview of optical access systems.

26.1 Evolution of Fixed-Access Networks and Impact on Society

The evolution of fixed-access networks has had a large impact on society during the past decades, starting with basic Internet services in the 1990s and the dotcom boom in the late 1990s.

26.1.1 Fixed Broadband Emerges

In the early 2000s, digital subscriber line (DSL) technologies enabled wide-scale adoption of residential

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_26

broadband services with capacities beyond 1 Mbit/s. This enabled media transfer (music, video, etc.) over the residential broadband connection, whereas early Internet services consisted mainly in transfer of text and images. Internet protocol (IP) emerged as the principal communications protocol, enabling convergence of network infrastructure for delivering different services. The wide-scale adoption broadband access unleashed the popularity of content sharing (e.g., BitTorrent, Gnutella), which heavily impacted the media industry (music, film, and gaming).

Broadband connectivity became an increasingly important service offering for telecom operators. The Internet and the value it offered drove continued evolution in broadband access towards higher speeds. However, for many operators, revenues were still dominated by traditional telephony services, while traffic volumes and investment costs were now heavily impacted by broadband access. Furthermore, the emergence of over-the-top voice services (e.g., Skype), where a service was delivered by a third party over the Internet subscription, started to threaten the traditional telephony business. Declining revenues from traditional telephony services combined with deteriorating price per bit for broadband access raised concerns for operators. Several operators adopted strategies to provide more service-oriented offerings. The ambition was to transform the operator role from a *dumb pipe* provider with flat-rate pricing to a *smart pipe* provider with pricing based on service value. To capture some of the value growth, network operators also started offering content-delivery services.

IP convergence was not just a threat to operators. It was also exploited to simplify networks in the evolution from, e.g., asynchronous transfer mode (ATM). High-capacity broadband connectivity combined with IP convergence enabled operators to offer triple-play services, where telephony, Internet, and television services were bundled and delivered digitally to residential subscribers over a converged IP-based broadband connection. The different triple-play services have different characteristics and requirements in terms of bandwidth, latency, jitter, and burstiness. While Internet connectivity was delivered as best effort, traffic engineering with quality of service (QoS) policies, allowed operators to provide service guarantees for voice and video services. This enabled a differentiation compared to third-party services that were delivered over the Internet subscription.

26.1.2 Transformation of Society

The impact of broadband access on society is hard to overestimate. Broadband access has been a foundation for digitalization of the society. The term e-business was originally coined for business that in some form relied on the Internet. Today the Internet is integral for almost any business. It has transformed how information is accessed and has had an important role in globalization. Services such as health care, banking, travel arrangements, and retail are all migrating towards online services. Social media is also playing an increasingly important role in the modern society. Online marketing and Big Data have become important tools for a wide range of industries. Today, several of the highest valued companies worldwide (Google, Facebook, Amazon, etc.) are online companies [26.1]. The business landscape has changed, and new business models have emerged that capture the value enabled by broadband connectivity, although most of the profits were captured by over-the-top players, i.e., content providers or online platforms (such as search engines and social media platforms) rather than infrastructure and broadband connectivity providers.

Broadband access has been found to be a key driver for economic growth [26.2] and with digitalization, society is also becoming increasingly dependent on broadband access for critical services. This has spurred national and regional agendas/targets for deployment of broadband network infrastructure. According to the *European Commission* Digital Agenda [26.3], by 2020, all European households should have Internet access at bitrates of at least 30 Mbit/s, and 50% or more at bitrates above 100 Mbit/s. It has also led to regulations on requirements for coverage, reliability, and availability, as well as requirements on unbundling and competition. Figure 26.1 shows the historic average penetration rate of fixed broadband access in OECD countries, which in 2017 passed 30% [26.4]. Among the fixed broadband connections, in 2017 on average 22% were based on fiber. However, regional differences are large, e.g., in countries such as Japan and South Korea, more than 75% of the fixed broadband connections are fiber based.

26.1.3 From Triple Play to IoT

Bandwidth requirements for residential broadband connectivity have continued to grow to today's technology capabilities of > 100 Mbit/s. Emerging services, such as ultra-high definition (UHD) TV, augmented reality (AR), and virtual reality (VR) will continue to drive bandwidth requirements beyond 100 Mbit/s. In parallel to the development of higher-speed residential access, broadband connectivity has evolved from interconnecting homes (with fixed broadband), to connecting people (with mobile broadband) and is now extending further to interconnecting devices, referred to as the Internet of things (IoT). Device-to-device and machine-to-machine connectivity is gaining new application areas such as

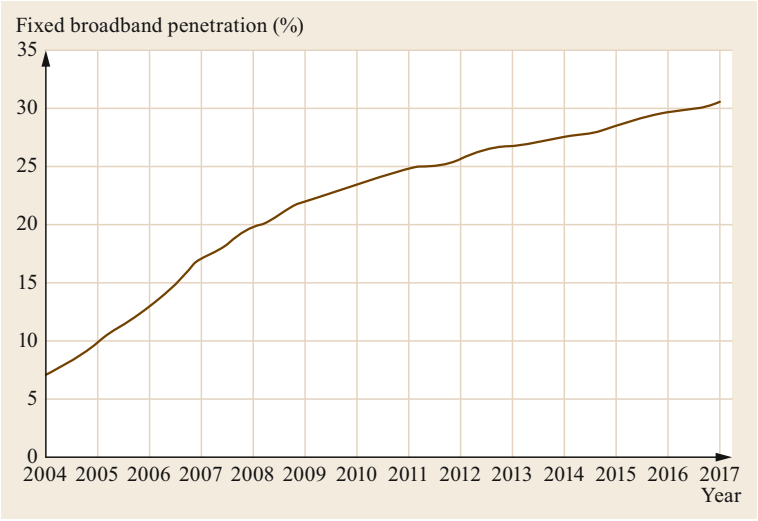


Fig. 26.1 Average fixed broadband penetration in OECD countries from 2004 to 2017 (after [26.4])

smart homes, connected vehicles, etc. and is bringing new challenges to the access networks. Connectivity is also seen to play an increasingly important role in the process and manufacturing industry. Industry 4.0 denotes the ongoing transformation towards increased exploitation of automation and data exchange within manufacturing.

Although, in both mobile broadband and IoT, the final access link will in most cases be wireless, fixed-access systems will continue to be a key part of the broadband infrastructure as an enabler for both mobile broadband and IoT. As a key part of a converged infrastructure, fixed-access systems will need to satisfy an even wider range of requirements on capacity ($> 10\,\text{Gbit/s}$), latency ($< 1\,\text{ms}$), and reliability (> 99.999% connection availability), with requirements coming from a wide range of end-user services as well as mobile transport.

26.1.4 Mobile Broadband and Convergence

Mobile communication networks have evolved several generations since they were first introduced in the 1970s. From the plain voice and text services in the first and second generation (1G and 2G) networks, today's 3G and 4G networks have successfully shifted to data centric services. The evolution of radio access networks (RANs) continues with emerging 5G standards. Mobile data traffic has increased 1000 times during the past few years [26.5], and this trend is not going to change. By 2020, the mobile traffic volumes are expected to grow rapidly, and the number of connected devices will be 10–100 times higher than today [26.6].

The next generation mobile network, referred to as 5G [26.6, 7], aims to address these issues and provide unlimited access to information for people and a large variety of connected devices. 5G deployment will increasingly rely on optical access technologies. Increasing traffic demand in the cellular networks leads to a need for densification of the mobile base stations, which in turn need to be interconnected by ultra-high capacity, low latency, and highly reliable links. This evolution brings new challenges to the access network segment. So far, the fixed and mobile access networks have evolved independently. The network infrastructure, equipment, topology, functions, and network requirements are largely different, which has led to independent investments and network operation. However, with the evolution in mobile networks towards higher capacity and site density, there has been an increasing interest in converged network solutions. Such converged network infrastructure would provide broadband access for fixed users as well as serve as a transport network for mobile networks.

26.1.5 Impact of Datacom on Telecom Networks

A major issue in the roll-out and build-out of broadband access infrastructure has been that revenues are concentrated to major over-the-top players, while infrastructure providers are facing large deployment and operational costs for building out and running the broadband access infrastructure. The access segment is highly cost sensitive, and cost is an important parameter for understanding the evolution of fixed-access systems and architectures. In the fixed-access community, different schemes for saving costs have been explored including architectural transformation through node consolidation [26.8], as well as convergence op-

portunities between fixed and mobile networks [26.9]. Recently, there has also been an increasing interest in network function virtualization (NFV), software defined networking (SDN), and open interfaces in the access segment. Trends from the datacom area with software routers/switches and generic network hardware, i.e., white boxes, are expected to increasingly impact telecom networks. Functions that today are implemented in dedicated network hardware (HW) (e.g., ASIC and FPGA) may, to an increasing degree, be virtualized, implemented in software (SW), and hosted in merchant silicon. The trend implies that network functions will, to a larger extent, be implemented in SW, while HW will be more generic. The separation of SW from HW allows the two to evolve with different life cycles facilitating SW upgrades and adding new SW features with shorter time-to-market. Combined with open interfaces, virtualization is also expected to affect the supplier ecosystems. Open interfaces are considered to foster innovation, where application developers are free to innovate on top of the interfaces. The network is increasingly seen as a platform for application developers to develop network applications. However, there is also concern that open interfaces may inhibit innovation across interfaces, given the resistance of modifying interfaces once defined. A major concern with virtualization is performance, since performance-critical functions are, in general, more efficiently implemented in dedicated HW optimized for a particular function, e.g., packet processing, baseband processing, encryption/decryption, etc.

Edge computing has emerged as an important concept in both fixed and mobile access [26.10, 11]. The trends of virtualization in telecommunication networks will in the long run lead to large-scale distributed cloud deployments. Such infrastructure could be exploited for other services besides telco workloads. The promise of edge computing is that alternative services can benefit from being hosted in a highly distributed edge cloud infrastructure. These include services that rely on extremely low latency control loops or services collecting massive amount of data that needs to be filtered at the edge of the network. For operators, the concept of the edge cloud is important, as it presents a vision where access networks and their proximity to customers will be of growing significance in future revenue generation. The edge cloud enables service differentiation compared to over-the-top providers, and access sites are increasingly seen as assets rather than a pure cost driver in the network infrastructure.

26.2 Access Network Architectures: Evolution and Trends

In many cases, fixed-access network architectures evolved starting from existing telephony networks, where fixed broadband access services are provided exploiting existing copper plants originally deployed for plain old telephony services (POTS).

26.2.1 Early Copper-Based Architectures

Copper cabling was deployed in the access segment between telecom central offices (CO) and subscriber premises. A brief introduction to copper access is important for understanding the evolution of fixed-access architectures. Several generations of ITU-T specifications for DSL provided increasing capacity of broadband access over twisted pair cables. DSL broadband services are provided over the copper access loop between a DSL access multiplexer (DSLAM), typically at the operator central office and a DSL modem at the customer premises. Asymmetric DSL (ADSL) technologies with ADSL, ADSL2, and ADSL2+ brought access capacity per subscriber in the Mbit/s range, with ADSL2+ providing a maximum subscriber rate of 24 Mbit/s downstream and 3.3 Mbit/s upstream. However, DSL subscriber rates are highly dependent on the length of the access loop, i.e., the cabling distance between the CO and the subscriber. Therefore, the distance to the CO determined which broadband services could be supported. For ADSL2+, the subscriber peak rate drops below 20 Mbit/s at 2 km distance and is further reduced to approximately 1 Mbit/s at 5 km. For POTS, the maximum loop length was 5 km using standard wiring. Despite dependence on distance, the capabilities of ADSL2+ enabled massive roll-out of fixed broadband services.

To meet the demand for even higher bandwidth, very-high-bitrate DSL (VDSL) and VDSL2 were specified in ITU-T. VDSL2 provides 100 Mbit/s symmetrical bitrates at source, which deteriorate to less than 50 Mbit/s at 0.5 km. Given the limited reach of VDSL2, in many cases it could not provide adequate service over the entire central office service areas. Techniques extending the reach of VDSL2 were developed, such as vectoring (noise cancelation techniques) and bonding (using two telephone lines available in many homes). These techniques could be used in some cases, while in other cases, active remote nodes were deployed to provide a termination point for VDSL2 located closer to the subscriber. This implied a densification of the

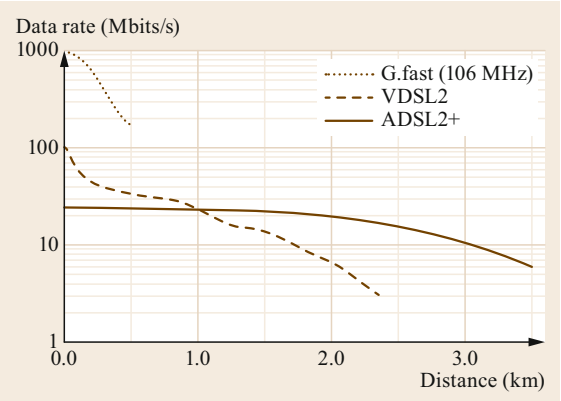


Fig. 26.2 Downstream maximum bitrate of ADSL2+, VDSL2, G.fast, and dependence on the length of the access loop from different measurements (after [26.12–14])

fixed-access network with active remote nodes in the CO service area.

Recently, a new generation of DSL was specified in ITU-T, referred to as G.fast. G.fast was developed for access loops of less than 0.5 km with a target bandwidth of 1 Gbit/s–100 Mbit/s, depending on the loop length. Due to the shorter reach, G.fast requires access sites that are closer to the customer than previous DSL generations. These sites should typically be served by fiber, resulting in fiber-to-the-x (FTTx) deployment scenarios with deeper fiber penetration. An advantage of DSL technologies is the ability to reuse existing copper-based infrastructure, where the last part of the access loop connecting to the subscriber is often the most costly to upgrade to fiber. DSL technologies have enabled reuse of existing infrastructure, reducing or postponing the need for costly fiber infrastructure deployment.

Figure 26.2 shows the maximum downstream bitrate for ADSL2+, VDSL2, G.fast, and dependence on the length of the access loop. The data in Fig. 26.2 is compiled from different measurements, as the relation between data rate and loop length depends on the quality of the copper plant and differs in different deployments.

26.2.2 Higher Speeds and Deeper Fiber Penetration

The architectural implication of increasing bandwidth in copper-based access networks has been a need to reduce the copper loop length, introducing active remote nodes at key distribution points (e.g., cabinets) in the CO service area. In many cases, the active remote nodes have been served by fiber-based backhaul to cater for aggregated bandwidth requirements, which in combination with increasing capacity demand per user has been driving deeper fiber penetration in the service areas [26.15–18]. Figure 26.3 shows a typical access/metro site hierarchy consisting of the home, building, cabinet, central office (CO), main CO, and core CO. The naming of the sites, as well as the number of levels in the hierarchy may differ for different countries/operators/areas. Fiber-based access networks can be categorized according to the degree of fiber penetration as: fiber-to-the-exchange (FTTEx), fiber-

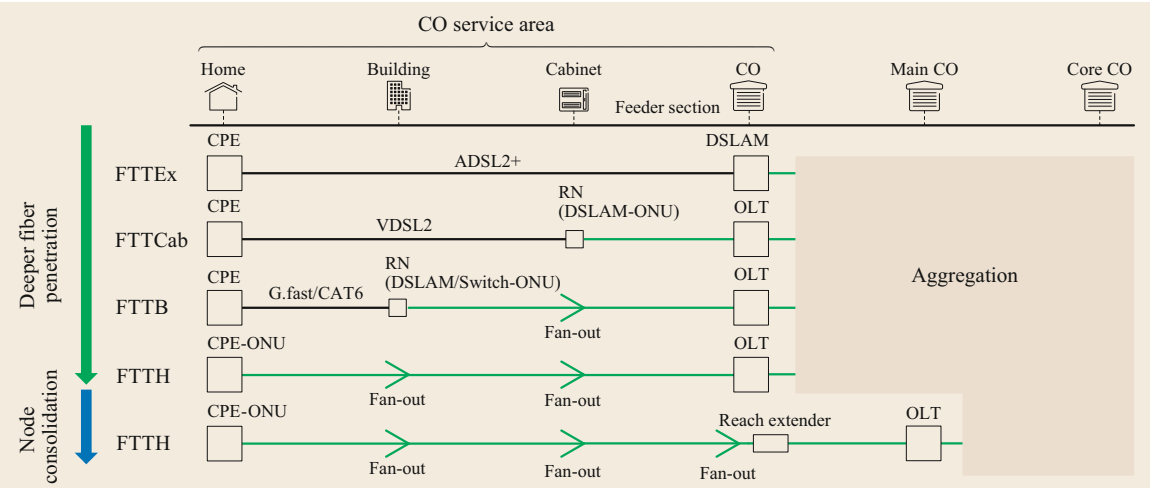


Fig. 26.3 Fixed-access evolution towards higher access bitrates and deeper fiber penetration. The bottom scenario shows node consolidation exploiting long reach optical access technologies for reducing the number of COs. (Abbreviations: customer premises equipment (CPE), DSL access multiplexer (DSLAM), optical line terminal (OLT), optical network unit (ONU), remote node (RN))

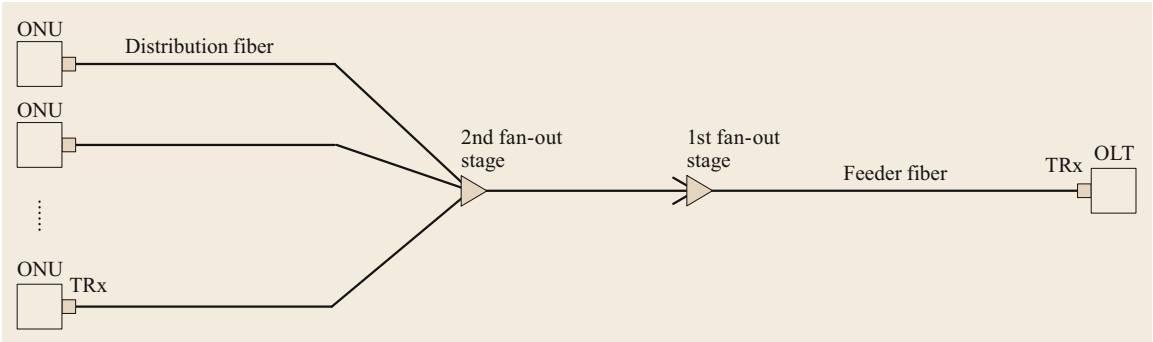


Fig. 26.4 Generic PON structure

to-the-cabinet (FTTCab), fiber-to-the-building (FTTB), and fiber-to-the-home (FTTH). The latter three categories are jointly referred to as FTTx. The architectural evolution towards higher access bitrates with deeper fiber penetration is illustrated in Fig. 26.3. Note that the optical access systems consist of an optical line terminal (OLT), typically placed at the CO, and optical network units (ONU), which are placed either at customer premises or at distributed access sites. The optical distribution network (ODN) is the fiber-based infrastructure that interconnects the OLT and ONUs.

26.2.3 FTTH: A Future Proof Infrastructure

Fiber supports higher bandwidth and longer reach compared to copper technologies. For this reason, FTTH infrastructure is often considered as the future-proof option for higher access bandwidths to come. Not only does FTTH provide support for higher bandwidths, it also reduces the need for active remote nodes in the network. Hence, the choice between FTTH and other variants of FTTx depends on a trade-off between fiber deployment costs and costs associated with active remote nodes. Despite being recognized as future proof, massive deployments of FTTH have in many areas been much delayed due to the high investment cost of fiber infrastructure. Regional differences are, however, quite large, with significant deployments in Asia, while regions such as Europe, to a large extent, are still exploiting existing copper plants.

The ODN of an FTTH deployment depends on the optical access technology. The structure of the ODN is important, as it may constrict or affect costs for future upgrades to higher capacity. An important group of solutions for large-scale deployments of optical access are PONs. There are several different types of PONs, but generally PONs are designed to save fiber in the feeder section (Fig. 26.3) without introducing costly active remote nodes. Different variants of PONs are reviewed in Chaps. 27 and 28.

Generally, a PON consists of multiple ONUs connected to an OLT via an ODN (Fig. 26.4). The PON ODN is commonly deployed in a tree-like topology with one or several stages of optical power splitters and/or wavelength filters. In the FTTH architecture, the ONUs are located at the customer premises, while the OLT is typically hosted at the CO. The ODN splitters or filters are located at distribution points in the CO service area (in Fig. 26.4, the *fan-out stages*). In PONs, the ODN is a shared transmission medium, and depending on the type of PON, different mechanisms are employed for sharing of the medium (i.e., time domain, wavelength domain, etc.). The ODN is a long-term investment that should support future technology generations. Considering the existing base of TDM-PON deployments (i.e., EPON and GPON), the most common ODN infrastructure today is a splitter-based PON ODN.

Besides splitter-based ODNs, other possible ODN deployments are point-to-point (PtP) fiber and filter-based PON ODNs. For splitter-based ODNs, careful planning of the splitter stages and splitter locations is needed to facilitate upgrades with increasing traffic and take rates.

26.2.4 Node Consolidation

With the transition to FTTH, some operators are considering exploiting the reach of optical technologies for consolidation of COs to a fewer number of COs with larger service areas (Fig. 26.3). A reduction in the number of CO sites may provide significant savings in fixed-access networks. However, node consolidation also leads to increased requirements on the optical access systems (longer reach, higher fan-out, higher power budget). Although optical communications support long reach, there is an impact on transceiver complexity (challenges with long-reach access systems are discussed in Chap. 30). Whether or not node consolidation is favorable depends on a detailed trade-off

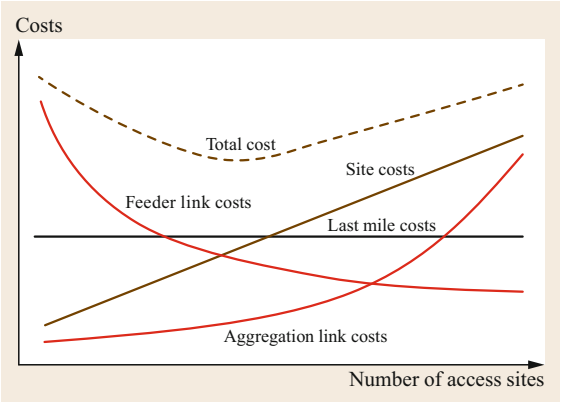


Fig. 26.5 The cost structure of the optical access with a sweet spot for node consolidation (after [26.19])

between site costs, fiber infrastructure costs, and technology costs for different scenarios. Plans for node consolidation should also consider the risks of higher upgrade costs when migrating to future optical access technologies, as reach typically deteriorates with higher bitrates. Figure 26.5 shows a schematic illustration of the cost structure in the access depending on the number of access sites. With node consolidation, site costs decrease due to fewer access sites. Aggregation links, which provide backhaul to access sites, will require higher capacity but will be both fewer and shorter, which results in reduced costs. Feeder links, i.e., the fibers between the traditional service area and the consolidated access sites, become longer with node consolidation, resulting in increased costs. Combined, these cost dependencies result in a sweet spot for a certain density of access sites.

26.2.5 Mobile Networks and New Requirements for Mobile Transport

Mobile transport requirements are expected to become increasingly important for fixed-access systems. However, until today, fixed-access systems have evolved without major impact from mobile access. Global mobile data traffic was 7% of the total IP traffic in 2016 and will be 17% of the total IP traffic by 2021 [26.20]. The number of mobile cell sites per CO service area are relatively few [26.21]. Hence, fixed-access systems were designed primarily for cost effective residential access services. Instead, dedicated systems and technologies were used for mobile backhaul including, microwave links, copper technologies, PtP fiber, and coarse WDM (CWDM) links. However, mobile networks are experiencing rapid growth in traffic. It is expected that fixed-access systems will need to provide the transport service for mobile networks to a greater extent. Organizations like next-generation mobile networks (NGMN) [26.22] and 5G infrastructure public private partnership (5G-PPP) [26.23] have defined aggressive performance targets for the 5G systems, including access bitrates up to 10 Gbit/s [26.7]. Those targets bring challenges for both the wireless technology and the wired backhaul/fronthaul/midhaul networks, referred to as mobile transport networks. To cater for this growth, not only the capacity per mobile cell site needs to increase (e.g., wider spectrum, beamforming, etc.), but also the number of mobile cell sites needs to increase (e.g., small cells). There are significant costs in adding new mobile cell sites as well as providing connectivity to such sites. Studies have shown that the significant cost of providing connectivity to a cell site may result in a situation where it is more cost effective to utilize the existing fixed-access network infrastructure and deploy a larger number of cell sites at locations where connectivity is available, rather than deploying a fewer number of cell sites at optimal radio locations, whilst extending connectivity to these locations [26.24]. Hence, the cost of mobile transport in future mobile networks may drive convergence between fixed-access and mobile transport, as discussed in Sect. 26.2.6.

Note also that one trend in mobile network evolution that impacts mobile transport is the increasing need for RAN coordination for denser networks. RAN coordination is the coordination of radio signals of neighboring mobile cells within a geographical area in order to improve radio performance or mitigate interference between cells. Several protocols have been proposed, such as, e.g., the coordinate multipoint protocol (CoMP) that will be used as an example below. RAN coordination typically puts requirements on the connectivity between cells in the coordination group. As shown in Table 26.1, there are multiple schemes, ranging from moderate coordination to very tight coordination, and these schemes result in a wide range of requirements on the transport network in terms of both required transport capacity and delay class (i.e., maximum latency between cells in a coordination group). Whereas moderate coordination is generally supported, tight and very tight coordination will have direct impact on the design of the mobile transport network, as they require high capacity and enforce very strict latency requirements.

Another aspect of RAN affecting mobile transport networks is the RAN deployment model. Traditional LTE deployment consists of eNodeB base stations at the cell sites with conventional S1 backhaul to the mobile core network. Internally, the eNodeB consists of digital units and radio units, separated by an interface, typically the common public radio interface (CPRI). These units may be integrated or separately deployed

Table 26.1 RAN coordination schemes associated capacity gain and requirements on latency between cells subject to coordination in long-term evolution (LTE) [26.21]

Coordination classification	Coordination feature	Max capacity gain[a]	Delay class
Very tight coordination	Fast UL CoMP (UL joint reception/selection)	High	0.1–0.5 ms
	Fast DL CoMP (coordinated link adaptation, coordinated scheduling, coordinated beamforming, dynamic point selection)	Medium	
	Combined cell	Medium	
Tight coordination	Slow UL CoMP	Small	1–20 ms
	Slow DL CoMP (e.g., postponed dynamic point blanking)	Small	
Moderate coordination	FeICIC	Small	20–50 ms

[a] Small: $\leq 20\%$, Medium: 20–50%, High: $\geq 50\%$

such that, e.g., the radio units are deployed in the cell towers close to the antennas, while the digital units are deployed at the base of the towers. An alternative deployment model, referred to as centralized-RAN (C-RAN), is to deploy the digital units at a central location enabling pooling of resources between cell sites and facilitating RAN coordination by co-location of digital units that belong to a common coordination group. This type of deployment implies that the mobile transport network needs to carry CPRI data between the central location and the cell sites, i.e., digital fronthaul. Another deployment model that has been proposed is analogue fronthaul, where, essentially, even the radio units are deployed at the central location. This reduces the cell site footprint but means that analogue signals need to be carried between the central location and the cell sites. With 5G, additional functional decompositions of the RAN have been considered, where internal RAN functions are grouped differently compared to today, resulting in new types of RAN nodes and new interfaces between nodes, resulting in different requirements on mobile transport networks. Details of digital and analogue fronthaul techniques are presented in Chaps. 31 and 32.

26.2.6 Convergence Between Fixed and Mobile Access Networks

Conditions for convergence between fixed and mobile networks depend both on convergence of sites, regarding placement of nodes and functions in the network, as well as on convergence of links, regarding transport/connectivity requirements between the sites. For example, stringent latency requirements in RAN deployments supporting tight coordination or mobile fronthaul impose requirements on the mobile transport as well as constrict the degree of node consolidation. This is because in these cases certain RAN functions need to be hosted sufficiently close to one or several cell sites. In a converged scenario, limitations to node consolidation in the mobile network will also impact the degree of node consolidation that can be achieved in the fixed-access network. Hence, the feasibility of convergence depends on whether an attractive compromise between fixed and mobile networks can be found.

Looking at the characteristics of today's mobile network architecture and the opportunities for node consolidation and convergence, Fig. 26.6 shows an illustration of the fraction of cell sites within 40 km distance of different types of CO sites. The distance 40 km corresponds to a round-trip time (RTT) of 0.4 ms, consistent with flavors of very tight RAN coordination. This shows that site consolidation to the CO or main CO level is feasible in the RAN, which means that converged architectures based on consolidation to these particular sites are possible.

Figure 26.7 shows the different types of convergence scenarios on the link level, exemplified with different systems (convergence happens on the optical system level, as the same optical system is used for fixed-access and mobile transport). This scenario provides maximum infrastructure reuse but may not be feasible if there are significant differences in the transport requirements. Fixed-access systems can generally provide support for basic mobile backhaul services, e.g., TDM-PON systems can support basic mobile backhaul services using static bandwidth allocation. Emerging deployment scenarios with the need for a tight RAN coordination and mobile fronthaul cannot, however, be supported by typical TDM-PON systems. Figure 26.7b shows convergence on the ODN level, where different optical systems/technologies are used for fixed-access and mobile transport exploiting the same ODN. This form of convergence enables sharing of fiber infrastructure, while sharing of active equipment is limited. Stringent mobile transport requirements were one of the

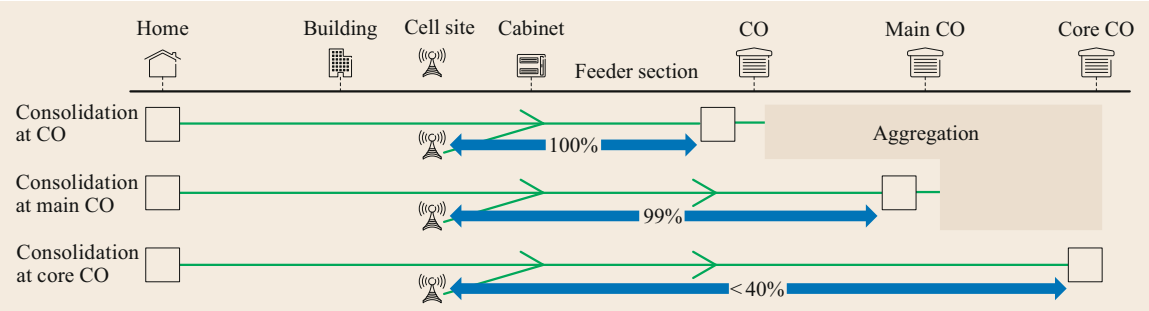


Fig. 26.6 Fraction of cell sites within 40 km distance to different CO sites in a typical European deployment (after [26.9])

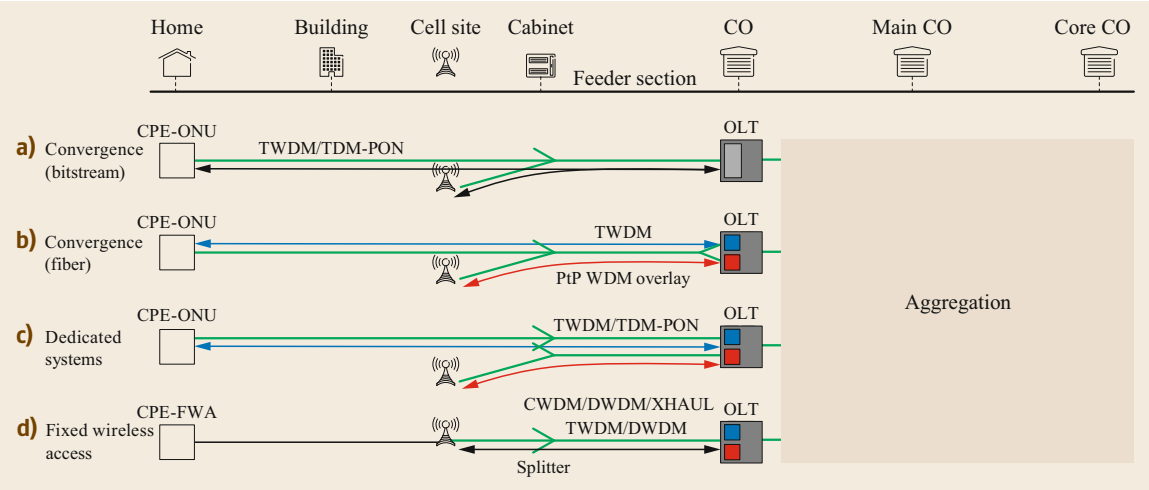


Fig. 26.7a–d Fixed mobile convergence scenarios with **(a)** optical system level convergence, **(b)** ODN convergence, **(c)** dedicated solutions, and **(d)** fixed wireless access. Abbreviations: dense WDM (DWDM), time and wavelength division multiplexing (TWDM)

reasons for introducing PtP WDM overlay support in NG-PON2, as discussed in Sect. 26.3.3. PtP WDM provides dedicated wavelength channels for high capacity and low latency transport that could be used for mobile transport, while fixed-access users are served by separate TDM wavelengths. One issue in this form of convergence is the additional complexity and optical power loss resulting from the convergence filters required to bring wavelengths for different services onto the same ODN. Figure 26.7c shows a scenario with dedicated solutions tailored for fixed-access and mobile transport, respectively. Sites and fiber ducts are shared between fixed and mobile deployments, while optical systems, including fiber infrastructure and active elements are dedicated. Figure 26.7d shows another form of convergence that will be commented on in the next section. Studies have shown that the optimum degree of convergence depends on a range of parameters, such as the subscriber density, density of small cells, and fiber availability, etc. [26.9].

26.2.7 Fixed Wireless Access: Integration of Fixed and Wireless Networks

The concept of fixed wireless access is based on exploiting wireless technologies for providing fixed-access services. Hence, instead of copper or fiber-based access, wireless technologies are exploited in the last mile such that each subscriber receives the fixed-access service through a wireless modem. Some of these fiber-wireless architectures are reviewed in Chap. 33. Conventional wireless technologies are limited in either capacity or reach. However, the evolution toward 5G has the potential to take fixed wireless access (FWA) to a completely new level, exploiting wireless technologies as an integral part of fixed access. Already today, in LTE with 40 MHz of bandwidth, there is often a working business case for FWA as an add-on improvement to fixed access [26.25]. 5G offers technology options that make it possible to use larger chunks of radio spectrum and provide consumers with

high bandwidth and low latency (1 ms). Compared with FTTH and other wireline solutions, FWA offers a variety of benefits, including significantly lower roll-out costs and rapid service rollout. This is because the bulk of the costs and most of the complexity involved in fixed-access deployments are associated with the last segment that reaches the user premises. The FWA scenario is depicted in Fig. 26.7d and can be seen as an intermediate step between FTTCab and FTTH.

26.3 Optical Access Systems/Technologies and Standards

There is a range of optical systems and technologies for optical access networks, from systems/technologies that are already standardized and deployed to research concepts that have been demonstrated or proposed for future access. Most of today's PON deployments belong to the family of time division multiplexing (TDM)-PONs. Several generations of TDM-PONs have been standardized in ITU-T and IEEE, as will be presented in Chap. 27. ITU-T standards include G-PON [26.26] and 10-gigabit-capable PON (XG-PON) [26.27]. IEEE standards include EPON/GE-PON [26.28] and 10G-EPON [26.29].

26.3.1 Legacy Systems

The standards consist of both a physical layer specification as well as a media access control (MAC) layer specification. The typical TDM-PON ODN was already shown in Fig. 26.4. Existing standards assume a single fiber working with one wavelength for downstream transmission and a second wavelength for upstream transmission. The downstream is based on TDM and the upstream is based on TDMA (time division multiple access). The theoretical user peak rate is equal to the PON line rate but the available PON capacity is shared through statistical multiplexing. Transmission is based on non-return-to-zero on-off-keying (NRZ-OOK), which allows for sufficient system reach at limited transceiver complexity. In TDM-PON systems, there is a dependency between system reach, fan-out, and subscriber rate that needs to be considered in deployment dimensioning. The available optical power budget, which is the difference between transmitter output power and receiver sensitivity, limits combined fan-out and reach. The ODN insertion loss increases with higher fan-out, leading to less power budget remaining for system reach. Higher fan-out also means more users sharing the total capacity of the PON, resulting in a higher degree of statistical multiplexing and lower average bandwidth per user. Typical deployments have an ODN fan-out of 1 : 32 or 1 : 64, but standards also provide support for higher fan-out up to 1 : 256. The physical layer specifications of PONs define several optical power budget classes. Higher optical power budget, through higher transmitter output power and higher receiver sensitivity, provides the better performance (in terms of combined fan-out and reach) required in certain deployments but also needs more complex and costly transceivers.

The MAC layer specifications of TDM-PONs define how the shared wavelength medium is accessed. As shown in Fig. 26.8, downstream transmission is based on broadcast-and-select, while upstream transmission is handled through time slot scheduling. Statistical multiplexing in the upstream direction is enabled through dynamic bandwidth assignment (DBA) where the size of the transmission slots for different ONUs are scheduled based on traffic demand and other policy parameters.

TDM-PONs provide an efficient solution for fixed access, as the design allows multiple ONU clients to share a single OLT transceiver interface exploiting statistical multiplexing. This limits the number of OLT transceivers needed in a network and reduces footprint at the CO. TDM-PONs exploit the bursty nature of access traffic, whilst still supporting high peak access rates. A drawback of TDM-PONs is the lack of support for low-latency services. Latency sensitive upstream traffic cannot be subject to DBA and must be handled through static bandwidth allocation, which eliminates statistical multiplexing gains in the upstream. This is problematic if a larger portion of the overall traffic is latency sensitive. More details will be provided in Chap. 27.

Figure 26.9 shows a range of optical access solutions, including legacy systems as well as futuristic concepts. Besides TDM-PON, other legacy options are those based on PtP fiber. PtP fiber architectures may exploit low-cost grey transceivers, i.e., the low-cost transceivers used in standard PtP fiber communication on typical wavelengths, such as 850, 1300, 1310, or 1550 nm. However, there are also drawbacks with such architectures. The flavor referred to as Homerun (Fig. 26.9) is based on PtP fiber links between the subscriber and the CO. A drawback of Homerun is the excessive amount of fiber required in the feeder section. In addition, a large number of fiber links need to be terminated at the CO, which leads to large CO footprint and increased operational complexity in handling large

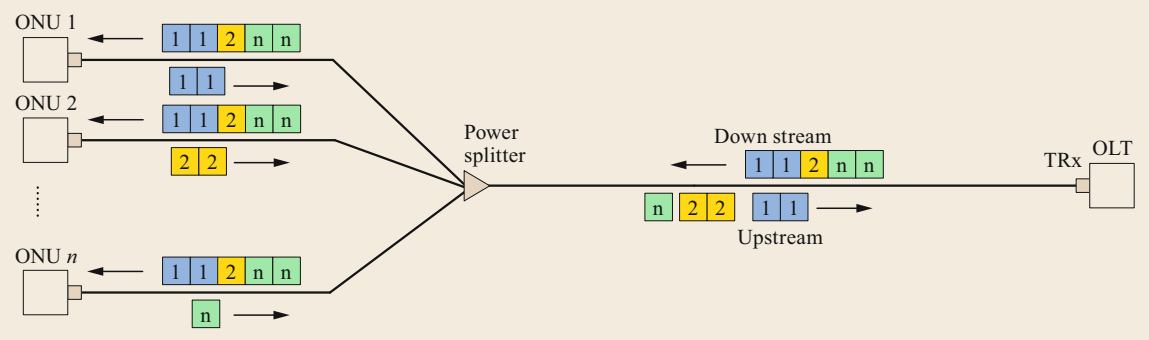


Fig. 26.8 TDM-PON with TDM downstream and TDMA upstream

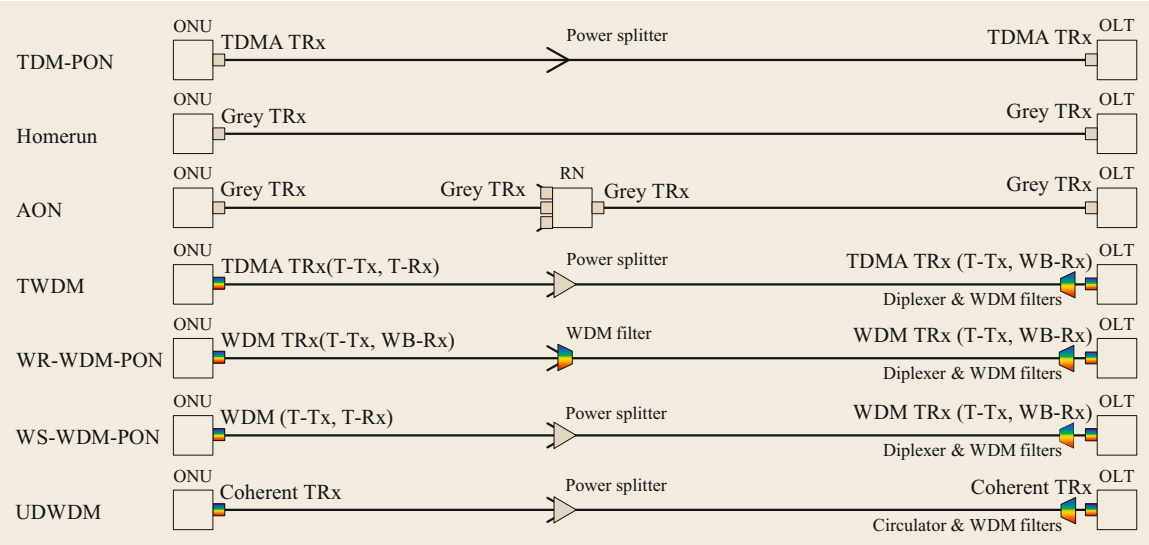


Fig. 26.9 Overview of optical access systems and architectures including TDM-PON, Homerun, AON, TWDM, WR-WDM-PON, WS-WDM-PON, and UDWDM. (Abbreviations: tunable transmitter (T-Tx), tunable receiver (T-Rx), wide band receiver (WB-Rx))

numbers of fiber and transceivers. An alternative approach to Homerun is active optical networks (AONs) with active remote nodes that aggregate traffic at distribution points between the customer and the CO. AONs do not reduce the total number of transceivers but reduce the number of transceivers at the CO and relax the requirements on available feeder fiber. The main drawback with AON is that it requires a large number of active remote nodes that need to be deployed and maintained, typically at high cost. Deployments of Homerun and AON exist but are more frequent in deployments of smaller scale (e.g., municipalities) rather nation-wide rollouts.

Up to access rates of 10 Gbit/s, PON standards in ITU-T and IEEE are based on TDM with increased bitrates for each new PON generation. Higher TDM-PON bitrates provide higher subscriber peak rates, and more bandwidth, which can be used either for increasing average bandwidth per user or for serving more subscribers per PON. However, in the evolution beyond 10 Gbit/s PONs, other options than TDM were considered, which will be discussed in the next section.

26.3.2 Beyond 10 Gbit/s PONs

NG-PON2 [26.30] is the first multiwavelength access standard. It is based on wavelength stacking for increasing PON bandwidth. NG-PON2 specifies time-and-wavelength-division multiplexing (TWDM) with 4–8 TDM-PON wavelengths of 10 Gbit/s each over a single fiber, as a primary alternative. The trade-off between cost and performance is critical in the specification of access PON systems. The cost of the ONU transceiver is particularly critical. Increased PON bitrates require

higher speed transceivers both at the OLT and ONU. In TDM-PON, the ONU transceivers must decode/encode the full PON bitrate, even though the subscriber may be enjoying merely a portion of the full PON bandwidth. Another issue in the evolution beyond 10 Gbit/s is the increased challenges with chromatic dispersion for higher bitrate NRZ-OOK, which results in a distortion of the optical signal with increased propagation distance. Chromatic dispersion limits combined fan-out and reach unless complex transceiver technologies for dispersion compensation are employed, e.g., based on digital signal processors (DSPs).

By employing TWDM, 10 Gbit/s transceiver technologies could be reused in NG-PON2. However, tunable transmitters are needed at the ONU to meet the requirement of colorless ONUs. The ONU is colorless if there is only one type of ONU transceiver despite multiple access wavelengths. The requirement of colorless ONUs is an important operational requirement to avoid the logistics associated with colored ONUs. In a multiwavelength system, this typically means that the ONU transmitter must be tunable in order to tune to an allocated wavelength. Depending on the system, the receiver may also need to be tunable, meaning that a tunable filter at the receiver determines the received wavelength at the ONU. In some systems, filtering is instead done in the ODN, and the ONU receiver only needs to be wide-band capable, i.e., it can receive signals on a wide range of wavelengths. The ability to reuse existing fiber deployment is an important requirement for next-generation systems. Considering the installed base of G-PON and EPON systems, support for legacy power splitter-based ODN is needed, which implies that also the ONU receivers need to be tunable. In NG-PON2 the ODN may be based on different combinations of power splitters and wavelength filters, such as arrayed waveguide gratings (AWGs). A purely power-split ODN has the highest flexibility concerning resource allocation but suffers from large insertion loss. ODNs containing AWGs can achieve longer reach but with less flexibility.

26.3.3 WDM-PON

An alternative track that has been considered for the evolution of fixed-access optical systems is wavelength division multiplexing (WDM)-PON. In WDM-PON, each subscriber is allocated a dedicated pair of wavelengths (one for downstream and one for upstream). The primary technology considered for WDM-PON is dense WDM (DWDM). DWDM is used in metro/-long haul links, but the technology has been considered too costly for the access. PONs based on DWDM with 40–80 wavelength channels at 1–10 Gbit/s each would be able to significantly increase aggregated PON capacity beyond that of TDM-PONs. However, the introduction of WDM in the access raises new challenges in realizing low-cost WDM components, including tunable transceivers. WDM-PONs constitute a large family of solutions, which can be categorized in different ways. One categorization is based on the ODN (Fig. 26.9) and whether it is based on power splitters, i.e., wavelength-selected (WS-)WDM-PON, or based on WDM filters, i.e., wavelength-routed (WR-)WDM-PON [26.31]. Solutions can also be categorized based on the ONU transmitters and whether they are based on tunable lasers or seeded reflective transmitters. In the latter approach, a common light source at the OLT is used, which is filtered by the ODN and a reflected light signal is modulated at the ONU for upstream transmission. The motivation for using seeded reflective transmitters is to avoid the complexity and cost of tunable ONU transceivers. However, this approach typically requires an ODN based on WDM filters, and PON performance in terms of reach is also affected.

WS-WDM-PON, which is based on passive optical power splitters, has a more limited reach compared to WR-WDM-PON due to the high insertion loss of power splitters. Each ONU is assigned one wavelength pair (one downstream and one upstream). Therefore, the maximum number of ONUs is equal to the number of available wavelengths. All wavelengths are available at each of the ONUs. Therefore, tunable receivers (e.g., tunable filters) are needed. In addition, tunable lasers are required for colorless transmitters. In WR-WDM-PON, there is one or several passive devices in the ODN that can multiplex/demultiplex wavelengths. These are typically AWGs which route single wavelengths or wavelength pairs to each ONU. The ONUs can be designed either with tunable lasers or seeded reflective transmitters and do not require tunable receivers.

In order to support latency critical services, NG-PON2 introduced support for dedicated wavelength services (1–10 Gbit/s). These are provided as an overlay (16 PtP WDM channels) co-existing with TWDM in the same ODN (shared spectrum). Part of the optical spectrum is used for PtP wavelengths, while part of the optical spectrum is used for TWDM services. Additionally, NG-PON2 supports a pure WDM-PON configuration, where the full NG-PON2 optical spectrum is used for PtP wavelength services.

However, despite their potential for very high capacity, a number of factors limit applicability of WDM-PONs for fixed-access networks. A general property of WDM-PONs is that there is no statistical multiplexing in the optical system as with TDM-PONs. There

is no sharing of optical transceivers at the OLT between different ONUs. With the evolution of photonic integrated circuits (PIC) and transceiver arrays, it may still be possible to reduce the cost per transceiver at the OLT. However, WDM-PON is primarily a technology for saving feeder fibers. It is efficient for providing high-bandwidth low-latency services but less efficient for bursty best-effort services. Another issue with WDM-PONs, which is more serious for WR-WDM-PON where OLT transceiver resources are tied to particular ONU end-points via the filtered ODN, is how to grow with increasing subscriber take rates. One avenue is to deploy OLT resources in advance, which can simplify provisioning but will result in high upfront costs. The other avenue is to deploy resources when needed, leading to low upfront costs but high operational costs for service provisioning. Furthermore, in WDM-PON, the maximum subscriber fan-out is equal to the number of WDM wavelengths, which depends on the allocated spectrum and grid spacing. Reduced grid spacing leads to higher complexity transceivers in order to mitigate interference and crosstalk. With DWDM, the fan-out is, in practice, limited to 80 channels at 50 GHz spacing exploiting the C/L bands. Higher fan-out requires advanced ultra-dense WDM (UDWDM), which is discussed in Sect. 26.3.4.

26.3.4 Future Outlook and Advanced PONs

Current research efforts are exploring technologies for higher-capacity PONs, such as the use of new modulation formats and denser WDM grids. In IEEE a 100G-EPON task force has been formed for 25, 50, and 100 Gbit/s PONs [26.32].

One avenue for increasing the TDM-PON bitrate beyond 10 to 25–40 Gbit/s is by exploiting low-complexity modulation formats that require less bandwidth and are more dispersion tolerant than NRZ-OOK [26.33]. Several flavors of intensity modulated direct detection (IM-DD) have been considered, such as quaternary pulse amplitude modulation (PAM-4), electrical duobinary (EDB), and optical duobinary (ODB). PAM-4 has two bits per symbol and requires half the baud rate of NRZ and is more dispersion tolerant. However, a more complex receiver is required with increased linearity and three threshold levels. EDB makes use of low-pass filtering. The transmitted data is full-rate NRZ, while at the receiver a converter is needed, which consists of two threshold slicers and an XOR gate to convert data back to NRZ. EDB is more tolerant to dispersion compared to NRZ, while the linearity requirement is relaxed compared to PAM-4. ODB is more dispersion tolerant than EDB but requires a more premium dual-drive Mach–Zehnder modulator (MZM) [26.33]. Dispersion is wavelength dependent, and in the C-band dispersion compensation will in most cases be needed for longer reach. There is a range of methods to mitigate dispersion, from chirp-based lasers to DSPs, with different performance and complexity.

Other flavors of PON that have been proposed and extensively researched include orthogonal frequency division multiplexing (OFDM)-PON, optical code division multiplexing (OCDM)-PON, and UD-WDM-PON. UD-WDM-PON is a variant of WS-WDM-PON exploiting coherent transceivers for ultra-dense channel spacing and long reach with the main drawback being costly coherent technologies. A comprehensive summary of such approaches is presented in Chap. 28.

26.4 Comparison of Architectural Aspects of Different Systems and Technologies

As the access segment is highly cost sensitive, cost is an important parameter for the evolution of fixed-access networks. There have been many techno-economic studies on the cost impact of different technology choices, including operational and capital expenditures (OPEX and CAPEX), as well as total cost ownership (TCO) [26.8, 21].

26.4.1 Optimizing Fixed-Access Architecture

This section summarizes important architectural aspects to be considered in comparing different optical access systems and technologies. This includes technical and operational aspects, as well and business and regulatory aspects. Technical capabilities such as reach, fan-out, and capacity are important for deployment costs. Aspects such as energy efficiency, footprint, active field equipment, resilience, operational complexity, control and management are important for operational costs.

The optimal deployment will depend on how costs are driven in a given scenario. This leads to regional differences depending on geographical characteristics, existing infrastructure, labor costs, etc. At the same time, volumes and uniformity are important drivers for reducing technology costs, limiting the degree to which solutions can be tailored for different markets and regions. Table 26.2 shows a qualitative comparison of the different solutions presented in Sect. 26.3, with re-

Table 26.2 Qualitative comparison of different optical access solutions

Criteria (per service area)	TDM-PON	Homerun	AON	TWDM	WR-WDM-PON	WS-WDM-PON	UDWDM
Support increasing subscriber traffic	●●		●	●●●		●	●
Bit rate per wavelength	●	●●●	●●●	●	●●●	●●	●●
Reduction in fiber count and length	●●●		●●●	●●●	●	●●	●●●
Reduction in number of interfaces	●●●			●●●		●	●
Reduction in active in field equipment	●●	●●●		●●	●●●	●●	●●●
Footprint in CO	●●●		●●●	●●●	●	●	●
Number of wavelengths per fiber				●	●●	●●	●●●
Legacy compatibility with fixed net	●●●		●	●●●	●	●●	●●
Re-use network infrastructure	●●●		●	●●●	●	●●	●●●
Support resiliency	●●		●●●	●●	●	●●	●●
Reduction in energy consumption	●●●	●●		●●●	●●	●	●
Reduction in failure rate	●●●	●●		●●●	●●	●●	●
Simple to service provision	●●●	●	●	●●●	●	●●●	●●●
Reduction in active shelves in CO	●●●		●●●	●●●	●		
Low latency (system level)	●	●●●	●●	●	●●●	●●●	●●●

Best – ●●● . . . ●● . . . ● – Worst

spect to various criteria, which will be discussed in Sects. 26.4.2–26.4.8.

26.4.2 Support for Traffic Evolution

Subscriber traffic volumes and take rates grow incrementally. Support for traffic growth differs between different systems. TDM-PON systems inherently provide flexibility to share bandwidth across a larger number of end-points. As traffic or subscriber take rates increase beyond certain thresholds, the ODN may be reconfigured with a lower split ratio. Operators may prepare the ODN for higher data rates by deploying necessary feeder fibers and relevant splitter stages. Figure 26.10 shows an example TDM-PON deployment in a 1 : 64 configuration, where the ODN is prepared with additional feeder fibers for a future upgrade to a 1 : 32 configuration as traffic and take rates increase. The additional splitter stage at the CO may reduce upfront investments in OLT equipment if initial subscriber rates are expected to be low.

For WDM and PtP solutions an abundance of bandwidth is typically provided already from start. The optical transceiver resources are dedicated per subscriber, and any subscriber upgrade in capacity requires an upgrade of the transceivers.

26.4.3 System Fan-Out and Reach

System fan-out and reach affect parameters such as fiber count and length, the number of interfaces, infield active equipment, and footprint at the CO, which in turn drive costs. However, the relationship also depends on the technology, where, for example, WDM fan-out (as opposed TDM) does not necessarily reduce the number of optical interfaces nor CO footprint.

High system fan-out increases sharing of feeder fibers, which is important in fiber-scarce scenarios. High system fan-out is typically more important in urban and dense urban scenarios, where the subscriber density is high and distances short. TDM and TWDM systems support high fan-out, e.g., NG-PON2 defines a maximum fan-out of 1 : 256 for TWDM. Studies show a significant TCO reduction for a fan-out up to around 1000 users, whereas a further increase limits significant cost savings and leads to a higher failure penetration range (i.e., more subscribers may be affected by a single failure). In terms of fan-out, WDM-PON is constricted by the wavelength spectrum and the wavelength grid. Technologies such as UD-WDM-PON that enable high WDM fan-out per feeder fiber suffer from high cost due to complex transceiver technology.

As described in Sect. 26.2.4, insufficient system reach may impact the number of required access nodes. System reach is crucial in rural deployments, where service areas may be large. Systems should typically provide passive long-reach transmission support of 20–40 km [26.8]. In power splitter-based systems the trade-off between fan-out and reach can be exploited to provide long reach at the expense of fan-out, where the available power budget is used for reach rather than fan-out. Standard TDM-PON systems support 40 km only with long-reach interfaces and a highly reduced power split ratio, i.e., lower than 1 : 8. In general, support for 60–90 km extended reach [26.8] for a backup path to the next closest CO (in the case of failure) can be achieved by using an active reach extender. WDM-PON systems can provide longer reach. WR-WDM

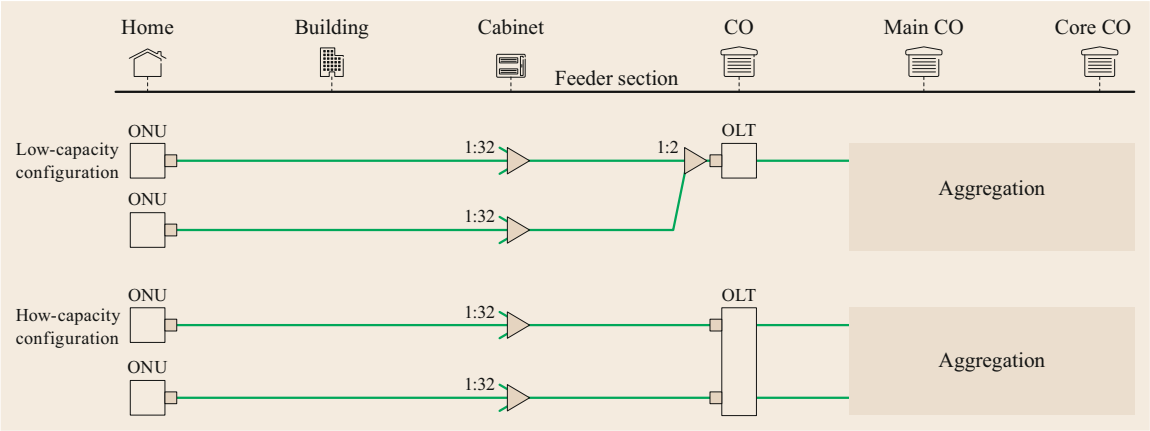


Fig. 26.10 TDM-PON where the ODN is prepared for future traffic growth to facilitate upgrade from 1 : 64 split to 1 : 32 split

PON may reach 60 km passively, with 80 channels using preamplifiers and avalanche photodiode (APD) ONUs [26.8].

26.4.4 Migration and Co-existence

Migration refers to shifts between technology generations, while co-existence refers to multiple technology generations co-existing simultaneously. Both migration and co-existence depend on the existing deployment and if costly ODN infrastructure deployment or upgrades are required. For example, migrating from a power splitter-based ODN to a filter-based ODN requires replacement of in-field filters. The assessment in Table 26.2 assumes that the existing infrastructure is a legacy GPON/EPON ODN. PON standards specified in ITU-T and IEEE are defined to support ODN co-existence between generations. By installing co-existence filters at the OLT side, two PON generations may co-exist in the same ODN, and customers can be migrated gradually, replacing *old* ONUs with *new* ONUs.

26.4.5 Resilience

Requirements on high availability and low failure penetration range impose the necessity for redundancy and resiliency mechanisms. To achieve an availability of $\geq 99.98\%$ (one-way), fiber path and OLT protection is required in all architectures in addition to the resiliency mechanisms in the aggregation network. Some business customers and some mobile cell sites may require an availability of $\geq 99.99\%$, which may lead to additional requirements on protection in the ODN infrastructure. The complexity of implementing fiber path protection differs between different solutions, as reflected in the assessment in Table 26.2.

26.4.6 Operational Aspects

Operational aspects include aspects such as service provisioning, maintenance, energy efficiency, etc. A fully automatic service provisioning, referred to as *zero touch* provisioning, is only possible where a full deployment and installation is done in the beginning, which leads to high up-front cost covering ODN infrastructure and in-house network installation. On the OLT side, there is a difference between solutions based on filtered ODNs and solutions based on power-split ODNs. In the case of filtered ODNs, OLT transceivers are either pre-deployed for each client enabling *zero touch* provisioning but at high upfront costs, or they are installed gradually at the expense of higher provisioning costs (e.g., for manual transceiver installation and fiber patching). In the case of power-split ODNs, shared OLT transceiver resources may be pre-deployed enabling *zero touch* provisioning for a certain number of clients but do not need to be pre-deployed for a particular client. Homerun and AON architectures based on PtP fiber links exhibit similar service provisioning properties to WDM-PONs based on filtered ODNs.

Both failure rates (which is a driver for maintenance costs) and energy consumption typically scale with active equipment and equipment complexity. This generally favors TDM-PONs compared to architectures that rely on PtP and WDM, which require a larger number of optical interfaces.

26.4.7 Mobile Transport

Mobile transport requirements were discussed in Sect. 26.2.5. Different optical access architectures offer different mobile transport and convergence possibilities. Tight RAN coordination, due to latency and jitter

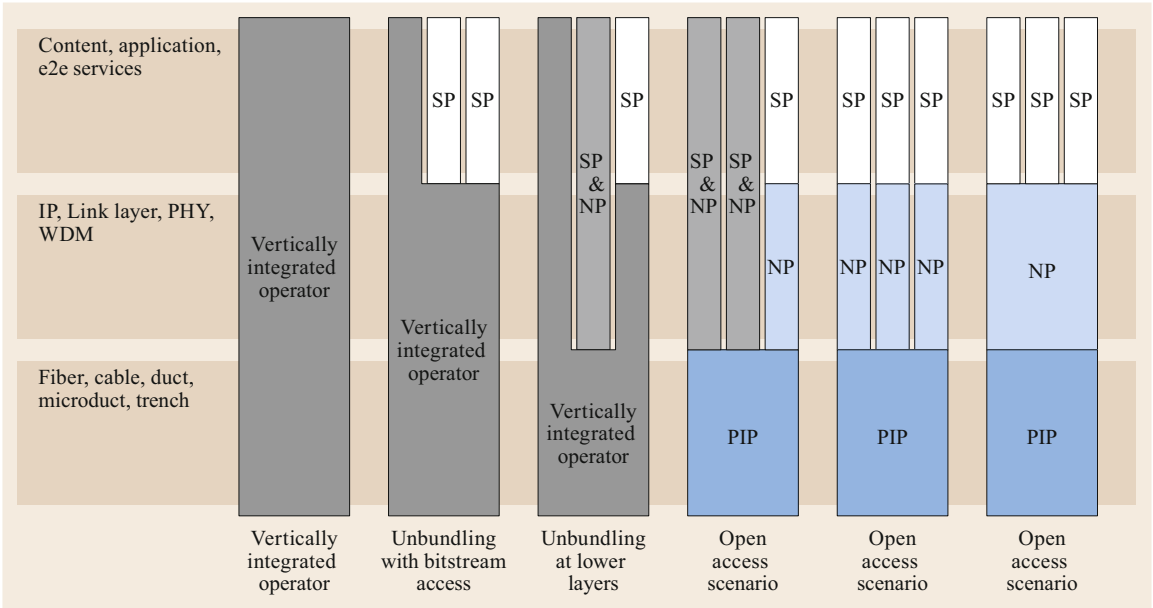


Fig. 26.11 Different business scenarios for fixed access with different roles and actors

requirements, is favorably supported by PtP or PtP WDM transport. TDMA-based solutions do not fulfill the requirements. Wavelengths may be carried over a dedicated ODN (e.g., WR-WDM PON) or a shared ODN (e.g., NG-PON2 with WDM overlay). Similar converged deployments with PtP WDM overlay may also be based on other flavors of TDM-PON (GPON and XG-PON). As described in Sect. 26.2.6, tight RAN coordination and associated latency requirements constrain the centralization of RAN resources to a distance of < 40 km. CAPEX studies [26.9] show that for aggregation of small cells via the fixed-access infrastructure, the convergence benefits of NG-PON2 increase with increasing small-cell density, increasing FTTH coverage, and reduced fiber availability. Dedicated solutions for mobile transport (e.g., WR-WDM PON) are favorable for low small-cell densities, low FTTH coverage, or high fiber availability. For higher capacity RAN transport, e.g., required by specific RAN functional splits, the benefits of convergence decrease, favoring dedicated RAN transport.

26.4.8 Cooperation Models

Besides the technical and operational aspects there are also important business and regulatory aspects to the deployment of fixed-access networks. Figure 26.11 shows a range of different cooperation scenarios for broadband access, with different roles that have been identified such as the physical infrastructure provider (PIP), the network provider (NP), and the service provider (SP). The PIP role covers responsibility for physical infrastructure such as fiber, cable, ducts, and trenches. The NP role covers responsibility for network equipment, including optical access systems, routers, etc., while the SP role covers responsibility for services that are delivered over the network infrastructure. As shown in Fig. 26.11, an actor can take on one or several of these roles.

Different cooperation models drive different requirements on fixed-access systems and architectures. Regulations on competition may require unbundling, i.e., vertically integrated operators are required to open the network at specific layers for competition. Two examples of unbundling are shown in Fig. 26.11, where the network of a vertically integrated operator is opened for competitors on the SP or NP levels. Different access systems and architectures enable different unbundling scenarios such as fiber unbundling, wavelength unbundling, and bit-stream access (BSA). Fiber unbundling means opening the fiber infrastructure for different NPs such that each NP uses dedicated fibers. This form of unbundling is good in terms of isolation between NPs but does not provide any sharing of fibers. Infrastructure sharing is limited to sites and ducts. Wavelength unbundling enables additional sharing of fiber infrastructure between NPs operating in different wavelength domains. This provides good isolation between NPs. However, this is only possible for architectures based on WDM, yet adds significant complexity if implemented at large scale. Another alternative is to open the network at a bitstream level. BSA

is possible with any access technology and typically the most cost effective, as it enables the highest degree of infrastructure sharing. Yet, it raises concerns regarding isolation between different providers. In all open access scenarios some form of coordination between NPs operating in the same area is necessary [26.8]. In open access scenarios, also shown in Fig. 26.11, multiple NPs can have access to all customers who can freely select the NP that gives the best offer. In such scenarios, the dynamicity of market share should be properly supported by the technology. There may be requirements on inter-NP isolation. If multiple NPs share the same infrastructure, there should be a mechanism to detect a misbehaving ONU (deliberate or not) and to react by switching it off. Furthermore, in the case of failure, the affected network equipment should be isolated.

26.5 Summary

Fixed-access networks have played an important role in society during the past decades. With growing capacity demand, fiber is penetrating deeper in the access, and optical access technologies are now driving the evolution of fixed-access networks. In Section D of this book, all important technologies and trends that characterize current and envisioned optical access are described. Chapter 27 reviews current TDM-PON technologies and standards that have been specified in ITU-T and IEEE, considering that wavelength stacking was used in the most recent PON standard (NG-PON2) for increasing PON capacity to 40 Gbit/s. Further advancements in PON technologies are described in Chaps. 28 and 29. Since from an architectural perspective, node consolidation has been seen as one avenue for cost reduction in access networks, Chap. 30 presents long reach PON technology, which is a main enabler for node consolidation. As optical access networks are expected to play an increasingly important role for mobile transport with rapidly growing traffic in mobile networks, Chaps. 31, 32 and 33 review different technologies through which optical technologies can support wireless access networks. Chaps. 34 and 35 present other emerging optical technologies that rely on free-space propagation. Chapter 34 discusses free-space optics, which has important applications in satellite communications. Chapter 35 presents the area of visible light communication (VLC) where visible light is exploited for high-speed short-range wireless connectivity. Finally, Chap. 36 overviews applications of optical communications and sensing for avionics.

References

26.1 Fortune Global 500, http://fortune.com/global500/ (2017)

26.2 World Bank Information and Communication for Development: Extending reach and increasing impact, http://www.worldbank.org/en/topic/ict (2017)

26.3 European Commission: Digital Agenda Scoreboard, https://ec.europa.eu/digital-single-market/en/digital-scoreboard (2013)

26.4 OECD: Broadband Portal, http://www.oecd.org/sti/broadband/broadband-statistics/ (2017)

26.5 AT&T: 2014 annual report. https://www.att.com/Investor/ATT_Annual/2014/downloads/att_ar2014_annualreport.pdf (2014)

26.6 J.F. Monserrat, H. Droste, Ö. Bulakci, J. Eichinger, O. Queseth, M. Stamatelatos, H. Tullberg, V. Venkatkumar, G. Zimmermann, U. Dötsch, A. Osseiran: Rethinking the mobile and wireless network architecture: the METIS research into 5G. In: *Proc. Eur. Conf. Netw. Commun. (EuCNC)* (2014)

26.7 E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, Y. Selén: 5G radio access. In: *Ericsson Technology Review* (Ericsson, Stockholm 2014)

26.8 M. Forzati, A. Bianchi, J. Chen, K. Grobe, B. Lannoo, C. Mas Machuca, J. Point, B. Skubic, S. Verbrugge, E. Weis, L. Wosinska, D. Breuer: Next generation optical access seamless evolution, J. Opt. Commun. Netw. **7**(2), 109–123 (2015)

26.9 COMBO: COMBO Deliverable D3.6, final architectural recommendations for FMC networks, http://www.ict-combo.eu/data/uploads/deliverables/combo_d3.6_wp3_30september2016_v1.0.pdf (2016)

26.10 European Telecommunication Standard Inst., Valbonne: Multi-access edge computing (MEC), http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing (2019)

26.11 Open Fog Consortium, Fremont: OpenFog, https://www.openfogconsortium.org/ (2019)

26.12 Ofcom: UK fixed-line broadband performance, https://www.ofcom.org.uk/__data/assets/pdf_file/0026/59273/bb-speeds-may2011.pdf (2011)

26.13 P.-E. Eriksson, B. Odenhammar: VDSL2: Next important broadband technology, Ericsson Rev. No. 1, https://www.ericsson.com/ericsson/corpinfo/publications/review/2006_01/files/vdsl2.pdf (2006)

26.14 Nokia: https://networks.nokia.com/solutions/g.fast

26.15 J. Libbenga: Europe breaks one million barrier FTTH Council Europe, https://www.ftthcouncil.eu/

resources/resources/europe-hits-one-million-ftth-broadband-connections?media_id=1211 (2008)
26.16 P.W. Shumate: Fiber-to-the-Home: 1977–2007, J. Lightwave Technol. **26**, 1093–1103 (2008)
26.17 G. Kramer, G. Pesavento: Ethernet passive optical network (EPON): Building a next-generation optical access network, IEEE Commun. Mag. **40**, 66–73 (2002)
26.18 B. Lund: PON architecture 'futureproofs' FTTH, Lightwave **16**, 104–107 (1999)
26.19 D. Breuer, F. Geilhardt, R. Hulsermann, M. Kind, C. Lange, T. Monath, E. Weis: Opportunities for next-generation optical access, IEEE Commun. Mag. **49**(2), S12–S14 (2011)
26.20 Cisco White Paper: The Zettabyte Era: Trends and Analysis (2017)
26.21 COMBO: COMBO deliverable D3.3: analysis of transport network architectures for structural convergence, http://www.ict-combo.eu/data/uploads/deliverables/combo_d3.3_pu.pdf (2015)
26.22 Next Generation Mobile Networks (NGMN): 5G white paper, https://www.ngmn.org/5g-white-paper/5g-white-paper.html (2015)
26.23 The 5G infrastructure Public Private Partnership (5G-PPP): 5G-PPP 5G Architecture, https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-5G-Architecture-WP-July-2016.pdf (2016)
26.24 P. Iannone, K.C. Reichmann, C. Ranaweera, M. Resende: A small cell augmentation to a wireless network leveraging fiber-to-the-node access infrastructure for backhaul and power. In: *Opt. Fiber Commun. Conf* (2013) p. Paper OTu2E.2
26.25 K. Laraqui, S. Tombaz, A. Furuskär, B. Skubic, A. Nazari, E. Trojer: Fixed wireless access on a massive scale with 5G, Ericsson Rev. **10**, 2 (2016)
26.26 ITU-T Rec. G.984.x, Gigabit-capable passive optical networks (GPON), https://www.itu.int/rec/T-REC-G.984.1/en (2008)
26.27 ITU-T Rec. G.987.x, 10-Gigabit-capable passive optical networks (XG-PON), https://www.itu.int/rec/T-REC-G.987 (2012)
26.28 IEEE 802.3ah GE-PON, https://standards.ieee.org/standard/802_3ah-2004.html (2004)
26.29 IEEE 802.3av 10G-EPON, https://standards.ieee.org/standard/802_3av-2009.html (2009)
26.30 ITU-T Rec. G.989.x, 40-Gigabit-capable passive optical networks (NG-PON2), https://www.itu.int/rec/T-REC-G.989 (2015)
26.31 K. Grobe, M. Roppelt, A. Autenrieth, J.-P. Elbers, M. Eiselt: Cost and energy consumption analysis of advanced WDM-PONs, IEEE Commun. Mag. **49**(2), S25–S32 (2011)
26.32 IEEE P802.3ca 100G-EPON Task Force, http://www.ieee802.org/3/ca/ (2017)
26.33 D. van Veen, V. Houtsma: High speed TDM PON beyond 10G. In: *Opt. Fiber Commun. Conf* (2016) p. Paper Tu3C.3

Björn Skubic

Ericsson Research
Ericsson AB
Stockholm, Sweden
bjorn.skubic@ericsson.com

Björn Skubic received his PhD in Physics from Uppsala University, Sweden, in 2007. He joined Ericsson Research in 2008 and has worked in areas such as fixed access networks, energy efficiency, and mobile transport. His current research interests include access networks and edge computing.

Lena Wosinska

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden
wosinska@chalmers.se

Lena Wosinska received the PhD degree in Photonics and the Docent degree in Optical Networks from the KTH Royal Institute of Technology, Stockholm, Sweden, where she has been a Full Professor of Telecommunication until October 2018. Currently, she is a Research Professor at the Chalmers University of Technology, Gothenburg, Sweden. She is the founder of the Optical Networks Lab (ONLab), which moved with her from KTH to Chalmers. Her research interests include fiber access and 5G transport networks, optical network control and reliability, and optical datacenter networks.

27. Current TDM-PON Technologies

Jun-ichi Kani, Doutje van Veen

This chapter describes many aspects of time-division multiplexing-passive optical network (TDM-PON) technology. TDM-PON is the architecture for optical access systems preferred by many network operators, due to its low operational and capital expenses. After the overview, topics related to the physical layer, such as burst-mode transmission and analog video distribution, are addressed. Then, topics in the higher layers of the network, such as the access control technologies, dynamic bandwidth allocation, security and privacy issues, protection switching and methods to improve energy efficiency like sleep modes, are described in detail for both ITU-T- and IEEE-based PONs. Finally, we conclude with a section on technologies beyond 10G PON.

The time-division multiplexing-passive optical network (TDM-PON) is an optical access system in which an optical line terminal (OLT) communicates with a number of optical network units (ONUs) through an optical distribution network (ODN) comprising optical fibers and optical splitter(s), as illustrated in Fig. 27.1. TDM-PON has been widely deployed worldwide because of its cost-effectiveness; it shares not only the optical fiber but also the OLT optics/electronics among a number of ONUs connected to the OLT.

27.1 Passive Optical Networks

For upstream transmission, TDM-PON relies on a scheme called time-division multiple access (TDMA) as shown in Fig. 27.1a. When the optical splitter combines upstream optical signals from all the ONUs, TDMA is employed to avoid collisions at the optical splitter. In the TDMA scheme, under the control of OLT, ONUs send their upstream signals as bursts using time slots that are different from each other. For downstream transmission, a time-division multiplexing (TDM) scheme is used as illustrated in Fig. 27.1b. Data and other information being sent to ONUs are multiplexed using a specified frame in the time domain. While the multiplexed downstream signal is optically broadcast to all the ONUs through the splitter(s), each ONU extracts the data and information directed to it after the reception of the downstream optical signal, i.e., after the optoelectronic conversion.

Although a single-stage optical splitter is illustrated in Fig. 27.1, optical splitting can be multistage. For example, it is possible to put the first-stage optical splitter inside the central office and the second-stage optical splitter at a remote node closer to the subscribers. A typ-

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_27

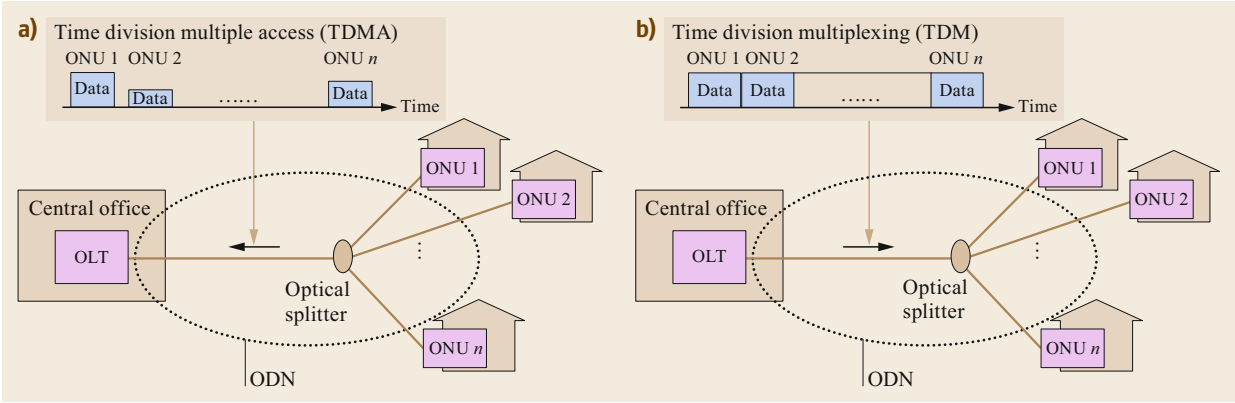


Fig. 27.1a,b TDM-PON; **(a)** TDMA for upstream and **(b)** TDM for downstream

ical total splitting number, i.e., the number of ONUs that can be physically connected to an OLT, ranges from 16 to 64, but can be even higher for PONs with short reaches. The point-to-multipoint optical connection composed of optical fibers and optical splitter(s) is called an optical distribution network (ODN).

TDM-PON was first explored in the late 1980s [27.1], and several TDM-PON options were studied in the 1990s with the aim of realizing fiber to the home (FTTH) [27.2]. In 1995, the full-service access network (FSAN) group was formed by several telecom carriers [27.3]. FSAN promoted the development of a common specification for TDM-PON, and the work resulted in the development of broadband PON (B-PON) in ITU-T over the period from 1998 to 2001 [27.4, 5]. B-PON offers 155.52 Mb/s bitrate upstream and 155.52 and 622.08 Mb/s bitrate downstream. It is based on asynchronous transfer mode (ATM) PON, and was initially called ATM-PON. While B-PON originally specified a frame structure to carry various types of user data, including Ethernet, in the early 2000s the concept of directly exchanging Ethernet frames in PON appeared [27.6]. This led to the development of the Ethernet PON (E-PON); E-PON offers 1.25 Gb/s bitrate both upstream and downstream as a part of the IEEE 802.3 (Ethernet) standard [27.7]. At the same time, FSAN/ITU-T developed Gb-capable PON (G-PON) based on a new generic frame to carry various protocols [27.8]; G-PON has 1.24416 and 2.48832 Gb/s bitrates for upstream and downstream. In the late 2000s, both FSAN/ITU-T and IEEE developed 10 Gb-class PON systems; those are 10 Gb-capable PON (XG-PON) [27.9] and 10 Gb Ethernet PON (10G E-PON) [27.10]. While XG-PON has 2.48832 Gb/s upstream and 9.95328 Gb/s downstream bitrates, a PON system having 9.95328 Gb/s bitrate for both upstream and downstream was additionally standardized as XGS-PON (10 Gb-capable symmetric PON) later in ITU-T [27.11]. The bitrate of 10G E-PON is 10.3125 Gb/s both upstream and downstream. All the PON systems mentioned here are based on TDMA (upstream) and TDM (downstream). Figure 27.2 summarizes the history of TDM-PON standardization, with rounded bitrate values.

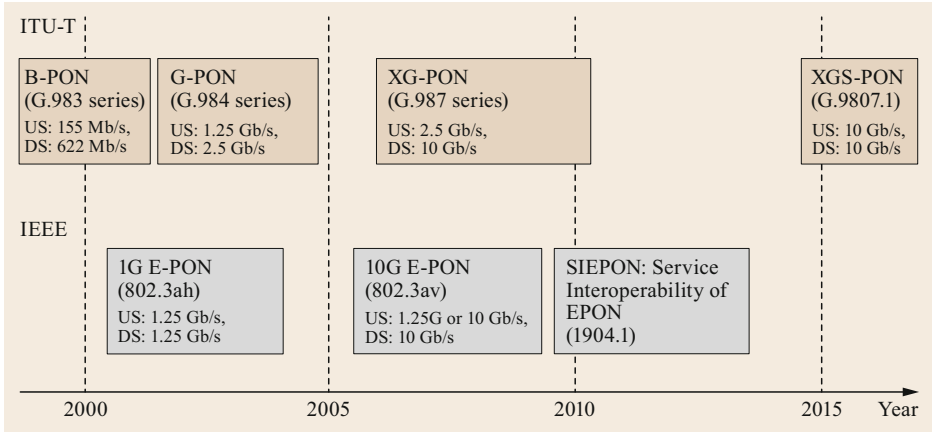


Fig. 27.2 History of TDM-PON standardization. (PON: passive optical network, B-PON: broadband PON, G-PON: gigabit-capable PON, XG-PON: 10 Gb-capable PON, NG-PON: next-generation PON, GE-PON: gigabit Ethernet PON)

27.2 Physical Layer Technologies

TDM-PONs such as G-PON and E-PON are based on a passive fiber distribution network, making them point-to-multipoint connections. In the downstream, data are distributed to the users based on time-division multiplexing, and in the upstream each ONU obtains access to the fiber via time-division multiple access schemes.

Downstream passive power splitting is very compatible with broadcasting video schemes; all ITU-T and IEEE PON standards reserved an extra downstream wavelength transporting radio-frequency (RF) video broadcasting from the OLT to all the ONUs.

To enable the upstream TDMA scheme, a burst-mode transmitter at the ONU and a burst-mode receiver at the OLT are needed. In the paragraphs below, details on the physical-layer technologies supporting upstream TDMA, RF video overlay and the outside fiber plant are discussed.

27.2.1 Optical Distribution Network

The ODN consists of a feeder fiber, passive optical power splitter(s) and fiber drops to each ONU.

The ODN is owned by the network operator, so in order to enable interoperability between different PON component vendors, the ODN needs to comply with a standard as well. The main parameters of the ODN are optical power loss (dB) because of splitter and fiber attenuation, and fiber distance (km). Maximum and minimum ODN loss are defined for different optical power budget classes, which correspond to attenuation ranges that need to be supported as defined in the available IEEE and ITU-T PON standards. The earlier ITU-T standards allow implementation of bidirectional transmission using wavelength-division multiplexing (WDM) on a single fiber or unidirectional transmission over two fibers. The majority of available PON standards set the upstream and downstream wavelengths with a large separation, which enables low-loss and low-cost 45° angle filters to multiplex and demultiplex the downstream and upstream wavelengths onto a single fiber. Therefore, the loss budget must be divided between fiber, splitter and WDMs (in the single-fiber case). The reach parameter of the ODN determines the maximum fiber length. The ODN needs to use standard single-mode fiber (SSMF) as described in ITU-T G.652.

Figure 27.3 shows an ODN with two stages of power splitting for a subset of ONUs, causing, e.g., ONU 1 and ONU 2 to experience a larger loss than ONU n. Maximum differential loss between ONUs is determined by the differential optical path loss parameter in the standard. The variation in output power from the ONU transmitters together with the differential loss define the maximum optical received power/minimum optical received power ratio, i.e. loud/soft ratio (LSR), as received by the OLT receiver. The OLT receiver needs to be able to receive consecutive loud and soft bursts correctly. This requirement is very important for the OLT burst-mode receiver (BM Rx), as described below in greater detail.

The physical-medium-dependent layer parameters of ODN are defined for B-PON and G-PON as listed in Table 27.1. Three original power budget classes are defined for B-PON/G-PON, namely class A, class B and class C. The power budget includes fiber loss, splitter loss and WDM losses to combine and separate up- and downstream wavelengths. Class B+ and class C+ were later added to the standard after realizing that three additional decibels were needed to accommodate WDM losses [27.12, 13]. In Table 27.2, equivalent parameters are listed for IEEE 1G E-PON, taken from [27.7]. In general, the IEEE PON standards have more relaxed parameters than the ITU-T PON standards.

In 2008, reach extension of G-PON based on regeneration or optical amplification was introduced in [27.14].

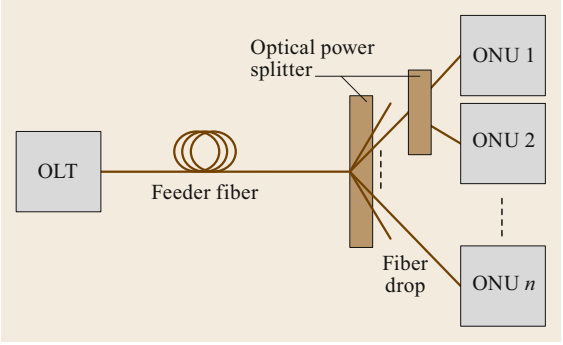


Fig. 27.3 Optical distribution network with two stages of power splitting

Table 27.1 Physical parameters of ODN for ITU-T B-PON/G-PON

Item	Unit	Specification
Fiber type	–	ITU-T G.652
Attenuation range	dB	Class A: 5–20 Class B: 10–25 Class B+: 13–28 Class C: 15–30 Class C+: 17–32
Differential optical path loss	dB	15
Maximum fiber distance	km	20 (10 as option)
Minimum optical return loss	dB	32

Table 27.2 Physical parameters of ODN for IEEE 1G E-PON

Item	Unit	Specification
Fiber type	–	ITU-T G.652
Attenuation range	dB	PX10-U: 5–20 (upstream) PX10-D: 5–19.5 (downstream) PX20-U: 10–24 (upstream) PX20-D: 10–23.5 (downstream)
Differential optical path loss	dB	13.5–15
Nominal fiber distance	km	10/20 (PX10/PX20)
Minimum optical return loss	dB	20

In Fig. 27.4, a basic PON architecture and an extended PON architecture with an optical trunk line (OTL) are shown. The extended maximum reach is up to 60 instead of 20 km, and an optical loss budget of more than 27.5 dB is supported in both the ODN and OTL. A logical reach of 60 km, which determines the maximum transmission delay that is still enabled by the PON protocol layer, was already defined in the original G-PON standard, so no changes were required for the higher layers of the network to accommodate the extension.

27.2.2 Burst-Mode Technology

The downstream TDM transmission scheme results in continuous data with constant amplitude and synchronous ONU receivers. The OLT transmitter sends a single data stream, which is received by all the ONU receivers.

In the upstream TDMA transmission scheme, the data stream is transmitted in burst mode. Every ONU sends bursts of data and the OLT receiver receives the bursts from all the ONUs. Each burst will therefore have a different optical received power and a different phase due to the different transmission distances between the OLT and each ONU. Adapting to the different optical received powers and aligning to the phase of each burst makes the burst-mode receiver quite a challenging component.

The maximum differential optical path loss and transmitter output power variations due to, e.g., manufacturing or temperature, result in an LSR equal to 20 dB. This is a very stringent parameter for the OLT because it needs to adapt its gain settings and settle very quickly at the beginning of each burst to level the up to 20 dB optical power differences between two consecutive bursts.

Hence, OLT needs to amplify bursts with varying power and it needs to recover the phase of each burst. One of the most complex components in a PON is therefore the OLT burst-mode receiver. Furthermore, the necessary settling time and phase-locking time increase the overhead, and thus lower the throughput of the upstream transmission.

In summary, for downstream transmission, the OLT Tx and ONU Rx can be implemented using conventional transmitter and receiver technology, as also used for point-to-point links. For the upstream direction, the ONU needs a burst-mode transmitter and the OLT needs a burst-mode receiver.

A burst-mode transmitter is based on a regular transmitter except that the data are sent in bursts. The laser is turned on only when a user is sending data. The laser needs to be turned on and off very rapidly to minimize extra overhead caused by the death time between consecutive bursts. Specialized burst-mode laser drivers have been developed to perform this task, as the bias current circuit in conventional burst-mode laser drivers is usually too slow.

Figure 27.5 shows the different parts of an upstream burst. Each upstream burst starts with a preamble to accommodate, e.g., settling effects and provide enough signal transitions to enable the burst-mode receiver to phase-lock to each burst. Ideally, this preamble is as short as possible to keep the extra overhead small in the upstream. Also, a guard band is needed to accommodate, for example, inaccuracy of the delay measurement during ranging and thus avoid burst overlap; the guard band between two bursts causes additional overhead.

The guard band can overlap with laser on and off times to minimize overhead. Finally, each upstream burst contains a delimiter that enables synchronization to the data in the payload section of the burst.

In general, the IEEE-standards-based E-PON has more relaxed burst-mode parameters because it focuses on cost, while the ITU-T-based PONs have more

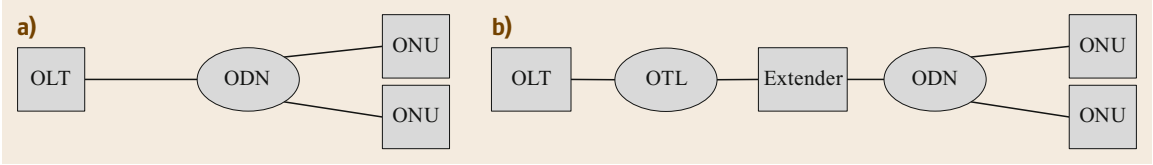


Fig. 27.4a,b A basic PON (a) and an extended PON (b)

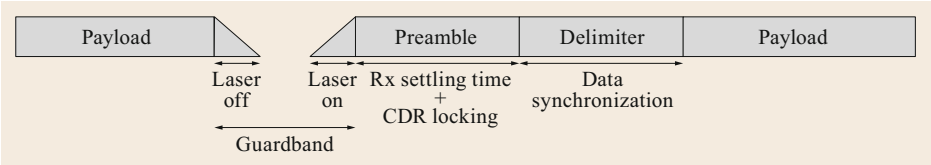


Fig. 27.5 Upstream burst-mode frame

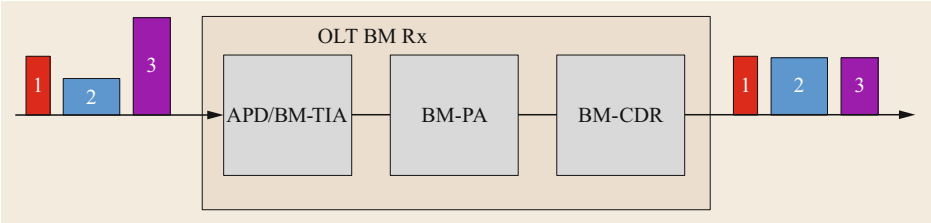


Fig. 27.6 Schematic overview OLT burst-mode receiver

stringent burst-mode parameters focusing on network throughput.

Burst-mode receivers can be DC-coupled or AC-coupled. DC-coupled receivers are preferred with regard to faster settling. AC-coupled receivers are easier to design but have a longer settling time due to the RC-time constant of the AC-coupling resistor and the circuit capacity in such a receiver. Figure 27.6 shows a schematic overview of a burst-mode receiver. It usually consists of an avalanche photo diode (APD) with a burst-mode transimpedance amplifier (BM-TIA). The BM-TIA levels and amplifies the input burst-mode diode current. It can be AC-coupled to filter the burst-mode DC component from the diode current or DC-coupled, with a fast gain control. To enable the fast gain control, the decision threshold and DC offset must be determined and measured in a short time. Inaccuracies in threshold and DC offset result in a sensitivity penalty relative to AC-coupled receivers. A reset signal provided from higher-protocol layers helps to improve the settling time of DC-coupled BM receivers. The next stage of the BM receiver is a burst-mode post-amplifier (BM-PA), which further levels the signal bursts and amplifies the signal such that the final part, the burst-mode clock and data recovery (BM-CDR), has a strong enough input signal. The preamble section of the burst is used to measure signal strength for the gain control in a DC-coupled BM receiver; it also provides transitions to enable fast locking of each burst by the BM-CDR, and these devices usually require an external reset signal.

Three types of implementation of the BM-CDR can be distinguished. The first is a fast-lock PLL (phase-locked loop)-based CDR (clock and data recovery) with lock times up to 100 ns, but with good output jitter characteristics. The second main type is an oversampling CDR, which is very fast; it needs only a few nanoseconds to lock but is relatively power-hungry due to the high-speed sampler needed. The third main type is a gated-voltage-controlled oscillator (VCO) type CDR; this is the fastest type, and it can lock in 1 bit, but it has a relatively large output jitter. Reference [27.15] gives a good overview of burst-mode technology. Large output jitter from the CDR causes errors in data recovery, so in general there is a trade-off between lock time and data recovery performance for BM-CDRs.

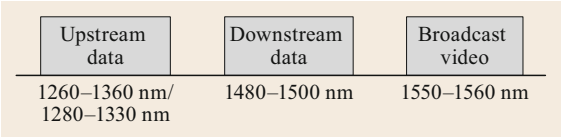


Fig. 27.7 Wavelength map in B-PON and G-PON with bidirectional data and downstream broadcast video

27.2.3 Video Overlay

ITU-T G.983.3 [27.12] defines an enhancement wavelength band, which can be used for new additional service capabilities, for example, broadcast video services and dense wavelength-division multiplexing (DWDM) digital data services.

Broadcast video has been the main service provided over the enhancement band. See Fig. 27.7 for the wavelength map in B-PON and G-PON.

When the standard was written, packet-based video technology was not yet mature, but the market for video services was large and growing. At that time, video distribution using the enhancement band was a nice compromise that also helped FTTH networks expand their market.

The video overlay wavelength carries an analog radio-frequency (RF) signal that contains television channels using 50–750 MHz electrical bandwidth. As analog channels need a high signal-to-noise ratio (SNR), the RF video overlay was placed in the low-

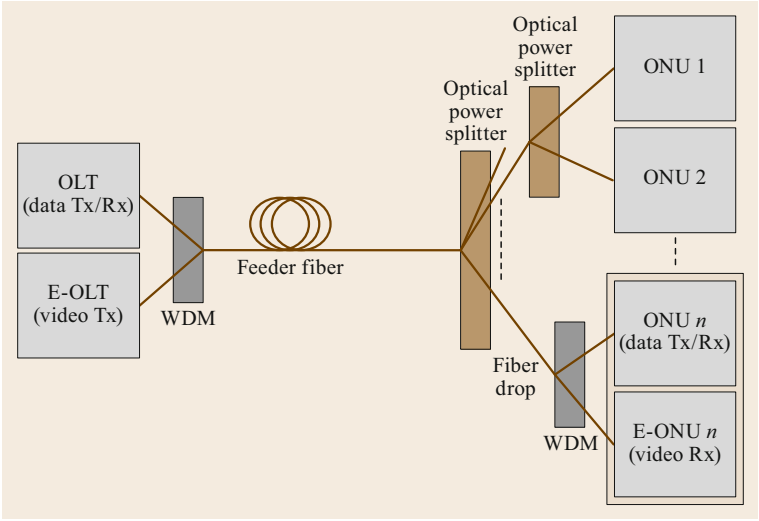


Fig. 27.8 PON architecture with a video overlay wavelength

loss-wavelength window around 1550 nm of standard single-mode fiber.

Figure 27.8 shows a PON architecture with a video overlay wavelength. At the ONU, the video signal is split off using a wavelength filter and received with a linear analog optical receiver, which converts it to an electrical RF signal that can be transported via coax-cable to the televisions/set-top boxes in the home. Upstream signals from the set-top box for interactive services use the basic upstream band of the PON.

The high SNR requirement of the RF video also calls for a lower analog receiver sensitivity and dynamic range compared with basic band digital receivers. A reduced differential ODN loss can be achieved, for example, by inserting the downstream RF video at the second stage of the splitter architecture.

Another negative side effect of the video overlay is that the shorter digital downstream wavelength (1490 nm) acts as a Raman pump for the downstream analog video signal at 1550 nm, resulting in an optical power penalty for the digital signal due to depletion. The analog video signal sees a reduction in the carrier-to-crosstalk ratio (CCR) due to the Raman interaction. The Raman effect is stronger at the lower frequencies (it is a low-pass effect), so the video channels with the lowest frequencies see the greatest effect; see also [27.16] for a more detailed overview of Raman pumping. Fortunately, the Raman pumping increases with fiber distance, while the optical power is attenuated along the fiber, so the overall carrier-to-noise ratio (CNR) of the video signal as a function of fiber distance is relatively constant [27.17]. One way to mitigate the Raman penalty for the analog video is to increase the modulation index of the lower-frequency video channels by a few decibels. In [27.17] it is experimentally shown that it works quite well. The penalty for the digital wavelength can be mitigated by launching slightly more optical power to compensate for the optical power depletion.

In [27.8], methods based on line codes applied to dummy packets that can be sent when system utilization is low are described (but are not mandatory in the standard). The line code, for example a Manchester code, shapes the spectrum of the digital signal such that the lower frequencies are suppressed, reducing overall optical power at these lower frequencies in the digital downstream signal at 1490 nm, and thus reducing the Raman pumping of the analog video signal at 1550 nm.

27.2.4 Wavelength Multiplexing

As mentioned earlier, bidirectional traffic can be implemented by wavelength multiplexing of an up- and downstream wavelength. The large wavelength distance between up- and downstream enables low-loss implementation of the needed filter with satisfactory isolation values to avoid penalties of crosstalk from one signal to the other. In the case of a video overlay, additional filters are needed to combine the downstream video wavelength at OLT and separate it at ONU. Also, so-called blocking filters are needed to block the video wavelength in ONUs that are not video-enabled; otherwise the digital receiver will also receive the video signal, which will jam the digital downstream data. The video overlay signal is relatively strong and is closer to the digital data wavelength than the upstream signal, so the filter slope and isolation values for these blocking filters are quite stringent. A successful diplexer technology for achieving satisfactory isolation and blocking filter values is called BiDi technology (bidirectional)

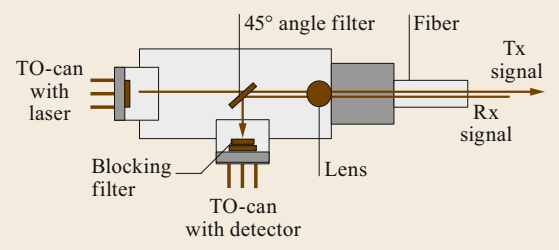


Fig. 27.9 BiDi technology for single-fiber bidirectional operation

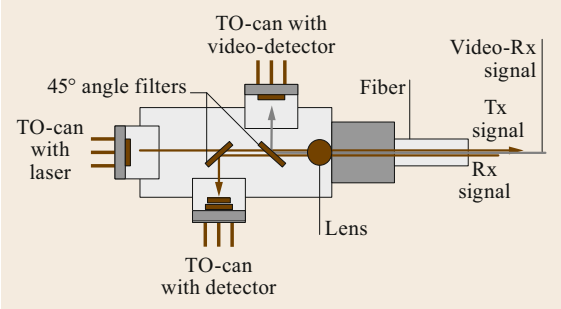


Fig. 27.10 3-port BiDi technology for single-fiber bidirectional operation and downstream video overlay

for just data, and three-port BiDi (triplexer) for additional downstream video, and is based on micro-optics. The BiDi module is compact and is based on a low-cost transistor outline (TO) package, namely a TO-can, containing a laser or detector and WDM filters at a 45° angle to separate the upstream and downstream wavelength bands. In the video overlay case, a second receiver and another filter are added. Figures 27.9 and 27.10 show a schematic of a 2-port and a 3-port BiDi, respectively.

27.3 Access Control Technologies in TDM-PON

In TDM-PON systems, a special mechanism is needed, implemented on top of the physical layer, to realize the point-to-multipoint communication between the OLT and several ONUs. This subsection provides its overview.

27.3.1 TDMA Control

One of the most unique and important technologies in TDMA-PON is the TDMA control mechanism, whose role is to avoid collisions among upstream signals from different ONUs. Figure 27.11 illustrates its general scheme. The left-hand side of the figure shows how downstream signals are sent from the OLT to three ONUs. As explained in Sect. 27.1 with Fig. 27.1, downstream signals to all the ONUs are multiplexed in the time domain. In addition to downstream data, the OLT sends upstream time-slot assignment information to each ONU. The right-hand side of Fig. 27.11 shows how upstream signals are sent from the three ONUs to the OLT. As illustrated, each ONU sends its upstream data in the time slot assigned by the OLT. This is done repeatedly, so that point-to-multipoint communication is established between the OLT and the number of ONUs.

27.3.2 Dynamic Bandwidth Assignment/Allocation (DBA)

DBA is the technology that enables flexible sharing of the total transmission bandwidth among all ONUs in TDM-PON systems. Thanks to DBA, a specific ONU can obtain a higher transmission bandwidth when other ONUs do not require bandwidth. While DBA can be implemented in various ways, two major methods are status-reporting DBA (SR-DBA) and traffic-monitoring DBA (TM-DBA). In SR-DBA, each ONU reports the buffer occupancy, i.e., the amount of the data waiting for the bandwidth allocation, to the OLT, in addition to the upstream data within the assigned time slot. The OLT assesses the requests received, i.e., reports, from all the ONUs, then assigns the bandwidth for each ONU considering priority, fairness, etc., and sends the time-slot assignment information to the ONUs as part of the downstream signals. In TM-DBA, OLT monitors statistical information regarding the amount of upstream traffic per traffic flow and predicts an optimal assignment in the next cycle using that information. Figure 27.12 illustrates the two methods.

Because fairness and throughput of the TDM-PON system are highly dependent on the DBA algorithm, there are many reports to optimize it [27.18–21].

27.3.3 TDMA Control in G-PON

G-PON implements the TDMA control mechanism in the G-PON Transmission Convergence (GTC) layer as specified in ITU-T Recommendation G.984.3 G-PON: Transmission convergence layer specification. In the GTC layer, OLT sends downstream GTC frames to all ONUs while each ONU sends upstream GTC frames

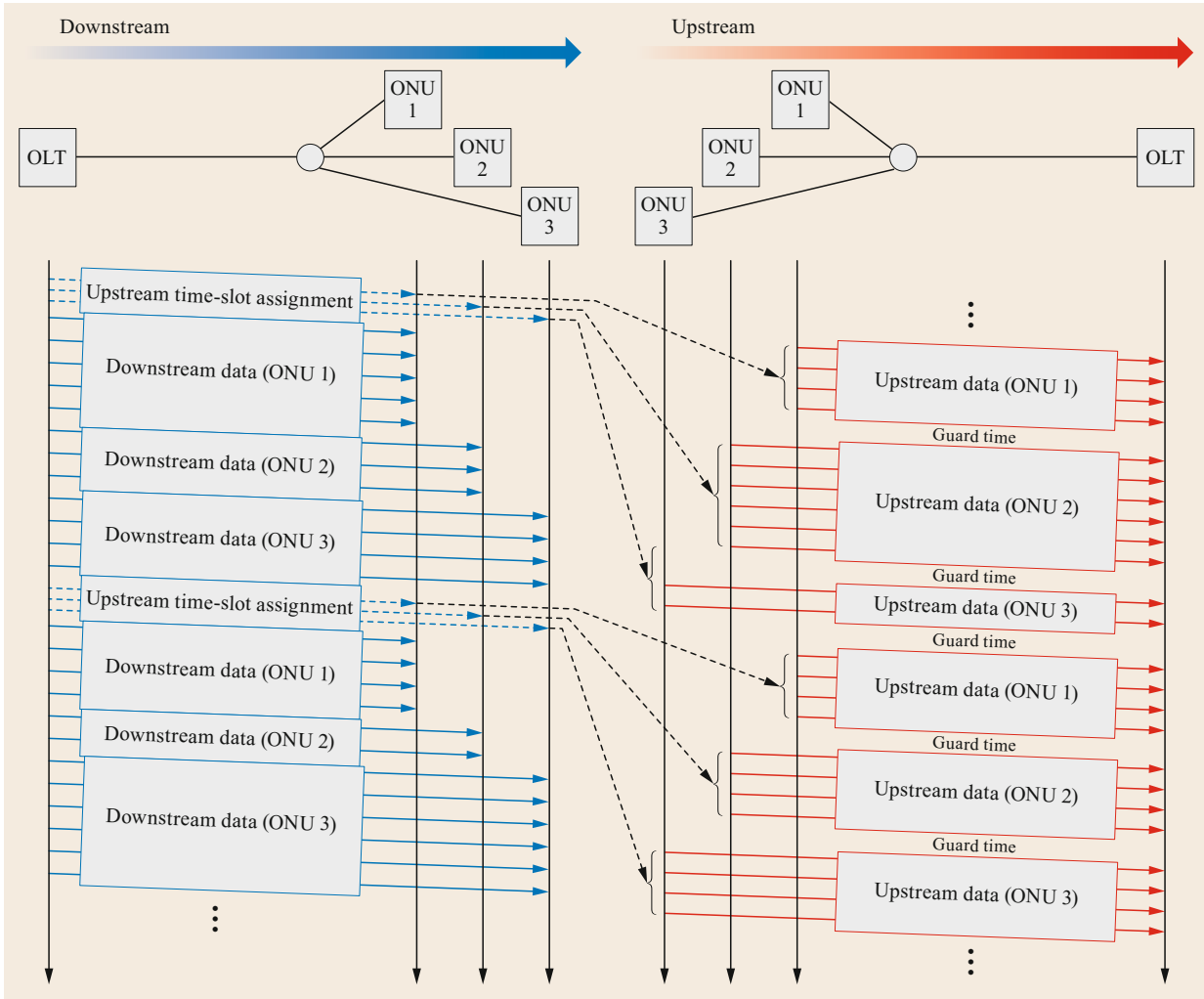


Fig. 27.11 TDMA control

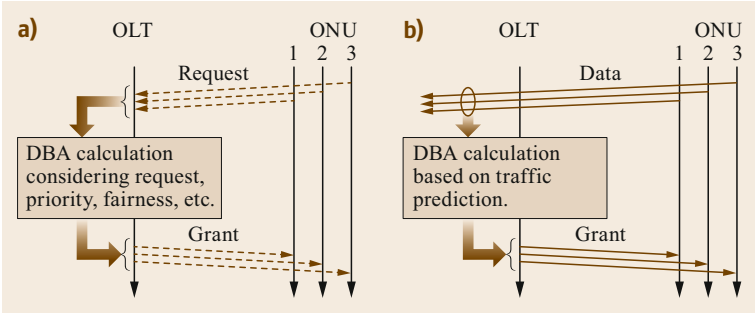


Fig. 27.12a,b DBA methods: **(a)** status-reporting DBA; **(b)** traffic-monitoring DBA

to the OLT. Figure 27.13 illustrates the frame structure of the downstream and upstream GTC frames, as well as the concept of the access control in G-PON. Downstream and upstream GTC frames have a fixed length of 125 μs. The downstream GTC frame consists of the GTC header section, called the physical control block downstream (PCBd), and the GTC payload section. The upstream GTC frame consists of multiple ONU bursts. Each ONU burst contains the upstream physical layer overhead (PLOu) section and one or more bandwidth

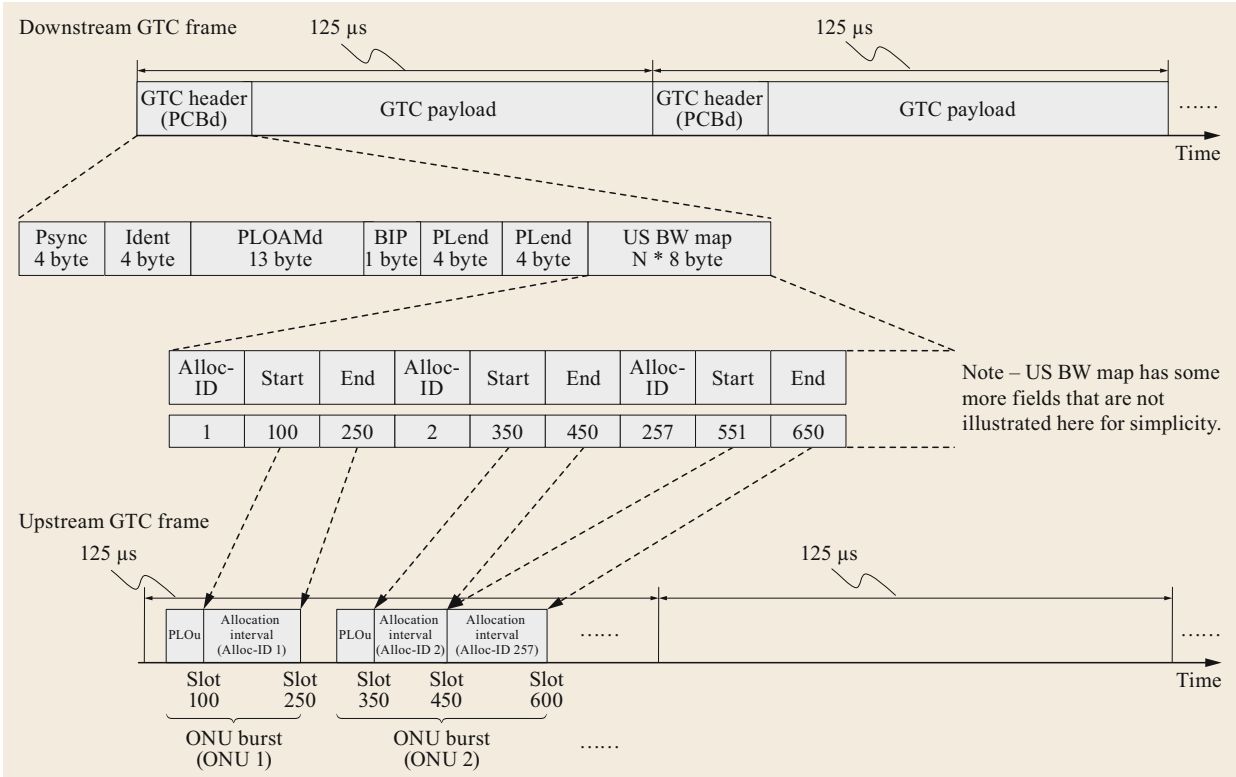


Fig. 27.13 Frame structure and concept of MAC in G-PON

allocation interval(s), each of which is associated with a particular allocation ID (Alloc-ID).

Transmission timing of each allocation interval is controlled using the upstream bandwidth map (US BW map) in PCBd in the downstream GTC frame. As shown in Fig. 27.13, the US BW map contains the start-time and end-time information of each allocation interval. Each ONU sends upstream ONU bursts according to this information, thus avoiding collision with other bursts and realizing TDMA in G-PON.

In addition to the US BW map, PCBd contains a number of other fields, including physical synchronization (Psync), IDENT, Physical-Layer Operations, Administration and Maintenance for downstream (PLOAMd), bit-interleaved parity (BIP) and payload length downstream (PLend). Psync is used for discriminating the starting position of each GTC frame, i.e., for the frame synchronization. IDENT contains a flag (used or non-used) for forward error correction (FEC) and the superframe counter. The superframe counter is incremented by one per frame and is used for synchronizing the timer of each ONU to OLT, and as a counter in the G-PON encryption process. PLOAMd is the field containing commands for operation, administration and maintenance (OAM) in the downstream-direction physical layer. The BIP carries FEC parity for detecting error(s) in reception. The PLend is the field used to specify the length of the preceding US BW map field.

While the length of the GTC frame is fixed as 125 µs, the G-PON encapsulation method (GEM) is specified on top of that to provide a variable-length framing mechanism for the transport of various types of client data as well as OMCI (ONU Management and Control Interface) messages for the management and control of ONUs. For example, Ethernet frames are encapsulated with GEM and mapped onto the GTC frames. Each GEM header contains a port ID, with which each ONU identifies the GEM frames it should receive. Figure 27.14 provides a general schematic of the GEM.

Figure 27.15 shows the structure of the ONU burst illustrated in Fig. 27.13. PLOAMu (physical-layer operations, administration and maintenance for upstream) is the field containing commands for OAM in the upstream in the physical layer. DBRu (dynamic bandwidth report upstream) is used to request bandwidth from the OLT. In other words, ONU uses DBRu to report the amount of traffic waiting for transmission in the specific Alloc-ID to OLT. OLT assesses the reported bandwidths

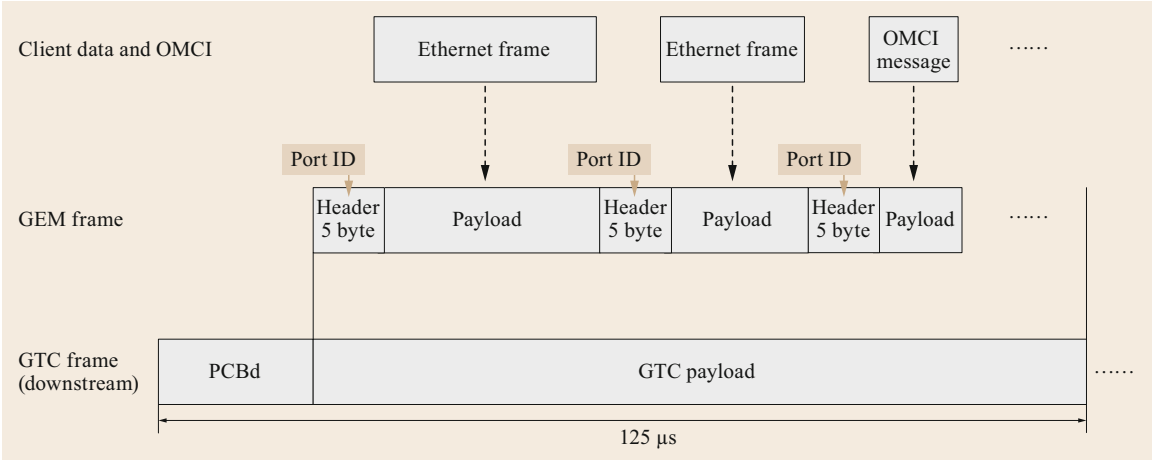


Fig. 27.14 G-PON Encapsulation Method (GEM)

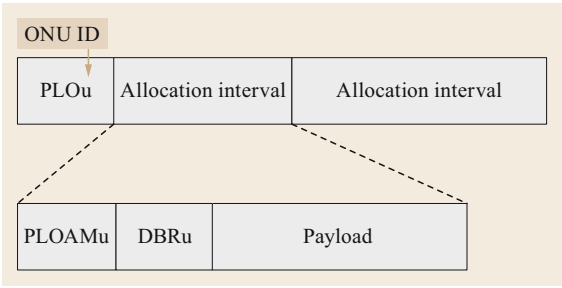


Fig. 27.15 ONU burst structure in G-PON

in assigning the next allocation intervals. This method is used to realize DBA in G-PON.

27.3.4 TDMA Control in E-PON

E-PON is specified as a part of the Ethernet specification in IEEE 802.3 Standard for Ethernet. Thus, how to implement the TDMA control in E-PON is specified as part of the Ethernet media access control (MAC) specification.

Figure 27.16 shows formats of the Ethernet frame and the E-PON frame, respectively. As shown, the Ethernet frame is transmitted in E-PON simply by adding a logical link identifier (LLID) to its preamble in front of the frame header that contains destination address (DA), source address (SA) and so on. A unique LLID or a set of unique LLIDs is allocated with each ONU for identification of the logical link(s) corresponding to the ONU. SFD and SPD are start-of-frame delimiter and start-of-packet delimiter to identify and synchronize to the frame. FCS is a frame check sequence for error detection information for the E-PON frame.

Figure 27.17 is a general schematic of multipoint MAC control, i.e., TDMA control in E-PON. As shown, transmission timing is controlled by GATE messages sent from OLT to ONUs. Each GATE message contains one or more sets of grant information, i.e., time-slot-allocation information, as assigned to a specific LLID. The grant information consists of start time and length (forming a time slot), which are precalculated in the OLT to avoid collision with other allocations.

Each ONU sends upstream data frames as well as REPORT messages within the granted time slots. Each REPORT message contains one or more sets of queue-length information, i.e., information regarding the amount of data waiting for transmission, of a specific LLID. Namely, each ONU uses the REPORT message to request bandwidth from the OLT. The OLT allocates the next time slots considering the reported queue lengths, and notifies the allocations in the next GATE messages to ONUs. This method can realize DBA in E-PON.

How to exchange GATE and REPORT messages in detail is specified as a part of the multipoint control protocol (MPCP) in E-PON. MPCP messages, including GATE and REPORT messages, are defined as one category of the Ethernet frames (IEEE 802.3 frames).

Operation, administration and maintenance (OAM) of ONUs are implemented through OAM messages, which are defined as a category of the Ethernet frames as well.

27.3.5 Process for Initial Connection of an ONU

While Sects. 27.3.1 to 27.3.3 describe how OLT communicates with already connected ONUs in TDMA-PONs, a special process is needed when one or more ONUs are initially connected, or reconnected, to OLT. This is because OLT needs to know the round-trip de-

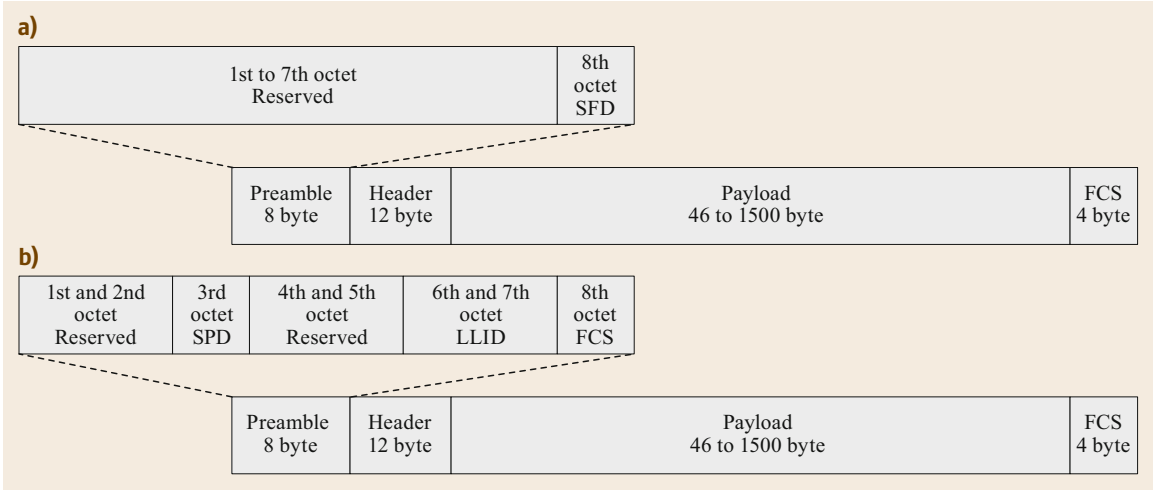


Fig. 27.16a,b Formats of (a) Ethernet frame and (b) E-PON frame

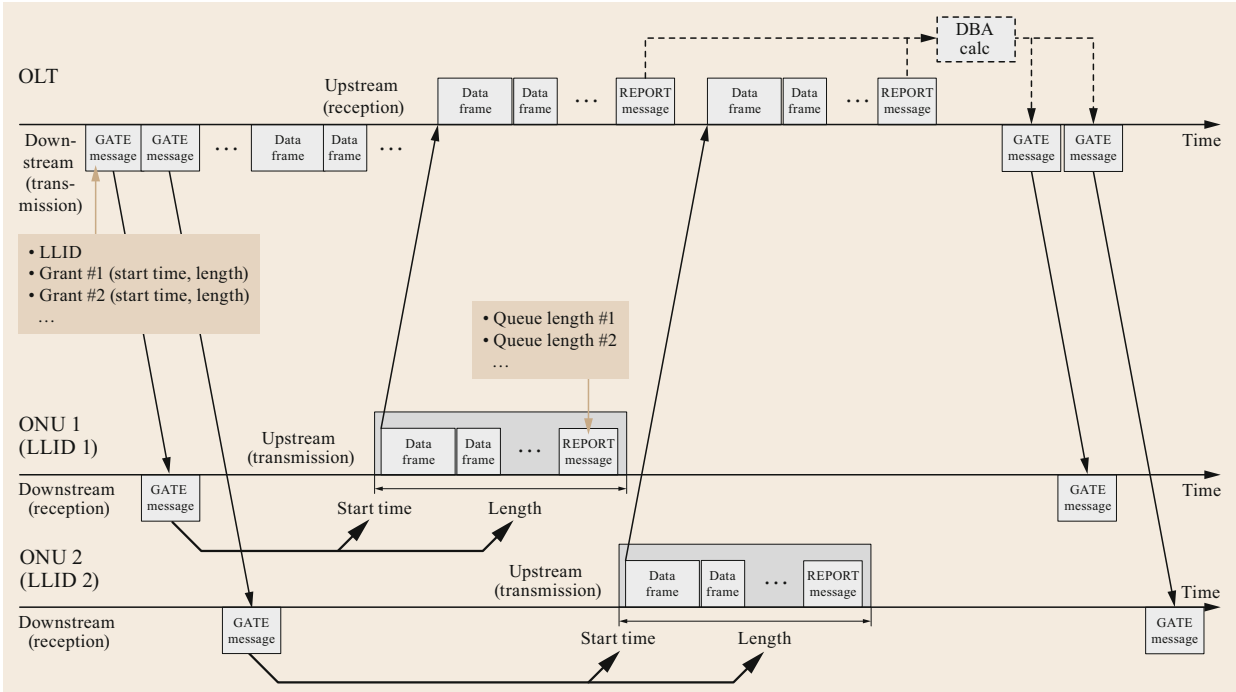


Fig. 27.17 Multipoint MAC control in E-PON

lay of each ONU for conducting the precise time-slot allocation necessary in TDMA. The round-trip delay, which includes round-trip propagation delay between OLT and an ONU, varies depending on the distance between the OLT and the ONU, and on processing delay in the ONU, both of which are initially unknown.

Figure 27.18 is a general schematic of the initial connection process in TDMA-PON, such as G-PON and E-PON. OLT periodically breaks its normal reception, i.e., stops allocating upstream time slots to already connected ONUs, and opens a short window for new ONUs to send a frame to the OLT. This window is called the quiet window in G-PON and the discovery window in E-PON. A new ONU identifies the window through reception of the serial number request as a broadcast bandwidth allocation in G-PON, and through reception of a broadcast message called the discovery gate in E-PON. The new ONU then sends

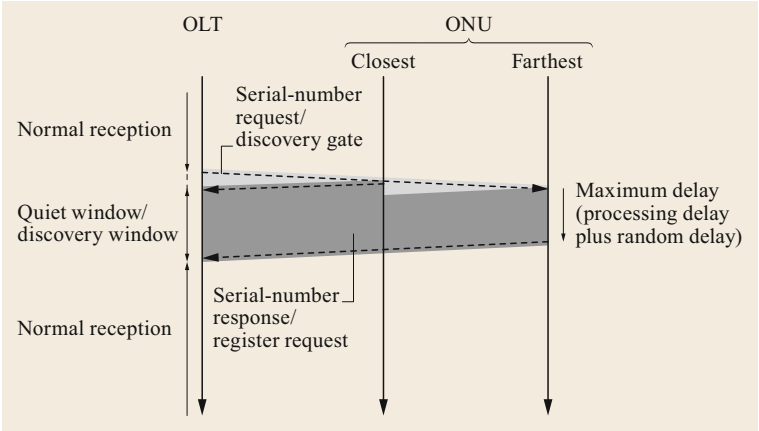


Fig. 27.18 Window for initial connection

a serial number response (G-PON) or a register request (E-PON). Here, collision can occur, especially when many ONUs attempt to newly connect at the same time. To reduce the probability of such a collision, a random delay is added before sending the serial number response/register request in the ONU. When a collision occurs in spite of that technique, the ONU retries sending the serial number response/register request at the next chance, i.e., the next quiet/discovery window.

Triggered by this initial process, some additional unicast communication runs between OLT and the ONU for further synchronization. In the case of G-PON, a process called *ranging* runs using another quiet window to determine the round-trip delay. In the case of E-PON, round-trip delay is captured in the initial process, and subsequent handshakes run with additional unicast communication. Thus, the synchronization where round-trip delay against new ONU(s) has been taken into account is completed and normal data exchange starts. The whole process before the start of normal data exchange is called activation in G-PON and discovery in E-PON.

27.4 XG(S)-PON and 10G E-PON

The basic idea of 10 Gb/s PON systems, i.e., XG-PON, XGS-PON and 10G E-PON, is to increase the transmission bitrate (line rate) of G-PON and E-PON (i.e., 1G E-PON) to 10 Gb/s. Specifically, the transmission bitrates of XGS-PON are 9.95328 Gb/s for both upstream and downstream, while those of XG-PON are 2.48832 Gb/s for upstream and 9.95328 Gb/s for downstream. The transmission bitrates of 10G E-PON are 10 Gb/s for downstream and two options, 1.25 and 10.3125 Gb/s, for upstream. Note that these are not the payload bandwidth, i.e., how much traffic the system can carry, but the transmission bitrate in the physical layer.

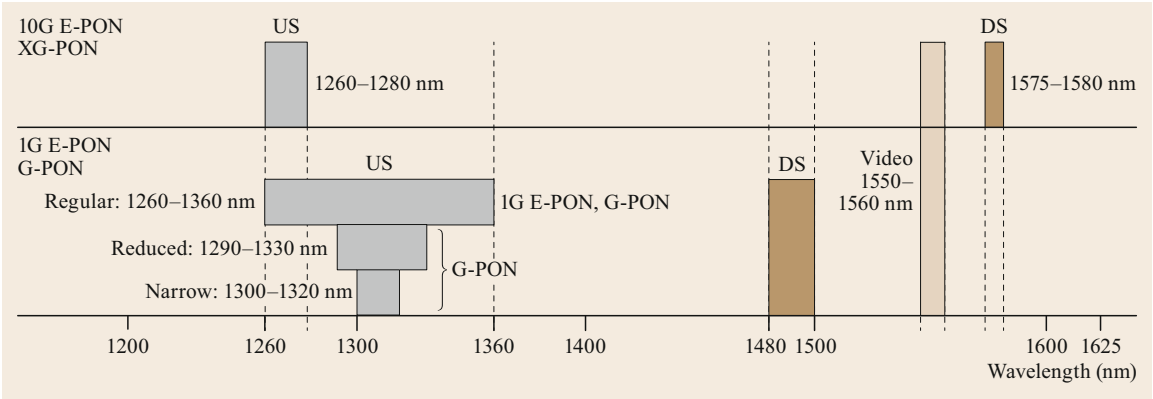


Fig. 27.19 Wavelength allocation of G-PON, XG-PON, E-PON and 10G E-PON (US: upstream, DS: downstream)

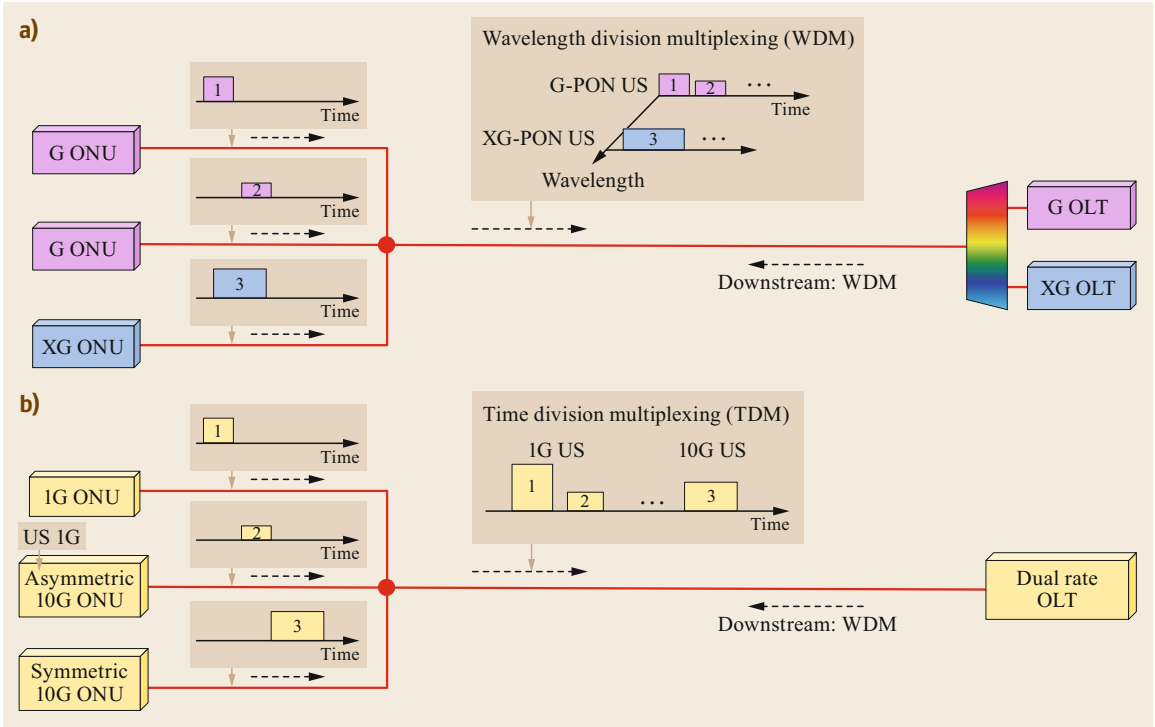


Fig. 27.20a,b Coexistence of (a) G-PON and XG-PON, and (b) 1G E-PON and 10G E-PON (US: upstream, DS: downstream)

Figure 27.19 summarizes the wavelength allocations of G-PON, XG-PON, 1G E-PON and 10G E-PON. Both the XG-PON and 10G E-PON downstream wavelengths are specified in the same region to avoid overlapping the G-PON and 1G E-PON downstream wavelengths. The XG-PON upstream wavelength is specified to avoid overlapping the reduced and narrow options of the G-PON upstream wavelength. Therefore, XG-PON can coexist with G-PON in the same ODN using wavelength-division multiplexing (WDM), provided the G-PON ONUs generate upstream signals in compliance with the reduced or narrow band options; Fig. 27.20a shows an implementation of this coexistence. On the other hand, the 10G E-PON upstream wavelength overlaps to the 1G E-PON upstream wavelength. As shown in Fig. 27.20b, TDM is assumed to accommodate the coexistence of 1G E-PON ONUs under the same ODN in the 10G E-PON. In terms of realizing the coexistence of 1G E-PON and 10G E-PON through TDM, the 10G E-PON upstream wavelength region can be the same as that of 1G E-PON. However, the 10G E-PON upstream wavelength region is specified to be the same as that of XG-PON rather than that of 1G E-PON. This allows several interesting possibilities, e.g., coexistence between G-PON (reduced or narrow option) and 10G E-PON and allocation of 1280–1360 nm to a future PON.

As for XGS-PON, the basic wavelength set (upstream and downstream) is defined to be the same as the XG-PON upstream and downstream wavelength regions. In addition, an optional wavelength set is defined the same as the G-PON upstream and downstream wavelength regions. That is, the basic and optional wavelength sets of XGS-PON allow coexistence with G-PON and XG-PON, respectively, through WDM.

A new feature introduced in XG(S)-PON and 10G E-PON is the mandatory support of forward error correction (FEC). All the FECs implemented in these PON systems use Reed Solomon (RS) coding [27.22]. Figure 27.21 illustrates the RS(255, 239) and RS(255, 223) codes, as examples of RS-code implementation. The code-block length is 255 byte in both cases, while the parity field is 16 and 32 byte for RS(255, 239) and RS(255, 223), respectively. The longer parity field provides stronger FEC in RS(255, 223), which improves the bit error ratio (BER) from 1×10^{-3} to 1×10^{-12} [27.23] while RS(255, 239) improves the BER from 1.8×10^{-4} to 1×10^{-12} [27.24], insofar as the bit errors are randomly distributed.

The 9.95328 Gb/s bitstream of XG-PON and XGS-PON, i.e., the XGS-PON upstream, the XG-PON

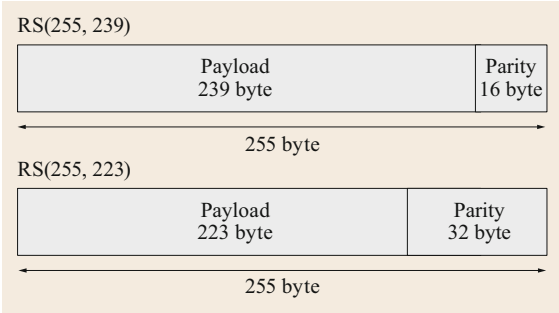


Fig. 27.21 FEC code blocks

downstream and the XGS-PON downstream, uses the RS(248, 216) code, which is generated as follows: the RS(255, 223) code is used as the base, and its payload is truncated after calculation of the parity bytes, so that a certain number of the FEC code blocks yields a physical layer frame (PHY frame) having a fixed length of 125 μs at the specified bitrate. In the 2.488 Gb/s upstream of XG-PON, the RS(248, 232) code is used; this is a truncation of RS(255, 239). The 10.3125 Gb/s bitstream of 10G E-PON, i.e., the 10G E-PON upstream and downstream, uses the RS(255, 223) code. The upstream FEC in XG-PON and XGS-PON can be turned on and off, in an ONU-by-ONU manner, through the selection of the burst profile, which allows OLT to set various parameters related to the ONU burst in each ONU. The burst profile is another new feature introduced in the standardization of XG-PON. The length and pattern of preamble and delimiter (see Sect. 27.2.2) can be set through the burst profile as well.

In XG-PON and XGS-PON, the PHY frame described above carries an XGTC (XG-PON transmission convergence) frame and FS (framing sublayer) frame, respectively, as the FEC payload. Both the XGTC frame and the FS frame are defined based on the GTC frame, i.e., the basic functions are the same, while they in-

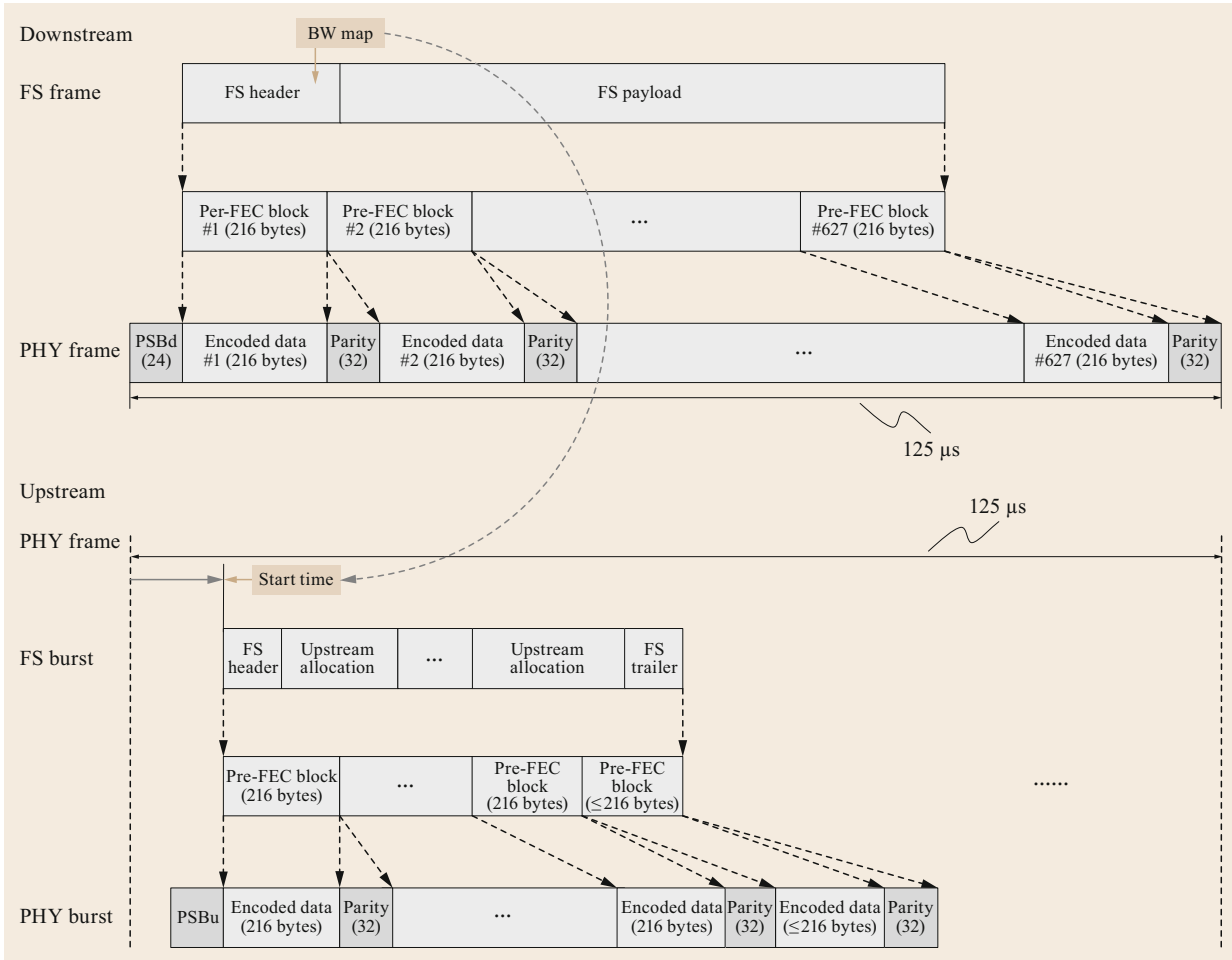


Fig. 27.22 Relationship of PHY frame to FS frame and FS burst in XGS-PON

Table 27.3 Physical parameters of ODN for XG-PON and XGS-PON

Item	Unit	Specification
Fiber type	–	ITU-T G.652
Attenuation range	dB	Class N1: 14–29 Class N2: 16–31 Class E1: 18–33 Class E2: 20–35
Maximum differential optical path loss	dB	15
Maximum fiber distance	km	20

clude a number of changes in their precise structures. As for the upstream burst from each ONU, the ONU burst is defined in G-PON (see Sect. 27.3.3). In XG-PON, XGTC burst and PHY burst are defined as the ONU burst before FEC encoding and that after FEC encoding. In XGS-PON, FS burst and PHY burst are defined as the ONU burst before FEC and that after FEC.

Figure 27.22 illustrates the relationship of the PHY frame to the FS frame and the FS burst described above for XGS-PON. PSBd and PSBu are downstream and upstream physical synchronization block sections, respectively. The PSBd contains a pattern for frame synchronization, the superframe counter (see Sect. 27.3.3) and a PON identifier (PON-ID) to identify the PON port among the huge number of PON systems possibly operated. PSBu contains preamble and delimiter (see Sect. 27.2.2).

Table 27.3 shows physical parameters of ODN for XG-PON and XGS-PON; see Sect. 27.2.1 for the same table for B-PON and G-PON. The ODN classes defined include Nominal 1 (N1), Nominal 2 (N2), Extended 1 (E1) and Extended 2 (E2), in which the acceptable optical losses between OLT and ONUs are 29, 31, 33 and 35 dB at maximum, respectively. Fiber type, differential optical path loss and fiber distance are defined the same as those for B-PON and G-PON. The introduction of FEC is one of the keys to keep accommodating such high optical losses when increasing the bitrate to 10 Gb/s.

Figure 27.23 illustrates the power budget design for XGS-PON class N1, in which the use of FEC is manda-

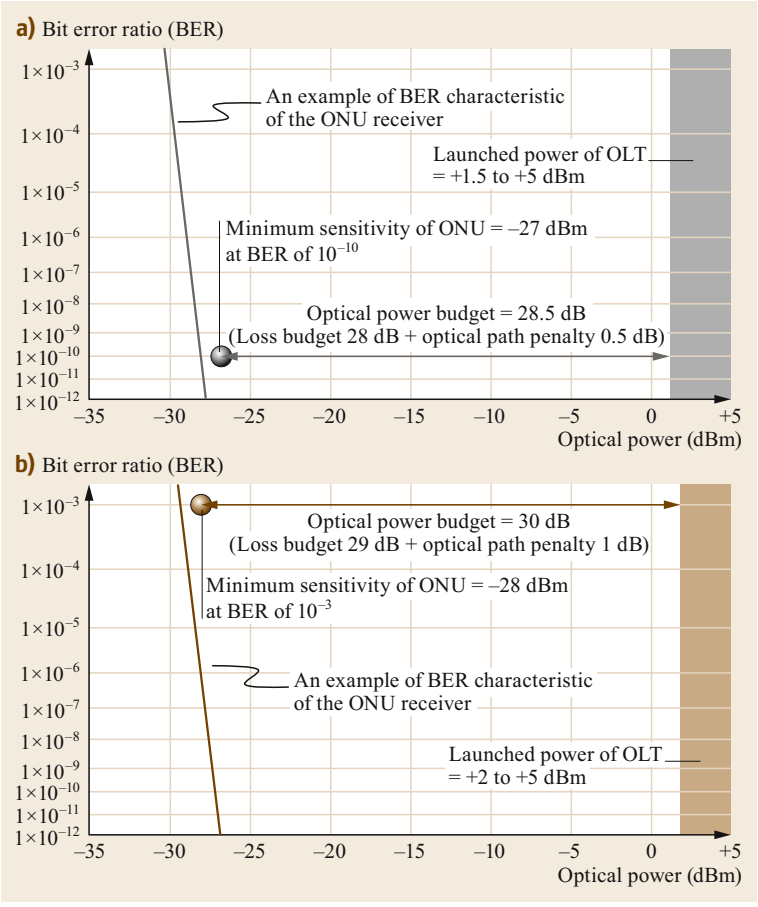


Fig. 27.23a,b Optical level design of **(a)** G-PON class B+ and **(b)** XGS-PON class N1

tory, comparing to that for G-PON class B+, in which FEC is not assumed to be implemented; only the downstream design is illustrated for simplicity. The use of FEC allows the system to accept a higher (worse) BER in the receiver, so a lower (better) minimum sensitivity can be specified. Therefore, the total optical power budget between OLT and ONU can be increased while keeping the transmitter launched power at a reasonable value for more cost-effective implementation. It should be noted that the optical power budget between OLT and ONU does not correspond to the optical loss budget, i.e., the maximum value in the attenuation range, between OLT and ONU, because the achievable sensitivity in the receiver gets worse when receiving the signal after the fiber transmission, e.g., due to the signal pulse broadening induced by the fiber dispersion. This degradation factor in the receiver sensitivity is defined as the optical path penalty.

27.5 Security and Privacy

There are several ways by which data of a legitimate PON user can fall into the hands of a malicious user, or by which a malicious user can pose as a legitimate user, and by which a malicious user can disturb the PON.

In the downstream, a PON is a broadcast network, which means that every user receives the data of every other user. Due to the directional nature of PON, the upstream direction is secure and can be used to send encryption keys. However, a malicious user could pretend to be a legitimate user and attack the PON in that way. Also, a malicious user could overwrite the signal of another user by transmitting at higher optical power.

Due to the nature of the fiber plant, reflections of signals resulting from dirty, damaged or open connectors could also end up at a malicious user.

One solution that resolves many of the security and privacy issues is ONU authentication and data encryption.

For example, the G-PON standard ITU-T G.984.3 requires the use of the Advanced Encryption Standard (AES). A description of the AES algorithm can be found in [27.25].

Encryption is done before FEC encoding, and only the GEM frame is encrypted, not the header.

The OLT initiates the key exchange by sending the Request_Key message in the PLOAM channel. The ONU responds by generating, storing and sending the key using PLOAM messages. After the OLT successfully receives the key, it stores it. Encryption keys need to be renewed for optimal security; this is also controlled by the OLT. The OLT chooses a frame number in the future, which will be the first frame using a new key, and communicates this to the ONU by sending the Key_Switching_Time message three times to ensure high probability of correct reception of the message. A similar security algorithm is specified for E-PON in [27.26].

As described earlier, in addition to the eavesdropping threat, malicious users can pretend to be legitimate users. Authentication is the method applied in standardized PONs to avoid this. Reference [27.26] describes three different mechanisms for authentication in E-PON. The first is authentication based only on the MAC address, the second is based only on the logical ID, and the third is based on MAC address or logical ID (if MAC address fails, logical ID can be used). Only after authentication is successfully completed is transmission of data frames allowed from OLT or ONU. Also, in G-PON, ONUs receive a unique ONU-ID assignment from the OLT after the ONU sends its serial number.

The final threat listed above is that of malicious users disturbing the PON by jamming the signal of other users. In [27.26] it is assumed that the E-PON infrastructure is physically secured and inaccessible in order to avoid such events, and no algorithms are described to protect against this threat.

27.6 Survivability

As illustrated in Fig. 27.24, various types of PON protection schemes are available for TDM-PON. Because a cut of the feeder fiber (the fiber between OLT and the splitter) or a failure in the OLT results in service breakdowns in all the ONUs connected to the OLT in PON, protection against such a cut/failure is very important.

Type A is a simple scheme that protects only the feeder fiber. As illustrated in Fig. 27.24, this is realized by setting a $1:2$ optical switch next to the OLT, using a $2:n$ optical splitter instead of a $1:n$ optical splitter, and employing working and protection feeder fibers between the optical switch and the $2:n$ optical splitter. No

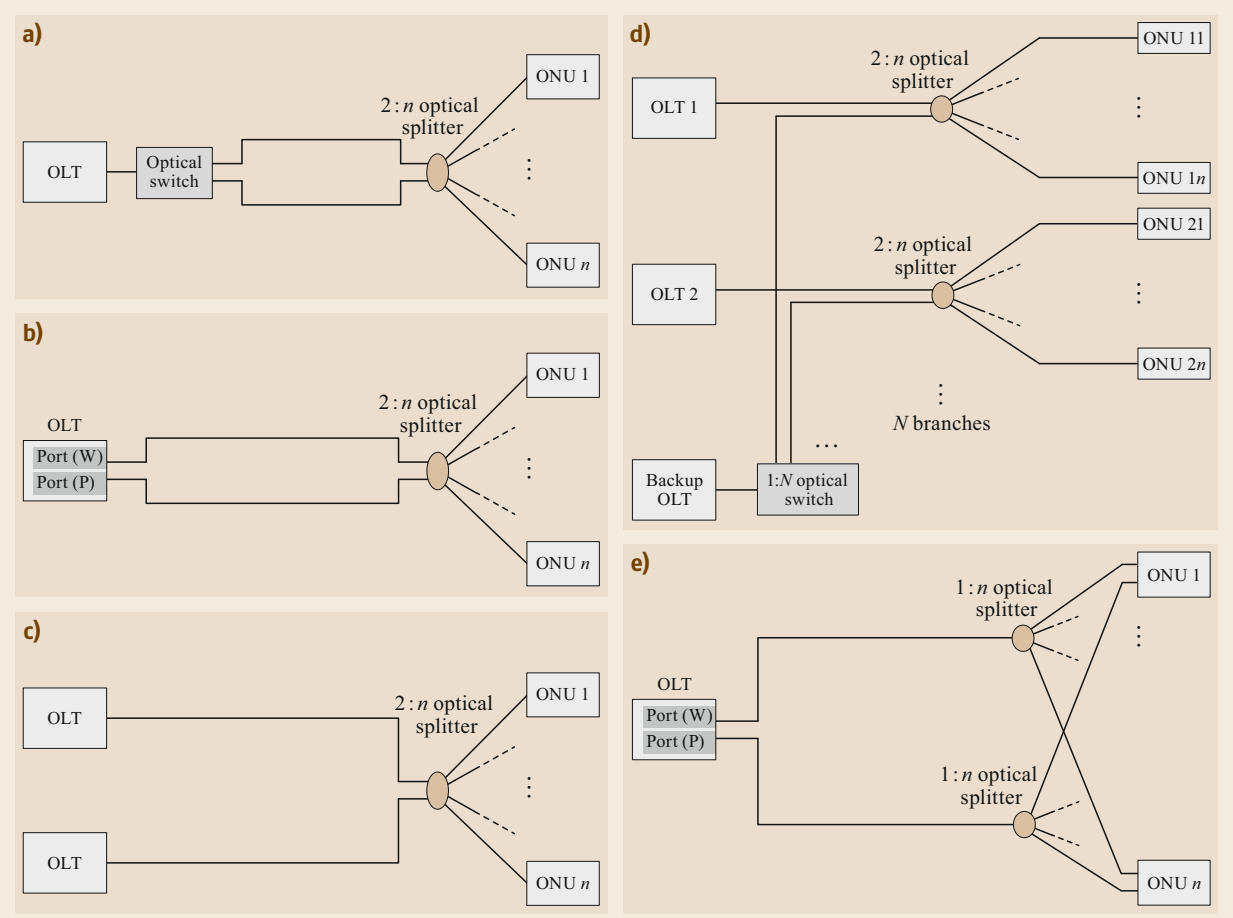


Fig. 27.24a–e PON protection options: **(a)** type A, **(b)** type B (basic), **(c)** dual-parent type B, **(d)** type B with N:1, **(e)** type C (W: working, P: protection)

additional component or function is necessary in either OLT or ONUs to realize this scheme.

Type B is a cost-effective scheme to protect both the feeder fiber and the OLT port while not requiring any additional ONU components or functions. In the case of type B, the two OLT ports, i.e., the working port and the protection port, must not generate the downstream signals simultaneously; otherwise the two downstream signals are mixed at the optical splitter and become unreadable, as can be understood from the figure. In this scheme, the protection port must stand by silently—the so-called cold standby—and start generating the downstream signal only after a failure is detected. Fast protection switching, e.g., within several tens of milliseconds, can be achieved by simplifying or omitting the remeasurement of round-trip delay between the protection port of the OLT and each ONU, and by properly controlling the working and protection ports of the OLT when a failure occurs [27.27, 28].

Setting the working and protection ports in different locations yields dual-parented type B protection. Dual-parented type B can secure the service even when all the equipment in one central office becomes unavailable, e.g., due to a disaster or a power breakdown. Assuming that the working and protection OLTs are connected to a common Ethernet switch in the backhaul network, Ethernet linear protection can be implemented to assist a proper protection switching [27.29].

Type B with $N:1$ is another variant of type B. The aim of this variant is to provide greater cost-effectiveness compared with the basic type B, as the backup OLT is shared among several PON systems. As illustrated in the figure, a $1:N$ optical switch is used to realize the sharing.

Type C duplicates the whole PON section between the OLT and each ONU. In this case, both the working port and the protection port of OLT can generate the downstream signals simultaneously be-

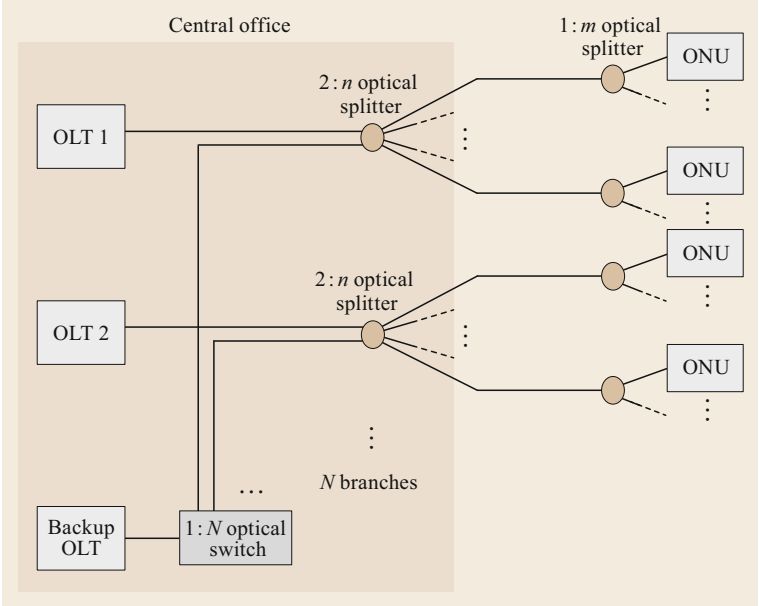


Fig. 27.25 In-office implementation of PON protection type B with $N:1$ to address OLT-card failure

cause the two downstream signals do not meet in the optical section, as can be understood from the figure. Here, the protection OLT can stand by while generating the downstream signal—the so-called hot standby. Moreover, depending on the implementation of the ONU, the protection port of each ONU can communicate with the protection port of the OLT independently from the communication between the working port of the ONU and the working port of the OLT, thus carrying extra traffic during normal operation.

Figure 27.25 illustrates an in-office implementation of type B with $N:1$ to address OLT-card failure [27.30]. Although this does not address the cut of feeder fiber, it can be a cost-effective solution to address OLT-card failure while keeping the existing non-duplicated outside plants (fibers and splitters) where PON systems have already been deployed.

27.7 Energy Efficiency in PON

Several reasons can be given to support high energy efficiency in PON. Sustainability is one reason, as it has been estimated that the energy consumption of the information and communications technologies (ICT) ecosystem has approached 10% of electricity generation worldwide, with the internet accounting for 10% of global consumption in 2013 [27.31], an amount comparable to that of the airline industry. Of all the networks that make up the internet, access networks have the highest energy usage per user, as they are shared among the least number of users, and home equipment is not even shared at all. Thus, any energy efficiency applied at the CPE (customer premises equipment) will have a significant impact on the overall energy consumption of the network.

Second, during emergencies, battery operation of user equipment might be needed, which also profits from lower power consumption of this equipment.

Third, energy efficiency is also important for economic reasons. OLTs are located at the central office (CO), where the achievable PON-port density (which determines cost per port) is limited by power consumption and resulting heat generation.

Finally, practical considerations such as limited power availability at remote node locations (e.g., FTT-curb, FTT-pole, FTT-distribution point) support the need for high energy efficiency of PON equipment.

In the following sections, we will describe the use of sleep modes in PON, which is a standardized method for increasing energy efficiency. We will also describe bit-interleaving PON, which has been proposed as a low-power-consumption PON architecture [27.32].

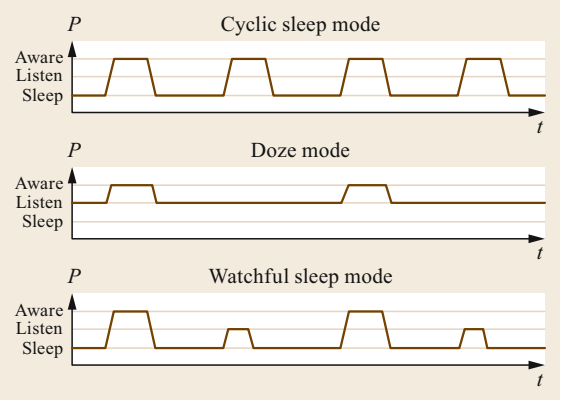


Fig. 27.26 Three sleep modes as defined in ITU-T G.987.3

27.7.1 Sleep Modes

Sleep modes are used to reduce the power consumption of a network. When a device does not need to transmit or receive data, it can go to sleep and thus consume less power.

Sleep modes are recognized as an effective method for reducing energy consumption and have been standardized to reduce energy consumption in a PON.

In ITU-T Recommendation Series G.987 (XG-PON), G.989 (NG-PON2) and G.984 (G-PON), three sleep modes have been standardized for PON: doze mode, cyclic sleep mode and watchful sleep mode. Watchful sleep mode was added later as a compromise between doze and cyclic sleep mode for an overall higher power savings.

ONU receivers can still receive downstream data when in doze mode, but in cyclic sleep mode they cannot; in the watchful sleep mode the receiver periodically becomes active. All three modes can be statically provisioned by the network management layer of the PON.

OLT and ONU together run a pair of power-management-state machines coordinated by signaling with PLOAM messages to operate the sleep mode algorithms. In summary, the ONU can have two main states, the Aware state and the LowPower state. In the Aware state, the ONU receives and transmits data. In the LowPower state, the ONU does not transmit data. Within the LowPower state, the ONU can be in Listen state (doze mode), in which an ONU can receive data, or it can be in the Asleep state (cyclic sleep mode), where the ONU receiver is inactive. Finally, the ONU can be in the Watch state (watchful sleep mode), where the ONU periodically checks the downstream traffic for wakeup indications from the OLT. See Fig. 27.26 for a schematic representation of the three sleep modes described in ITU-T G987.3 and G.989.3 [27.33, 34].

The OLT can enable and disable the different sleep modes in real time using the Sleep-Allow message. The ONU can signal the intent to change power mode with a Sleep_Request message. In addition, the OLT can expedite waking up a dozing or sleeping ONU using a forced wake-up indication (FWI) bit in the BW map allocation structure, in case it needs to.

IEEE also standardized sleep modes for Ethernet-based PONs. In IEEE standard 1904.1 [27.26], dynamic and static power-saving modes are described. In the current draft of 1904.1, two types of sleep modes are described for E-PON. The Tx sleep mode, where the ONU Tx is sleeping but the ONU Rx is still active (similar to doze mode in ITU-T PON), and TRx sleep mode, where both ONU Tx and Rx are sleeping. The sleep control signaling in E-PON is based on messages delivered using extension MAC control frames.

The SLEEP_ALLOW message is sent by OLT to request the ONU to enter the power-saving mode. The ONU sends the SLEEP_ACK message in response to OLT when the ONU is capable and is configured. The ONU can request a SLEEP_ALLOW message from the OLT by sending a SLEEP_INDICATION message. In E-PON, OLT and ONU can both initiate power-saving modes.

Similar to the forced wake-up in ITU-T PONs, E-PON ONUs should support the early wake-up function.

There are several challenges related to sleep modes. The main challenge is the trade-off between power savings and the quality of the user experience (or quality of service, QoS). The on/off transition time is significant, limits overall power savings and increases delays in the network. Sleep modes also need large packet buffers, which again increases power consumption. The watchful sleep mode has solved some of these challenges because it has the advantages of both the doze mode (i.e., better user experience due to fast transition between low- and high-power mode, but lower power savings) and the cyclic sleep mode (i.e., higher power savings, but worse user experience). For example, in [27.35] it has been shown that the watchful sleep mode indeed outperforms the doze mode and the cyclic sleep mode with respect to energy efficiency for low-traffic-load cases. Also, the downstream delay has been reduced, which is a main parameter for QoS. The downstream delay is even lower in the doze mode, but this is at the expense of the energy efficiency.

27.7.2 Other Power-Saving Schemes

In addition to sleep modes, other power-saving methods have been proposed in the research literature.

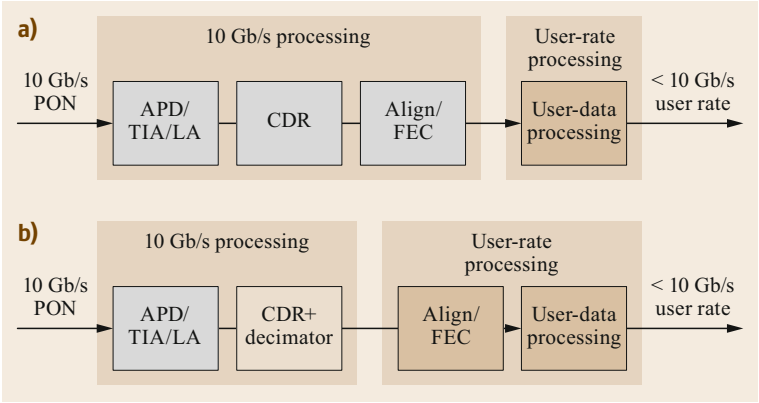


Fig. 27.27a,b ONU receive path architecture in standard TDM-PON **(a)** and bit-interleaving PON **(b)**

For example, in 2012, the GreenTouch Consortium [27.32] announced and demonstrated the bit-interleaved passive optical network (Bi-PON) concept, which can be used in addition to sleep modes to further reduce power consumption.

As in the downstream of a PON, all the data are passively split and distributed to all the users; each user, in order to recover its data, must process all the data before it can filter out the data that was meant particularly for its use. From an energy consumption point of view, this is very inefficient, because the large majority of the CPE processing is on data that are meant for other users. In the upstream, the processing at the CPE is more efficient because only user traffic of this CPE is handled.

In bit-interleaving PON, each user selects the relevant bits, and drops bits meant for other users immediately after the CDR, enabling further processing of the data at the user rate instead of the aggregate data rate. In other words, user traffic is already separated in the physical layer instead of in the protocol layers. Figure 27.27 shows how bit-interleaving PON compares with standard TDM-PON. In a conventional TDM-PON, the clock and data are recovered by a CDR immediately after the optical receiver front-end consisting of an APD, a transimpedance amplifier (TIA) and a limiting amplifier (LA). After the CDR, the data are aligned, and functions such as forward error correction (FEC) are operated on the full-rate bitstream. After FEC, the higher layers can process the data and determine which part of the data is meant for the particular user and can be further processed at the lower user rate.

In the case of bit-interleaving PON, instead of a conventional CDR, a CDR with a decimator is used. The decimator automatically drops data not meant for the particular user. As a result, the data can be processed at the lower user rate immediately after the clock and data recovery stage. Functions such as FEC and alignment can be run at much lower bitrates, and therefore the power consumption at the ONU is substantially reduced.

Dynamic power consumption is proportional to bitrate (frequency) and the amount of switching activity, making the bit-interleaving scheme quite effective, as has been shown in reference [27.36, 37]. Dynamic power consumption is the power consumption of a device when it is in operation, minus the power consumption when it is not in operation (static state). Static power consumption is mainly caused by leakage currents of the transistors in the circuit. In [27.36] a field-programmable gate array (FPGA) implementation was reported in which the dynamic downstream processing power of a 10G Bi-PON ONU was at least 30 times lower than a standard XG-PON ONU under all traffic conditions, including cases where sleep modes were also applied. In [27.37], an application-specific integrated circuit (ASIC) implementation using a 0.13 nm complementary metal–oxide–semiconductor (CMOS) is described; the paper estimates that integration reduces the dynamic power consumption ratio between Bi-PON and standard TDM-PON beyond 35 and up to 180 times for a 10 Mb/s user rate. To make the bit-interleaving principle work, an alternative PON protocol needs to be implemented, but as of now such a protocol has not yet been standardized.

In addition to sleep modes and bit-interleaving, lower power consumption can be achieved with the use of smaller-scale CMOS technology. The dynamic power consumption, primarily, will be lower because of higher integration and lower voltages. However, with the smaller-scale CMOS technology, leakage currents increase, which limits the reduction of the static power consumption.

27.8 Technologies Beyond 10G PON

Research into the topic of increasing the line rate of TDM-PON has been quite active in recent years. As cost targets are very stringent, in optical access the focus has been on reusing 10G optical components via the use of higher-order modulation schemes such as duobinary or PAM-4, or by applying electronic equalization methods to achieve 25, 40 or 50 Gb/s transmission [27.38]. Research activity has also sparked the IEEE PON standardization group to start standardizing a PON system with a 25 Gb/s line rate [27.39, 40]. Discussion and investigation into higher line rate has also started in the ITU-T PON standardization group FSAN.

Another direction of TDM-PON evolution is to realize a longer reach and/or higher split ratio, keeping the line rate as 10 Gb/s or higher. While optical amplification and digital coherent reception are the candidate technologies to realize these goals, issues include how to accommodate the burst-mode upstream signals and how to minimize the system cost. There are several challenges in realizing burst-mode optical amplification and burst-mode digital coherent reception for TDM-PON applications [27.41]. One approach for minimizing the system cost is to avoid using high-cost components at the ONU side, while sharing the cost of the optical amplifier and/or the digital coherent receiver among a large number of ONUs at the OLT side [27.42].

References

27.1 J.R. Stern, J.W. Balance, D.W. Faulkner, S. Hornung, D.B. Payne, K. Oakely: Passive optical local networks for telephony applications and beyond, Electron. Lett. **23**(24), 1255–1256 (1987)

27.2 K. Kumozaki: Optical access systems: Present state and future directions, NTT Tech. Rev. **6**(5), 1–7 (2008)

27.3 FSAN: https://www.fsan.org/

27.4 ITU-T Recommendation G.983 Series: Broadband passive optical network (2001)

27.5 F. Effenberger, H. Ichibangase, H. Yamashita: Advances in broadband passive optical networking technologies, IEEE Commun. Mag. **39**(12), 118–124 (2001)

27.6 G. Kramer, G. Pesavento: Ethernet passive optical network (EPON): Building a next-generation optical access network, IEEE Commun. Mag. **40**(2), 66–73 (2002)

27.7 IEEE Std 802.3ah-2004: IEEE standard for information technology – Local and metropolitan area networks – Part 3: CSMA/CD access method and physical layer specifications amendment: Media access control parameters, physical layers, and management parameters (2004)

27.8 ITU-T Recommendation G.984 Series: Gigabit-capable passive optical network (G-PON) (2008)

27.9 ITU-T Recommendation G.987 Series: 10 Gigabit-capable passive optical network (XG-PON) (2012)

27.10 IEEE 802.3av-2009: IEEE standard for information technology – Local and metropolitan area networks – Specific requirements – Part 3: CSMA/CD access method and physical layer specifications amendment 1: Physical layer specifications and management parameters for 10 Gb/s passive optical networks (2009)

27.11 ITU-T Recommendation G.9807 Series: 10 Gigabit-capable symmetric passive optical network (XGS-PON) (2016)

27.12 ITU-T Recommendation G.983.3 – Amendment 2: A broadband optical access system with increased service capability by wavelength allocation (2005)

27.13 ITU-T Recommendation G.984.2 – Amendment 2: Gigabit-capable passive optical networks (G-PON): Physical media dependent (PMD) layer specification amendment 2 (2008)

27.14 ITU-T Recommendation G.984.6: Gigabit-capable passive optical networks (GPON): Reach extention (2008)

27.15 X.Z. Qiu: Burst-mode receiver technology for short synchronization. In: *Opt. Fiber Commun. Conf.* (2013), https://doi.org/10.1364/OFC.2013.OW3G.4

27.16 M.R. Phillips, D.M. Ott: Crosstalk due to optical fiber nonlinearities in WDM CATV lightwave systems, J. Lightwave Technol. **17**(10), 1782–1792 (1999)

27.17 M. Aviles, K. Litvin, J. Wang, B. Colella, F.J. Effenberger, F. Tian: Raman crosstalk in video overlay passive optical networks. In: *Opt. Fiber Commun. Conf.* (2004)

27.18 G. Kramer, B. Mukherjee, G. Pesavento: IPACT a dynamic protocol for an Ethernet PON (EPON), IEEE Commun. Mag. **40**(2), 74–80 (2002)

27.19 M.P. McGarry, M. Maier, M. Reisslein: Ethernet PONs: Asurvey of dynamic bandwidth allocation (DBA) algorithms, IEEE Commun. Mag. **42**(8), S8–S15 (2004)

27.20 B. Skubic, J. Chen, J. Ahmed, B. Chen, L. Wosinska, B. Mukherjee: Dynamic bandwidth allocation for long-reach PON: Overcoming performance degradation, IEEE Commun. Mag. **48**(11), 100–108 (2010)

27.21 M.-S. Han, H. Yoo, B.-Y. Yoon, B. Kim, J.-S. Koh: Efficient dynamic bandwidth allocation for FSAN-compliant GPON, IEEE J. Opt. Netw. **7**(8), 783–795 (2008)

27.22 I.S. Reed, G. Solomon: Polynomial codes over certain finite fields, J. Soc. Ind. Appl. Math. **8**(2), 300–304 (1960)

27.23 IEEE 802.3av-2009: IEEE standard for information technology – Local and metropolitan area net-

works – Specific requirements – Part 3: CSMA/CD access method and physical layer specifications amendment 1: Physical layer specifications and management parameters for 10 Gb/s passive optical networks (2009)
27.24 ITU-T Recommendation Series G Supplement 39: Optical system design and engineering considerations (2008)
27.25 ISO/IEC 18033-3: Information technology – Security techniques – Encryption algorithms – Part 3: Block ciphers (2010)
27.26 IEEE P1904.1/D3.1: IEEE draft standard for service interoperability in Ethernet passive optical networks (SIEPON) (2016)
27.27 ITU-T Recommendation Series G Supplement 51: Passive optical network protection considerations (2017)
27.28 T. Nishitani, J. Mizoguchi, H. Mukai: Experimental study of Type B protection for a TWDM-PON system, J. Opt. Commun. Netw. **7**(3), A414–A420 (2015)
27.29 ITU-T Recommendation Series G Supplement 54: Ethernet linear protection switching (2015)
27.30 T. Sakamoto: Protection schemes beyond currently defined in FTTx. In: *Natl. Fiber Opt. Eng. Conf.* (2013), https://doi.org/10.1364/NFOEC.2013.NM2I.6
27.31 M.P. Mills: The cloud begins with coal: Big data, big networks, big infrastructure – An overview of the electricity used by the global digital ecosystem, http://www.cepi.org/news/cloud-begins-coal-overview-electricity-used-global-digital-ecosystem (2013)
27.32 GreenTouch: Mission to deliver the architecture, specifications and roadmap to increase network energy efficiency by a factor of 1000 compared to 2010 levels, https://s3-us-west-2.amazonaws.com/belllabs-microsite-greentouch/index.html (2015)
27.33 Recommendation ITU-T G.987.3: 10-Gigabit-capable passive optical networks (XG-PON): Transmission convergence (TC) layer specification (2014)
27.34 Recommendation ITU-T G.989.3: 40-Gigabit-capable passive optical networks (NG-PON2): Transmission convergence layer specification (2015)
27.35 D.A. Khotimsky, D. Zhang, L. Yuan, R.O.C. Hirafuji, D.R. Campelo: Unifying sleep and doze modes for energy-efficient PON systems, IEEE Commun. Lett. **18**(4), 688–691 (2014)
27.36 D. Suvakovic, H. Chow, D. van Veen, J. Galaro, B. Farah, N.P. Anthapadmanabhan, P. Vetter, A. Dupas, R. Boislaigue: Low energy bit-interleaving downstream protocol for passive optical networks. In: *IEEE Online Conf. Green Commun.* (2012) pp. 26–31
27.37 C. Van Praet, H. Chow, D. Suvakovic, D. Van Veen, A. Dupas, R. Boislaigue, R. Farah, M.F. Lau, J. Galaro, G. Qua, N.P. Anthapadmanabhan, G. Torfs, X. Yin, P. Vetter: Demonstration of low-power bit-interleaving TDM PON, Opt. Express **20**, B7–B14 (2012)
27.38 D.T. van Veen, V.E. Houtsma: Proposals for cost-effectively upgrading passive optical networks to a 25G line rate, J. Lightwave Technol. **35**(6), 1180–1187 (2017)
27.39 V. Houtsma, D. van Veen, E. Harstead: Recent progress on standardization of next-generation 25, 50, and 100G EPON, J. Lightwave Technol. **35**(6), 1228–1234 (2017)
27.40 IEEE P802.3ca 50G-EPON Task Force: Physical layer specifications and management parameters for 25 Gb/s and 50 Gb/s passive optical networks, http://www.ieee802.org/3/ca/index.shtml (2020)
27.41 R. Koma, M. Fujiwara, J. Kani, S. Kim, T. Suzuki, K. Suzuki, A. Otaka: Demonstration of real-time burst-mode digital coherent reception with wide dynamic range in DSP-based PON upstream, J. Lightwave Technol. **35**(8), 1392–1398 (2017)
27.42 D. Qian, E. Mateo, M.-F. Huang: A 105 km reach fully passive 10G-PON using a novel digital OLT. In: *Eur. Conf. Exhib. Opt. Commun.* (2012), https://doi.org/10.1364/ECEOC.2012.Tu.1.B.2

Jun-ichi Kani

NTT Access Network Service Systems Laboratories
NTT Corporation
Yokosuka, Japan
junichi.kani.wb@hco.ntt.co.jp

Jun-ichi Kani has been engaged in the research and development of optical communication systems for metro and access applications in the NTT Corporation since 1996. Dr Kani is currently a Senior Distinguished Researcher and heads the Access Systems Technologies Group in NTT Access Network Service Systems Laboratories, Yokosuka, Japan.

Doutje van Veen

Bell Labs Access Lab
Nokia
Murray Hill, NJ, USA
dora.van_veen@nokia-bell-labs.com

Doutje van Veen joined Bell Labs in 2000 and is currently a Distinguished Member of Technical Staff and part of the Fixed Networks leadership team in the Bell Labs Access Lab in Murray Hill, NJ, USA at Nokia. Dr van Veen has been the lead engineer in a team that has developed several early and record-breaking system demonstrations. Her current research is focused on next-generation high bit-rate PON.

28. Emerging PON Technologies

Josep Prat, Luca Valcarenghi

This chapter focuses on recent advances in optical access networks, which are also commonly termed FTTH (fiber-to-the-home) or PON (passive optical networks). The last decade has seen dramatic growth in these networks; they have gone from being almost nonexistent to being a worldwide presence, providing true broadband Internet connections to end users. In recent years, research and development efforts have been directed into enhancing fiber resources and increasing the number of users and the aggregate capacities of PONs. This chapter is divided into two parts. The first (Sects. 28.1–28.3) discusses the second generation of PONs (NG-PON2), which improve upon the first generation by enabling increased bit rates and including a WDM overlay. The second part (Sects. 28.4–28.6) explores proposed longer-term approaches (a future third generation of PONs) with higher scalability and flexibility that have undergone proof-of-concept testing.

Since they were first standardized in the 1990s (B-PON ITU-T G.982, 1996), passive optical networks (PONs) have represented a means of supplying broadband access at low cost. Because of their passive nature and their utilization of optical fibers, they provide a good tradeoff between cost, reach, and capacity. PON speeds have evolved very quickly from the hundreds of megabits per second offered by B-PON to the tens of gigabits per second offered by NG-PON2, increasing in speed by almost one order of magnitude per decade. There are two major ongoing standardization efforts for PONs: one conducted by ITU-T based on the evolution of B-PON, and the other by IEEE in relation to their standard 802.3 (also known as ethernet PONs). In this chapter, we focus mainly on ITU-T PONs, and the latest ITU-T PON standard, NG-PON2, in particular.

Section 28.1 provides an overview of the architecture and the different functions that each NG-PON2 layer provides. NG-PONs utilize multiwavelength transmission in the physical layer, and the advantages of this are highlighted along with the physical aspects that must be considered when attempting to maximize performance. The requirements of tunable transmitters and receivers at the ONUs are discussed in depth. Transmission convergence (i.e., access) layer protocols are reviewed, including specifics relating to the utilization of tunable transmitters and receivers at the ONUs.

Section 28.2 is dedicated to an emerging topic in NG-PON2: energy savings at the optical line termi-

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_28

nation (OLT). While the same techniques for saving energy at the ONUs can be applied in NG-PON2, it is also possible to save energy at the OLT by (de)activating transceivers as a function of the traffic and (de)aggregating ONUs to transceivers.

Finally, Sect. 28.3 focuses on the evolution of PONs towards wavelength division multiplexed (WDM) PONs. In particular, the relationship between NG-PON2, which includes point-to-point transmission, and NG-PON2 is highlighted.

28.1 Hardware, PHY, and MAC of NG-PON2

The International Telecommunication Union's Telecommunication Standardization Sector recently standardized the 40 GB capable Next-Generation PON 2 (NG-PON2) [28.1]. Only a general functional reference architecture is specified in [28.1]. Details about the PHY (physical interface)MAC (medium access control) layers are included in other standards, and some of them are left to vendors. This section provides an overview of the main functionalities implemented in the layered architecture of NG-PON and the hardware utilized in them.

28.1.1 NG-PON2 Architecture

The NG-PON2 architecture is defined in [28.1] and depicted in Fig. 28.1. The ITU-T has directed a great deal of effort into guaranteeing backward compatibility with previous technologies such as XGPON and GPON. Thus, the optical distribution network (ODN) in this architecture is based on a passive splitter.

The most common NG-PON2 architecture is predicted to be the time and wavelength division multiplexed (TWDM) passive optical network (PON), formerly known as the hybrid TDM-WDM PON, as shown in Fig. 28.2. Tunable point-to-point WDM is supported by the NG-PON2 system. The devices utilized are an OLT featuring line cards (LCs), a wavelength multiplexer, a passive splitter, and several ONUs featuring tunable transceivers [28.2]. The line cards, which are also known as optical subscriber units (OSUs), have fixed transceivers that transmit and receive at different wavelengths. This architecture has several advan-

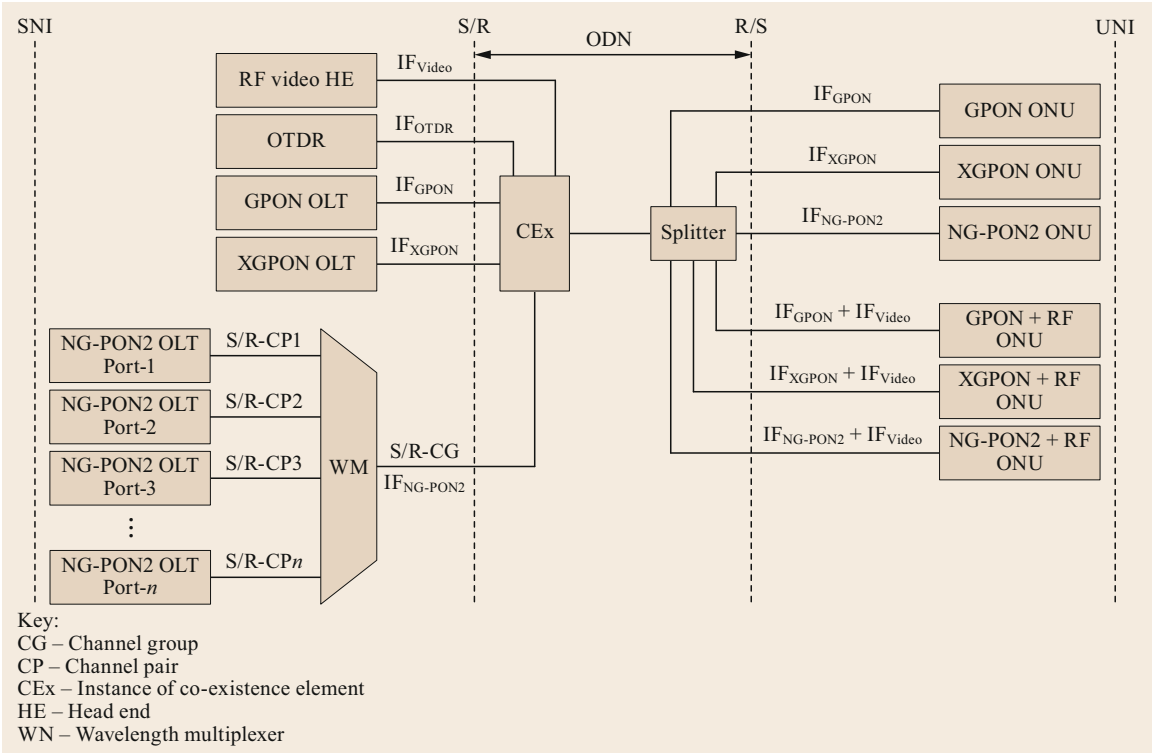


Fig. 28.1 The NG-PON2 architecture as in [28.1]

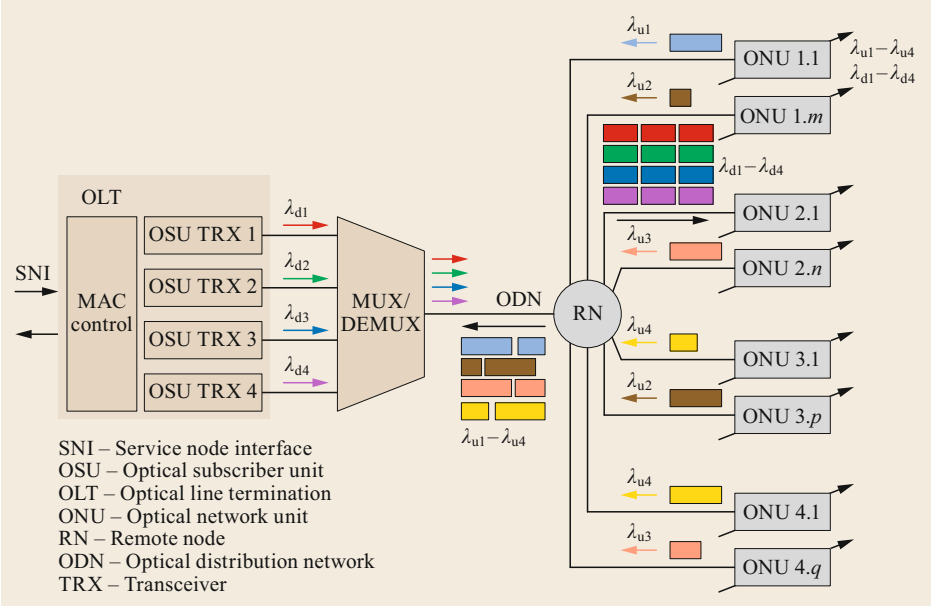


Fig. 28.2 A time and wavelength division multiplexing passive optical network (TWDM PON)

tages. The tunable transceivers of the ONUs allow high flexibility and reconfigurability. Indeed, all of the ONUs are capable of transmitting and receiving at any wavelength, so they can communicate with any OLT transceiver/OSU. Additionally, the utilization of ONUs with tunable transceivers makes them colorless, which facilitates ONU installation and management.

NG-PON2 standardization is based on three standards [28.3, 4]: ITU-T Rec. G.989.1 [28.1], ITU-T G.989.2 [28.5], and ITU-T G.989.3 [28.6]. ITU-T Rec. G.989.1 [28.1] defines the general requirements of 40 Gb capable passive optical network (NG-PON2) systems and their architecture; ITU-T G.989.2 [28.5] deals with the physical media dependent (PMD) layer; and ITU-T Rec. G.989.3, which specifies the TC layer for XGPON, deals with the specific transmission convergence (TC) layer for NG-PON2. This section briefly summarizes ITU-T Rec. G.989.1, while the two subsequent sections outline ITU-T G.989.2 (including the hardware involved) and ITU-T Rec. G.989.3, respectively.

28.1.2 NG-PON2 Physical Layer and Hardware

The physical media dependent (PMD) layer of NG-PON2 is specified in ITU-T G.989.2 [28.5]. This standard establishes the characteristics of hybrid time and wavelength division multiplexing (TWDM) channels, referred to as a TWDM PON. In addition, the characteristics of optional tunable point-to-point wavelength overlay channels, also known as a point-to-point wavelength division multiplexing (PtP WDM) PON, are described.

Among the important information included in the standard document is the wavelength allocation plan for NG-PON2, which is shown in relation to existing legacy PON technologies; see Fig. 28.3. Table 28.1 reports the NG-PON2 allocation plans for both TWDM and PtP WDM. If a particular subset of the spectrum in either band is unused by a TWDM PON and/or legacy systems, a PtP WDM PON is permitted to make use of that particular subband in the upstream and/or downstream direction. However, isolation requirements must be considered. Moreover, when a TWDM PON and a PtP WDM PON are both present, the wavelength channels of both technologies may occupy adjacent wavelength bands. However, the TWDM and PtP WDM channels must not be interleaved.

As reported in [28.7], NG-PON2 systems support a minimum aggregate capacity of 40 Gb/s in the downstream (DS) direction and 10 Gb/s in the upstream (US) direction. From a per-wavelength-channel perspective, TWDM offers three DS/US line rate combinations: the 10/2.5 Gb/s base case and optional symmetric rates of 10/10 and 2.5/2.5 Gb/s. Three line rate classes (around 1, 2.5, and 10 Gb/s) are specified for PtP WDM to transport ethernet, SDH/OTN, and CPRI services. Each NG-PON2 system supports a minimum of 256 addressable ONUs per ODN. Optical parameters are specified assuming four and eight bidirectional wavelength channels for TWDM and PtP WDM, respectively. However, the specification anticipates a future increase in the number of wavelength channels for both technologies.

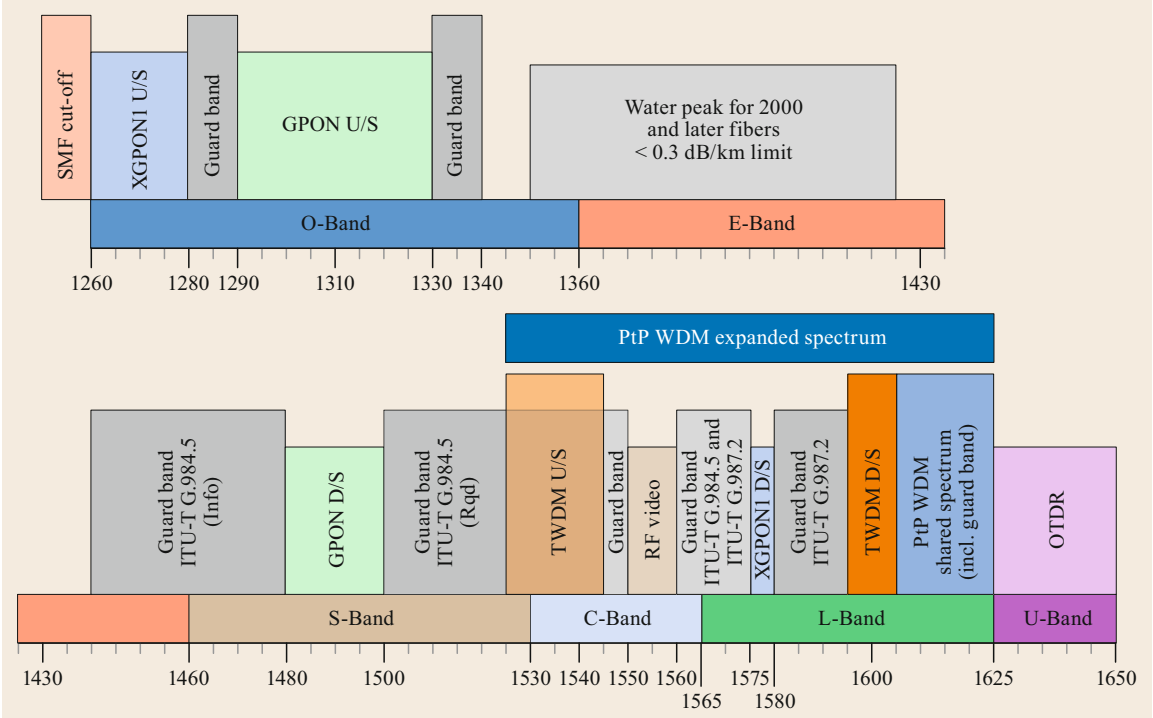


Fig. 28.3 NG-PON2 wavelength plan in relation to coexisting technologies; note that U/S refers to upstream and D/S to downstream. (After [28.5])

Table 28.1 NG-PON2 wavelength allocation plan from [28.5]

TWDM PON		PtP WDM PON
Downstream	**Upstream**	**Upstream/downstream**
1596–1603 nm	Wideband option:1524–1544 nm Reduced band option:1528–1540 nm Narrow band option:1532–1540 nm	Expanded spectrum:1524–1625 nm Shared spectrum:1603–1625 nm

The physical requirements embodied in the NG-PON2 PMD specification (i.e., ITU-T G.989.2) are impacted by optical link design topics, including the optical distribution network characteristics, degradation due to Raman fiber nonlinearity, interchannel crosstalk tolerance, and the wavelength-tuning capabilities of optical network units.

NG-PON2 systems have the ability to support power-split (PS) and wavelength-routed (WR) *optical distribution networks (ODNs)*, as well as hybrids of the two. While support for power-split-only ODNs is mandatory and wavelength-split ODNs are permitted, only ONUs with wavelength selection features are allowed in both cases. Two classes of ODN architectures are considered for PtP WDM PONs. A wavelength-selected ODN (WS-ODN) relies on tunable filters to provide a wavelength selection capability in the ONUs. A wavelength-routed ODN (WR-ODN) possesses an intrinsic wavelength routing capability due to the presence of wavelength splitters in the ODN. The hybrid configuration mixes power splitters, as featured in WS-ODNs, with a cyclic arrayed waveguide grating (CAWG), as featured in WR-ODNs. The ODN type influences the optical path loss budget of the ODN. In general, the use of wavelength splitters rather than power splitters leads to lower insertion losses. ITU-T G.989.2 specifies the ODN optical path loss classes reported in Table 28.2. In addition, minimum and maximum fiber distances are defined, as reported in Table 28.3. To learn more about ODNs, the reader is referred to [28.7].

One of the most important features of NG-PON2 is the ability of the ONUs to tune their upstream/downstream transmission/reception wavelength. This feature paves the way to bandwidth allocation algorithms that work in both the time and the wavelength domains and can better utilize the network capacity, but whose effectiveness depends on the physical characteristics of

Table 28.2 ODN optical path loss classes (ODN classes)

	Class N1	Class N2	Class E1	Class E2
Minimum optical path loss (dB)	14	16	18	20
Maximum optical path loss (dB)	29	31	33	35
Maximum differential optical path loss (dB)	15			

Table 28.3 ODN fiber distance classes

Fiber distance class	Minimum fiber distance (km)	Maximum fiber distance (km)
DD20	0	20
DD40	0	40

Table 28.4 NG-PON2 ONU tuning time classes

Tuning time class	Tuning time
Class 1	$< 10\,\mu s$
Class 2	$10\,\mu s$ to 25 ms
Class 3	25 ms to 1 s

the devices. In particular, the ONU transmitter/receiver tuning time is the main parameter. In ITU-T G.989.2, the tuning time is defined as:

> ... the elapsed time from the moment the tunable device leaves the source wavelength channel to the moment the tunable device reaches the target wavelength channel ...

ITU-T G.989.2 specifies the three tuning time classes reported in Table 28.4. In [28.7], it is reported that the slowest (class 3) tunable devices may utilize thermal effects to change their operating wavelengths, and are best suited to applications in which tuning operations are performed infrequently or when a short service interruption is tolerable. Semi-static load-sharing and power-saving mechanisms based on channel changes are supported by this class of devices. Class 2 tunable devices feature faster channel tuning, meaning that sub-50 ms channel protection schemes can be applied. They also enable the implementation of dynamic load-sharing and dynamic power-saving schemes. Class 1 tunable devices, which have the shortest tuning times, may facilitate dynamic wavelength and bandwidth allocation in the system.

Another important parameter to consider in an NG-PON2 design is the optical path penalty (OPP). In ITU-T G.989.2, the OPP is defined as the apparent degradation of receiver sensitivity due to impairments during transmission as well as the apparent increase in ODN loss due to Raman depletion.

The optical link budget for a particular ODN optical loss class is only satisfied if the difference between the minimum mean launched optical power of the transmitter (in dB relative to 1 mW) and the receiver sensitivity (in dB relative to 1 mW) is equal to or exceeds the sum of the maximum OPP (in dB) and the maximum optical path loss (in dB) specified for the given ODN optical loss class.

The OPP accounts for the effects of reflections, intersymbol interference, mode partition noise, fiber dispersion, and fiber nonlinearities. Among these, we consider chromatic dispersion and Raman nonlinearity caused by high optical power and multiwavelength operation in this chapter. Chromatic dispersion is the phenomenon in which the phase velocity of a wave depends on its frequency. Raman nonlinearity can result in nonlinear crosstalk and signal depletion for certain wavelengths. Chromatic dispersion occurs due to the spectrum of the lasers utilized for transmission. In NG-PONs, both directly modulated DFB lasers (DML) and externally modulated lasers (EML) are considered. ITU-T G.989.2 defines maximum OPP values in dB for both downstream (i.e., at the ONU) and upstream (i.e., at the OLT) reception, assuming the use of EMLs. These maximum OPP values are reported in Table 28.5.

Electronic dispersion compensation (EDC) can be used in the OLT (ONU) transmitter to achieve the OPP specified in optical interface parameter tables. EDC may also be used at the ONU (OLT) receiver.

As stated in [28.7], Raman nonlinearity in the fiber of a NG-PON2 system has two distinct effects: first, US TWDM channels are depleted by the counterpropagating DS TWDM channels; second, modulation crosstalk between any copropagating optical signals separated in optical frequency by 1–40 THz. The power depletion is accounted for by recommending an increase in OPP in the NG-PON2 PMD layer. Results presented in [28.7] show the worst-case Raman depletion of US TWDM channels for all optical path loss (OPL or ODN) classes and both 4- and 8-channel cases as a function of fiber distance.

Other types of impairments that need to be taken into account in TWDM-PONs include crosstalk due to imperfect laser spectra (side-mode suppression), high dynamic power ranges (up to almost 30 dB in an 8-channel system), and burst-mode-induced laser wavelength drift (tens of GHz in the worst case).

As reported in [28.7], the optical crosstalk terms for the NG-PON2 PMD layer relate to the in-band interferometric crosstalk generated when optical power wanders into the spectral regions of other channels. In

Table 28.5 Maximum OPP values for chromatic dispersion

Rate (Gb/s)	Direction	Interface	Maximum OPP (dB)				Notes
2.48832	Downstream	ONU receiver (optical interface R)	1.0				
9.95328	Downstream	ONU receiver (optical interface R)	2.0				
2.48832	Upstream	OLT receiver (optical interface R)	N1	N2	E1	E2	
			1.0	1.0	1.5	1.5	With Raman effects (DD20, four channels)
			1.5	1.5	2.0	2.0	With Raman effects (DD40, four channels)
			1.0	1.5	2.0	2.0	With Raman effects (DD20, eight channels)
			2.0	2.5	3.5	3.5	With Raman effects (DD40, eight channels)
9.95328	Upstream	OLT receiver (optical interface R)	N1	N2	E1	E2	
			1.5	1.5	2.0	2.0	With Raman effects (DD20, four channels)
			FFS	FFS	FFS	FFS	With Raman effects (DD40, four channels)
			FFS	FFS	FFS	FFS	With Raman effects (DD20, eight channels)
			FFS	FFS	FFS	FFS	With Raman effects (DD40, eight channels)

general, the most stringent requirements for the spectral characteristics of a TX (transmitter) are imposed in the US, mainly because of the large power level difference at the OLT RX (receiver) between signals from different ONUs. In [28.7], it is stated that two crosstalk terms are important when characterizing an enabled channel: the interference caused by similar signals, i.e., TWDM-to-TWDM and PtP WDM-to-PtP WDM, and the interference caused by dissimilar signals, i.e., TWDM-to-PtP WDM and PtP WDM-to-TWDM crosstalk. In ITU-T G.989.2, two terms are utilized to describe the aforementioned cases: out-of-channel power spectral density and out-of-band power spectral density.

The out-of-channel optical power spectral density (OOC-PSD) defines the maximum power spectral density a NG-PON2 TX is permitted to emit outside the spectral interval corresponding to its current operating wavelength channel. This PSD is specified at the appropriate reference point (S/R-CG for the DS direction, R/S for the US direction). In the NG-PON2, a penalty of 1 dB due to OOC interference in the US direction is accounted for. The corresponding value in the DS direction is as low as 0.1 dB because of the bandpass filter function of the wavelength multiplexer (WM).

The out-of-band optical power spectral density (OOB-PSD) defines the maximum power spectral density a NG-PON2 TX is permitted to emit outside the specified operating wavelength band. In both transmission directions, a penalty of 0.1 dB due to OOB interference is accounted for in the NG-PON2 standard.

The last part of this section focuses on two of the most important devices in a TWDM-PON: tunable transmitters and tunable receivers. As stated in [28.8], the expected primary specifications for tunable transmitters (T-TX) in ONUs are 4-channel tunability with an assumed 100 GHz channel spacing in the C band, a line rate of 2.5 (most likely for residential services) or 10 Gb/s (most likely for business and mobile services), and burst-mode operation. Table 28.6, originally reported in [28.8], summarizes different types of tunable transmitters.

There are three different approaches to making transmitters tunable: electronic, thermal, and mechanical. The most prominent electronically tuned lasers are various implementations of the DBR laser. Among mechanically tunable lasers, there has been particular progress in the development of microelectromechanical vertical cavity surface-emitting lasers [28.9, 10]. The narrow-band (approx. 5 nm) tunable DFB laser is a notable type of thermally tunable laser. So far, however, only lasers based on electronic tuning have achieved significant commercial success.

Table 28.6 Possible tunable transmitter categories (from [28.8])

Type	A	A′	B	C
Device	Heater-integrated DFB-LD	EML (TEC inside)	DBR-LD (short cavity type)	4-ch EML array + selector SW
Tuning control	Thermal	Thermal	Electrical	Electrical (SW)
Tuning time	Subsecond order	Subsecond order	< 5 ns (LD chip level)	< 100 ns (burst-response time included)
Modulation	Direct	External	Direct	External
DCT[a] for 10 Gb/s operation	Needed	Not needed	Needed	Not needed

[a] Dispersion compensation technique

Table 28.7 Possible tunable receiver technologies (as reported in [28.8])

Type	X	Y	Z
Device	Heater-integrated TFF + PD	DEMUX + 4-ch APD array + selector SW	DEMUX + SOA gates + PD
Tuning control	Thermal	Electrical (electric SW)	Optical selector SW
Tuning time	Seconds	< 20 ns	< 1.5 ns (chip level)

Type A is a conventional uncooled direct modulation laser. The advantages of this laser are its small size, low cost, and easy burst-mode operation. Its main drawbacks are the need for an integrated heater on a chip for tunability and its slow (subsecond) tuning time, as the wavelength is tuned by changing the chip temperature using the heater. Another disadvantage of type A is that dispersion compensation techniques (DCTs) are needed for 10 Gb/s direct modulation operation in the C band, such as a burst-mode electric dispersion circuit (EDC) in the OLT and pre-emphasis in the ONU.

Type A′ is an electroabsorption modulator integrated distributed-feedback laser diode (EML), which has similar characteristics to type A because there is generally a thermoelectric cooler (TEC) inside the module. However, it does not need DCTs for 10 Gb/s operation as it has an integrated modulator.

Type B is a distributed Bragg reflector (DBR)-LD. Its advantages are its small size and fast tuning time, on the order of nanoseconds (chip level); however, just like type A, it needs DCTs for 10 Gb/s operation in the C-band.

Type C consists of a 4-channel EML arrayed device, an optical multiplexer, and an electric wavelength channel selector SW. Its main advantages are its short tuning time, elimination of mode hopping, and that it does not require DCTs, for the same reason as for type A′. Its main drawback is its large footprint. However, this issue can be solved through small scale factor integration.

Turning our attention to tunable receivers, the following is stated in [28.8]:

> … the expected primary specifications for tunable receivers (TRX) in the ONUs are 4-channel tunability with an assumed 100 GHz channel spacing in the L-band and a line rate of 2.5 or 10 Gb/s.

Because TRX technologies are less mature than T-TX technologies, TRX represent a major challenge at the device level.

Table 28.7 summarizes possible tunable receiver technologies (as originally reported in [28.8]). Type X is based on a thin-film filter (TFF) with integrated heater and a p-intrinsic-n (p-i-n) PD. Type X comes in a small package (e.g., TO-Can) and is inexpensive. It is therefore an attractive solution, but its tuning time is limited to the subsecond range due to the heater control.

Type Y consists of a 4-channel APD-arrayed device, an optical demultiplexer (DEMUX), and a switch (SW). Type Y has a short tuning time of < 20 ns [28.8]. Finally, type Z is based on a DEMUX, semiconductor optical amplifier (SOA) gates, and a PD. Wavelength selection is performed by the SOA gates, leading to a very fast tuning time of 1.5 ns. The difference between type Y and type Z is that wavelength selection is performed at the electric level in type Y and at the optical level in type Z.

28.1.3 NG-PON2 MAC

The NG-PON2 system contains two types of wavelength channels: TWDM channels and PtP WDM channels [28.11]. TWDM channels provide point-to-multipoint connectivity using conventional TDM/TDMA PON mechanisms. PtP WDM channels provide point-to-point (PtP) connectivity using an externally specified synchronous or asynchronous mechanism. The NG-PON TC layer specifications for the two types of wavelength channels differ in some respects. For TWDM channels, the NG-PON TC layer specification includes a full protocol stack. For PtP WDM channels, the specification relies on framing, encapsulation, and management, which are specified for the underlying point-to-point client. Note that the description of TWDM management channels is extensively based on the XG-PON TC layer specification [28.12].

As stated in [28.11], the NG-PON2 TC layer specification relies on interoperability between the OLT channel terminations (OLT CT) within the NG-PON2 system to enable multiwavelength operations and wavelength channel mobility. The inter-channel termination protocol (ICTP) helps to achieve this interoperability. At any point during the ONU's activation cycle, it can be instructed to execute a scheduled handover from one TWDM channel (source) to another TWDM channel (target). The OLT CTs involved in the handover operation coordinate the effort via the ICTP. Additionally, the ICTP makes use of the primitive set outlined in [28.6] and is specified normatively by the Broadband Forum (BBF) in WT-352.

A new concept, the auxiliary management and control channel (AMCC), is proposed in the NG-PON2 TC layer specification. The AMCC is a low-rate channel that allows the OLT CT and an ONU to exchange infor-

mation without allocating a separate wavelength carrier while avoiding interference with in-band data communication.

The NG-PON2 medium access control, which is specified in [28.6], is known as the transmission convergence (TC) layer. As stated in [28.6]:

> ... The transmission convergence (TC) layer is the protocol layer of the NG-PON2 system that is positioned between the physical media dependent (PMD) layer and service clients. It builds on the recommendation ITU-T G.987.3, with modifications for NG-PON2 specific features ...

ITU-T G.989.3 deals not only with *medium access control* (i.e., upstream time-division multiple access and dynamic bandwidth assignment mechanisms) but also with *service adaptation* (e.g., the (XG-PON) encapsulation method (XGEM)), *framing* (i.e., downstream frame and upstream burst format specification), the physical interface (PHY) adaptation sublayer (e.g., synchronization, forward error correction, and scrambling), physical layer operation (e.g., the administration and management (PLOAM) messaging channel), the optical network unit (ONU) activation cycle state machine, performance monitoring, protection, signaling mechanisms and protocols to support ONU power management, and channel management. Similar functions, when applicable, are also defined for PtP operations.

The TWDM TC layer is composed of three sublayers that implement some of the aforementioned functions: the TWDM TC *service adaptation sublayer*, the TWDM TC *framing sublayer*, and the TWDM TC PHY *adaptation sublayer*. These sublayers are shown in Figs. 28.4 and 28.5, which are taken from [28.6]. In the downstream direction, the interface between the TWDM TC layer and the PMD layer is represented by a continuous bitstream at the nominal line rate, which is partitioned into 125 μs frames. In the upstream direction, the interface between the TWDM TC layer and the physical medium dependent (PMD) layer is represented by a sequence of precisely timed bursts. In particular, the relationship between the PHY frame and PHY burst in the US direction is depicted in Fig. 28.6.

As stated in [28.6], the duration of an upstream PHY frame is 125 μs, which corresponds to 38 880 bytes (9720 words) at an upstream rate of 2.48832 Gb/s and to 155 520 bytes (38 880 words) at an upstream rate of 9.95328 Gb/s. Each ONU determines the time at which a particular upstream PHY frame is initiated by appropriately offsetting the starting point of the respective downstream PHY frame. The sequence of upstream PHY frame boundary points provides a common timing reference shared by the OLT CT and all the ONUs on the PON, but those points do not correspond to any specific event (unlike the downstream PHY frame boundary points, where the transmission or receipt of a PSBd starts). In the upstream direction, each ONU transmits a series of relatively short PHY bursts and remains silent, disabling the transmitter, in between the bursts. An upstream PHY burst consists of an upstream physical synchronization block (PSBu) and a PHY burst payload represented by the upstream FS burst, whose content may be protected by FEC and is scrambled. The OLT CT uses the BWmap to control the timing and duration of the upstream PHY bursts in order to ensure that upstream transmissions by different ONUs do not overlap. The upstream PHY bursts of each ONU are referenced to the start of the appropriate upstream PHY frame. An upstream PHY burst belongs to upstream PHY frame N as long as this burst is specified in the BWmap transmitted with downstream PHY frame N. If this is the case, the first byte of the FS burst header is transmitted within the boundaries of PHY frame N. The PSBu portion of an upstream PHY burst may be transmitted within the boundaries of the previous PHY frame. An upstream PHY burst belonging to a particular upstream PHY frame may extend beyond the trailing boundary of that frame. To prevent upstream transmissions from colliding with and jamming each other, a guard time between upstream bursts from different ONUs is specified by the OLT. The guard time accommodates the TX enable and disable times and includes a margin for individual ONU transmission drift. The recommended minimum guard time is 64 bit.

The service adaptation sublayer performs TC layer service data unit (SDU) (i.e., a user data frame such as an ethernet frame or MPLS packet, or a protocol message of the ONU management and control interface) encapsulation using the XG-PON encapsulation method (XGEM) as specified in G.987.3. The protocol data unit (PDU) of the TWDM TC service adaptation sublayer is represented by an XGEM frame, which consists of an XGEM header and XGEM payload. The XGEM header is an 8 byte protected field that identifies a logical connection or traffic flow to which the encapsulated SDU belongs and other functions relating to PDU composition. The XGEM payload may contain a complete TC layer SDU or, if necessary, a fragment of a SDU. An example of a TCP segment encapsulated into an XGEM PDU is reported in [28.11].

For the *framing and PHY adaptation* sublayer, the SDU of the *framing sublayer* (*FS*) is represented by a sequence of one or more XGEM frames that form the FS payload. The PDU of the framing sublayer or, equivalently, the SDU of the PHY adaptation sublayer is represented by the FS frame and the FS burst for the downstream and upstream directions, respectively.

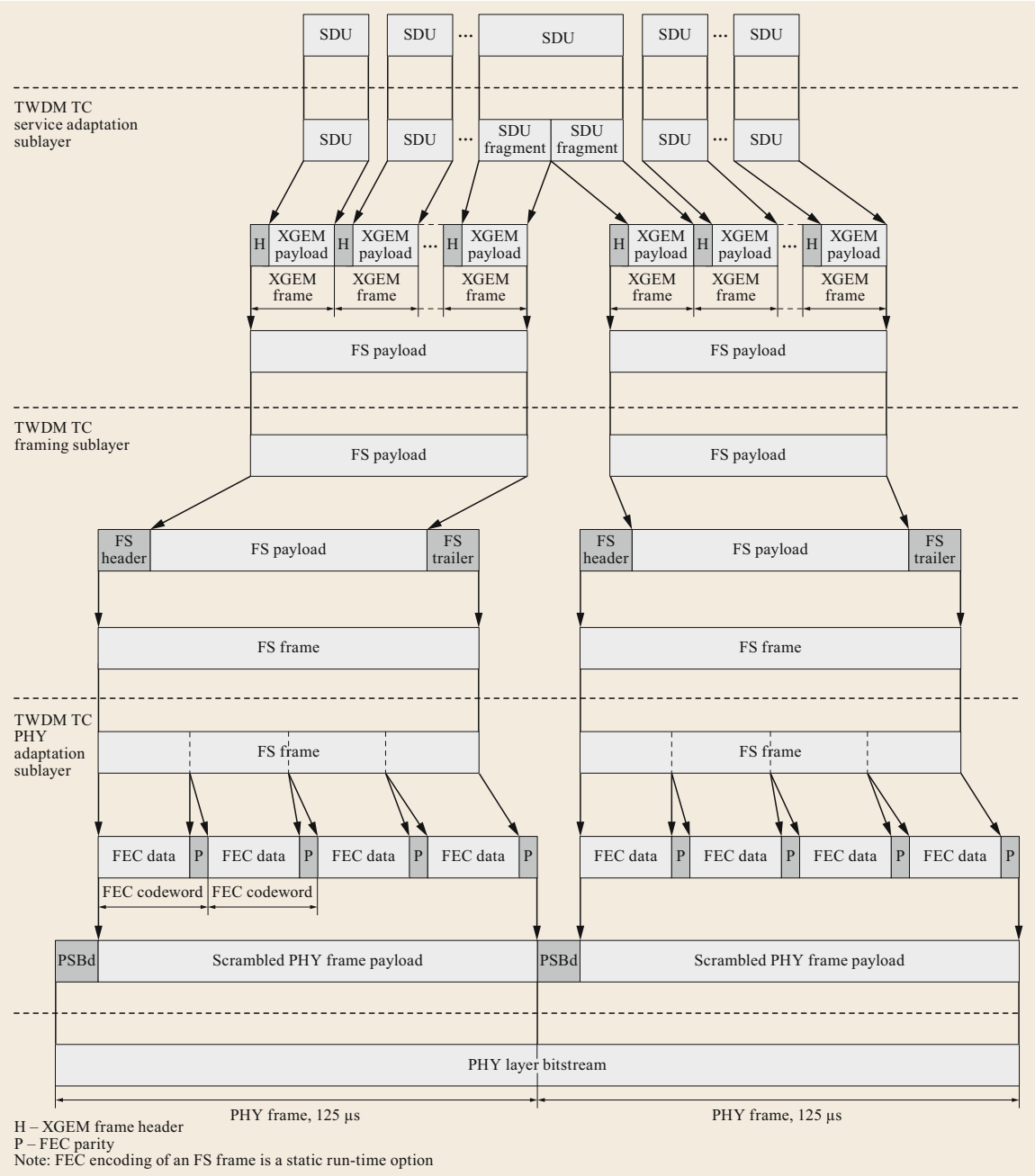


Fig. 28.4 Downstream SDU mapping into PHY frames (after [28.6])

The structure of the FS as well as the PDU of the *PHY adaptation sublayer* are reported in [28.11]. It should be noted that the size of the PHY frame payload is fixed for a given downstream line rate (155 496 octets for 9.95328 Gb/s or 38 856 octets for 2.48832 Gb/s), so the size of the downstream FS frame is fixed for a given line rate and FEC status.

Upstream bandwidth assignment in TWDM PON systems is based on BWmap partition. The BWmap is transmitted every 125 µs and contains a sequence of 8-octet allocation structures (AS), each specifying a transmission grant to a particular ONU. The format of an AS in a TWDM PON is identical to that in a G.987 XG-PON. The grant recipient within an ONU is

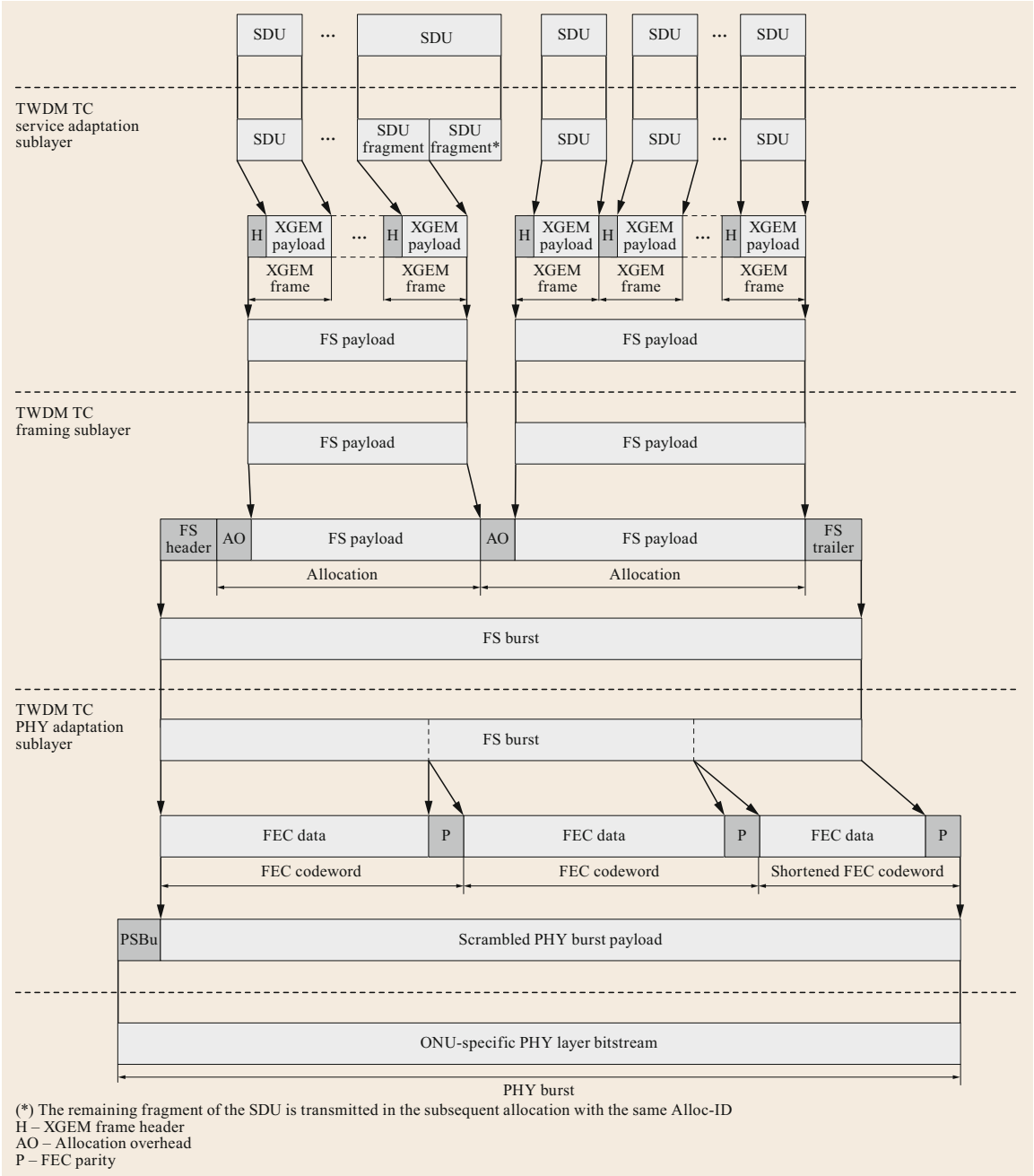


Fig. 28.5 Upstream SDU mapping into PHY frames (after [28.6])

identified by the Alloc-ID. The grant recipient can be either a managed transmission container (T-CONT) associated with a specific group of user data traffic flows or an internal structure that is created automatically and is intended for the management of traffic sourced by the ONU itself. Within an AS, a 1 bit forced wakeup indication (FWI) flag is employed by the ONU power management control mechanism to implement energy-efficient scheme operations.

The auxiliary management and control channel (AMCC) for the PtP WDM channels is specified in ITU-T G.989.3. The PtP AMCC TC layer protocol stack reuses the TWDM TC specification and is composed of three sublayers: the service adaptation sub-

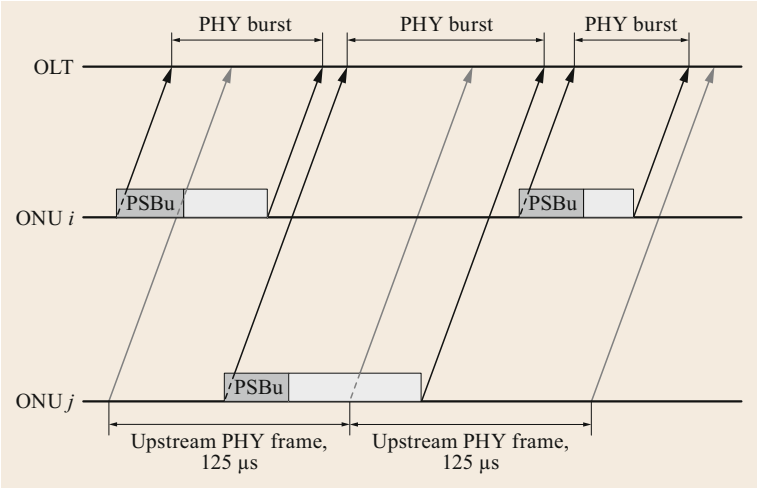


Fig. 28.6 Relationship between PHY frames and PHY bursts in the upstream direction

layer, the framing sublayer, and the PHY adaptation sublayer. The SDU of the PtP AMCC TC FS is represented by a sequence of one or more XGEM frames. The PtP AMCC FS frame format and its transformation into the PHY adaptation sublayer PDU are invariant with respect to the direction of transmission, following the pattern laid down for the TWDM TC layer in the downstream direction.

For TWDM channels, the OLT CT and ONU are required to support forward error correction (FEC) in both the downstream and the upstream directions, but FEC is subject to ON/OFF control. For PtP WDM channels, when the AMCC operates in transcoded mode, FEC encoding is an integral part of transcoder operation. The use of FEC when the AMCC operates in transparent mode requires further study [28.11].

NG-PON2 management is based on the physical layer operation, administration and maintenance (PLOAM) messaging channel. The NG-PON2 PLOAM specification extends that specified in G.987.3 for the XG-PON. In particular, the NG-PON2 PLOAM channel supports legacy functions such as burst profile announcement, ONU activation, ranging, registration, data encryption key exchange, and ONU power management. In addition, G.989.3 introduces new PLOAM functions: system and channel profile announcement, ONU wavelength channel handover signaling, wavelength adjustment and calibration, protection configuration, optical power leveling, and rate control. All of the aforementioned messages facilitate the advanced NG- PON2 features detailed in the following section.

28.2 Energy Efficiency in TWDM PONs

In TWDM-PONs, energy can be saved by dynamically reconfiguring the number of active optical subscriber units (OSUs) (i.e., OLT transceivers, OLT channel termination (OLT CT) in [28.6]) based on the network load [28.13–15]. However, the energy savings that can be achieved without severely impacting network performance (e.g., the average delay) are limited by many factors: the traffic threshold at which OSUs are de/activated, the optimality of the dynamic wavelength and bandwidth allocation (DWBA) [28.16], the ONU transceiver tuning time [28.17, 18], and the OSU turn-on time. Some solutions to the aforementioned problems have already been proposed. An automatic load balancing dynamic wavelength and bandwidth allocation (DWBA) algorithm was proposed in [28.19] as a means to reduce the impact of the ONU transceiver tuning time. In [28.20], the first demonstration of a hitless λ-tuning sequence was published for a scenario with an OLT equipped with four OSUs and two ONUs. In addition, solutions for reducing DWBA tuning-time overhead penalties were recently proposed. In [28.21], a DWBA is proposed that minimizes the number of wavelength reallocations. This, in turn, minimizes the tuning-time overhead penalties while facilitating the fair distribution of traffic to OSUs. OSUs and ONUs may implement a synchronous cyclic sleep (SCS) mode as an alternative to DWBA. In this mode, the OSU sleeps for a cycle once all transmissions to and by the

ONUs assigned to that OSU have been performed. At the same time, each ONU implements cyclic sleep outside its assigned slot.

The following section discusses these issues further and delineates possible solutions. Some of the proposed solutions are also evaluated using a time-driven simulator.

28.2.1 Factors That Can Affect Reconfigurable TWDM PON Performance

This section overviews how some factors can impact reconfigurable TWDM performance in terms of both energy efficiency and average frame delay. Figure 28.7 summarizes the factors considered: whether and when to trigger reconfiguration (i.e., reconfiguration triggering), deciding which OSUs to (de)activate (i.e., OSU selection), the assignment of ONUs to OSUs, and ONU transmission/reception scheduling at each wavelength (i.e., wavelength and bandwidth allocation). The average frame delay is defined as the time that elapses between the arrival of a frame at the sender (either an ONU or an OSU) and its receipt at the receiver, averaged for all of the frames that were successfully transmitted between ONUs and OSUs and vice versa. The average energy efficiency (a percentage) is the average energy saved at the OLT when OSUs are dynamically turned ON and OFF upon network reconfiguration as compared to when the OSUs are always ON.

Reconfiguration Triggering

Reconfiguration triggering is impacted by the load threshold at which OSUs are (de)activated and by the reconfiguration triggering time:

1. *Load threshold at which OSUs are (de)activated.* Careful selection of this threshold and, in turn, the number of OSUs (i.e., wavelength pairs) to keep active can lead to more energy-efficient TWDM-PON performance. Situations may arise where there is a high load on the active OSUs after reconfiguration, causing delay peaks [28.14, 22, 23].
2. *Reconfiguration triggering time.* The reconfiguration triggering time is the time that elapses between the point at which the network overcomes the load threshold that necessitates reconfiguration and the point at which reconfiguration procedures start. Following load variation and after checking whether network reconfiguration is a possibility (e.g., choosing to turn OFF some OSUs), reconfiguration may be triggered after a certain delay (e.g., to avoid reconfiguring the network too often when there are brief peaks in traffic). However, this delay between the decision to reconfigure the network and the moment that the reconfiguration is triggered may impact network performance. This is especially true when the ONU tuning time is null.

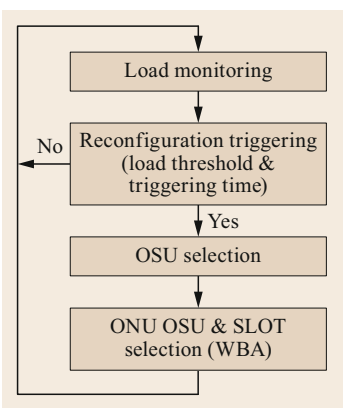


Fig. 28.7 Energy-efficient TWDM-PON decision flow

OSU Selection and DWBA Optimality

Selecting the OSUs to be kept active and those to be switched into sleep mode (i.e., *OSU selection*) can also impact network performance. If the number of ONUs assigned to an OSU before reconfiguration is not considered in the decision, a large number of ONUs might need to be tuned, potentially increasing the average frame delay. DWBA optimality might also impact reconfigurable TWDM PON performance. The problem of optimally assigning ONUs to wavelengths and slots can be reduced to a bin-packing optimization problem, which is NP-hard, as stated in [28.24]. However, the problem formulation should also consider the architectural implementation of the TWDM-PON. For example, ONUs featuring a single transceiver cannot transmit/receive simultaneously at two different wavelengths. Thus, even when the traffic generated by the ONUs is greater than the capacity of a single wavelength, only one OSU can be exploited. Other parameters to be considered are the ONU tuning time, the OSU turn-on time, and the RTT. Last but not least, DWBA suboptimality can be caused by the discrete nature of the number of wavelengths together with the constraint that the ONU cannot transmit/receive simultaneously at two different wavelength pairs. Consider the following example. Three wavelengths are available and only two ONUs are present. In this case, even if the ONUs generate traffic that uses all of the capacity provided by the three wavelengths, each ONU can only exploit the capacity of a single wavelength at most. More generally, therefore, an ONU can exploit the capacity of one wavelength at most (i.e., 10 Gb/s for NG-PON2). This situation is true irrespective of the traffic the ONU generates.

ONU Transceiver Tuning Time

The factor that probably affects the DWBA performance the most is the ONU transceiver tuning time [28.18]. In particular, while the tuning time of the tunable transmitter can be significantly reduced, the tunable receiver tuning time remains a problem. The ONU tunable receiver architectures proposed in [28.17] have the potential to achieve rapid tuning. In them, all transmitted wavelengths are received by an optical demultiplexer and then photodiodes. The desired wavelength is selected electronically. Although this solution leads to a very short tuning time, it implies considerable hardware and energy expenses. In addition, the ONU tuning process may impact not only the performance of the tuning ONU but also the transmissions from other ONUs. Assume that an ONU finishes its tuning process before the time slot assigned to it. In this case, if the laser is active, the ONU can disturb the transmission of another ONU utilizing the same wavelength. Thus, specific architectural countermeasures for the tuning ONU need to be adopted, such as only enabling laser transmission at the beginning of the assigned slot. Finally, periodically tuning an ONU between two wavelengths may cause unexpected delays. Indeed, when some ONUs tune to a wavelength already utilized by other ONUs, the tuning process can exploit the transmission of the latter set of ONUs in TDM fashion. On the other hand, when the ONUs tune to a wavelength that is not currently being utilized, the tuning may not be performed in parallel with another ONU transmission, so an additional delay is introduced.

OSU Turn-On Time

Another factor that increases not only the average frame delay but also the energy consumption is the OSU turn-on time. Even though this factor is often neglected, OSUs may require some time to wake up after a sleep period, just as ONUs do. Therefore, energy is consumed during this phase as well.

28.2.2 Energy-Efficient TWDM PON Performance Evaluation

In this section, various results from energy-efficient TWDM PON performance evaluations in relation to the aforementioned parameters are presented.

Optimizing Reconfiguration Triggering in TWDM PONs with a Fluctuating Load

In [28.25], it is shown that there is an optimal time to trigger network reconfiguration following load variations in TWDM-PONs. Simulations and experiments show that it is better to reconfigure the network as soon as the traffic increases. However, the preferred option if the ONU tuning time is non-negligible and there are rapid traffic variations is to avoid reconfiguring the network, even if this results in increased energy consumption.

The TWDM-PON system model depicted in Fig. 28.2 is considered when conducting a performance evaluation. This model has the following characteristics:

- The EPON multi-point control protocol (MPCP) defined in IEEE standard 802.3-2012 is the network control protocol. This protocol, depicted in Fig. 28.8, also includes the extensions proposed in [28.26] to include the wavelength pairs through which ONUs and OSUs communicate in both GATE and REPORT messages.
- The time is assumed to be slotted, as this increases the simplicity without losing generality.
- The upstream (US) and the downstream (DS) transmission directions are locked.

The decision flow is depicted in Fig. 28.7. The following decisions are taken by the OLT:

- *Load monitoring*. The OLT monitors the average network load periodically (period: T_{var}). The average network load is defined as the ratio of the sum of the average frame arrival probability to the number of OSUs.
- *Reconfiguration triggering*. The monitored load is utilized by the reconfiguration triggering method to decide whether reconfiguration must be initiated. The policy for deciding how many OSUs are ON is as follows. If $\rho < 0.25$, only one of the four OSUs available is activated. If $0.25 \leq \rho < 0.5$, two OSUs are activated. If $0.5 \leq \rho < 0.75$, three OSUs are activated. If $\rho \geq 0.75$, all four OSUs are activated.
- *OSU selection*. An offline WBA is performed to assign ONUs to OSUs and slots. OSUs are activated in order of increasing OSU ID and deactivated in order of decreasing OSU ID. The WBA first assigns the slots in a cycle and then the wavelength pairs to the ONUs. Irrespective of the frames queued, one slot is assigned to each ONU. The remaining slots that belong to the active OSUs are then assigned to the ONUs in proportion to the number of queued frames, utilizing a method based on the largest remainder method (also known as the Hare–Niemeyer method).

Finally, the OLT sends GATE messages to the ONUs. These messages carry the wavelength(s) and the slot(s) that the ONUs are assigned to in the next cycle. Upon receipt of the GATE messages, each ONU tunes to

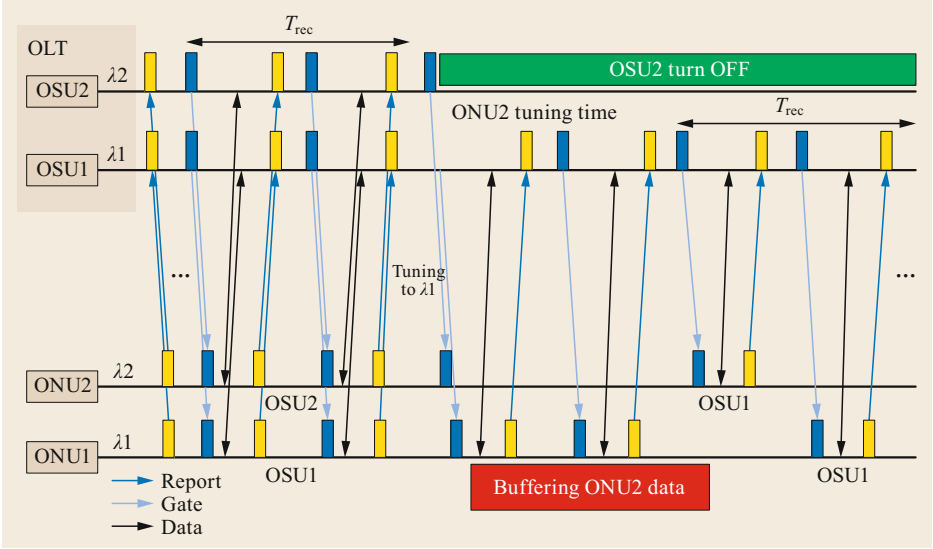


Fig. 28.8 MPC protocol time chart. © IEEE 2016. Reprinted, with permission, from [28.25]

the wavelength of its first transmission slot if needed. ONUs that do not tune (e.g., ONU1) keep transmitting while the other ONUs tune. In addition, as shown in Fig. 28.8, OSUs are turned OFF during the ONU tuning process; the OSUs can be OFF for the ONU tuning time (T_t) if they do not need to serve other ONUs.

A first performance evaluation is carried out using a custom-built time-driven simulator. The simulations involve four OSUs (i.e., four wavelength pairs), four ONUs, a fixed cycle time T_c, and the duration of four slots T_s. Initially, each ONU generates frames based on a Bernoulli process with the same probability p for each ONU. The frame transmission time is assumed to be one slot. Then, every T_{var}, $\Delta p_i \in [-\Delta p, \Delta p]$ is computed for each ONU i, and this value is added to p to simulate traffic variation (note that, in all cases, $p_i = p + \Delta p_i \in [0, 1]$). As previously described, the average network load ρ is computed every T_{rec} based on the received REPORTs.

The performance parameters of interest are the average frame delay and the average OLT energy savings. The average frame delay is defined as the time between the arrival of a frame at the sender (either an ONU or an OSU) and its receipt by the receiver averaged across all the frames successfully transmitted between ONUs and OSUs and vice versa. The average energy efficiency (a percentage) is the average energy saved at the OLT when OSUs are dynamically turned ON and OFF during network reconfiguration with respect to the energy used when the OSUs are always ON. The power used by the OSU during ON and OFF times is 6 and 1 W, respectively.

Fig. 28.9 shows that, in general, if reconfigurations are delayed with respect to load variations (i.e., T_{rec} is large), the average frame delay increases. However, for

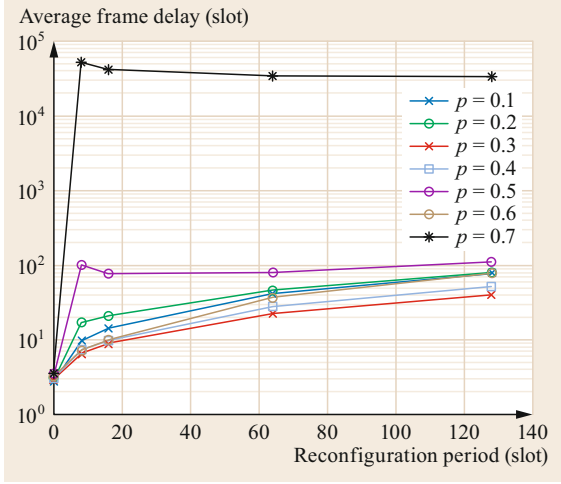


Fig. 28.9 Average frame delay versus reconfiguration period T_{rec} and initial frame arrival probability p, with $T_t = T_s$, $T_{var} = 4T_s$, and $\Delta p = 0.3$. © IEEE 2016. Reprinted, with permission, from [28.25]

some average network loads close to the reconfiguration triggering threshold (i.e., $p = 0.7$ or $p = 0.5$), the average frame delay presents peaks for small values of T_{rec}, even if the tuning time overhead (e.g., $T_t = 1$) is short. For example, for $p = 0.7$ and $p = 0.5$, the maximum average frame delay is reached for $T_{rec} = 8$ slots. Reconfigurations will therefore be triggered shortly after load variations. However, if T_{rec} is short, the tuning-time overhead penalizes frequent reconfigurations. For $p = 0.5$, the value of T_{rec} that minimizes the average frame delay is $T_{rec} = 16$ slots.

As depicted in Fig. 28.10, experiments are run by emulating the behavior of one OLT consisting of two

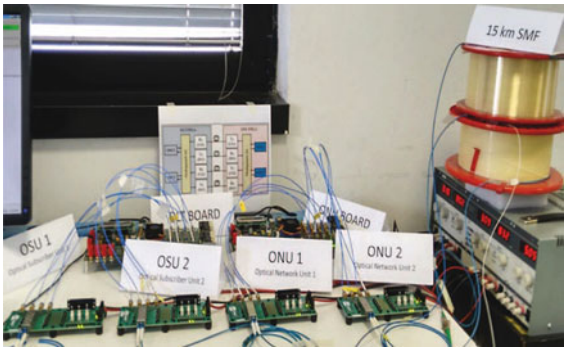


Fig. 28.10 TWDM-PON testbed. © IEEE 2016. Reprinted, with permission, from [28.25]

OSVUs and two ONUs in FPGAs (i.e., Stratix IV GT edition EP4S100G2F40I1N). The OLT and ONUs are connected to four SFP+ evaluation boards equipped with four optical transceivers, facilitating two pairs of DS/US optical transmission channels. Traffic is generated by each ONU independently according to a Poisson process with the same average arrival rate p. Load variation is emulated by changing the TWDM PON capacity: one of the OSUs is turned off periodically (period: T_{rec}). Therefore, the network behaves as a TWDM PON (i.e., there are two active OSUs) for one period, and it behaves as a TDM PON (i.e., the two ONUs share the only active OSU capacity equally) for the next period. Figure 28.11 shows the average frame delay as a function of the reconfiguration period T_{rec} with different frame arrival probabilities p (note that low frame arrival rates are considered in the experiments because of FPGA memory limitations). If reconfigurations are performed, there is an optimal value of T_{rec} (i.e., T_{rec} = 100 ms) that minimizes the average frame delay. As detailed in [28.27], for all of the values of $T_{rec} > 0$ considered, the average energy efficiency achieved is constant (e.g., 27% when $p = 0.02$) for a given value of p because the overall amount of time for which the OSUs are ON/OFF is the same.

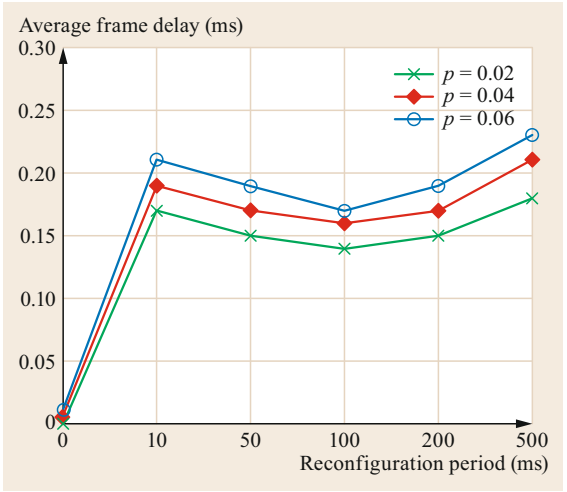


Fig. 28.11 Average frame delay versus T_{rec}, with T_t = 1 ms. © IEEE 2016. Reprinted, with permission, from [28.25]

Reducing Delay Penalties in an Energy-Efficient TWDM PON Through Reconfiguration Threshold Adaptation

As shown in the previous section, if the reconfiguration triggering threshold is close to the maximum load supported by the active OSUs, the average frame delay exhibits some peaks. The impact of the threshold at which OSUs are turned ON/OFF on the average frame delay is studied in [28.28]. An adaptive threshold policy is proposed in order to reduce the impact of reconfiguration on the average frame delay with limited degradation of the energy efficiency.

The same protocol and scenario utilized in the previous section and a newly proposed adaptive threshold policy are now considered. The policy is based on the decision curve depicted in Fig. 28.12. When the average network load decreases, a different set of thresholds is utilized (i.e., thresholds THi_D) compared to those used when the average network load increases (i.e., thresholds THi_U). In particular, when the average network load exceeds THi_U, all of the ONUs are immediately activated. This is done to cope with sudden load increases. The number of active OSUs is slowly decreased if the average network load decreases. Moreover, the thresholds THi_D are set to values that are below the maximum load that active OSUs can handle to enable the network to accommodate average network load fluctuations around the reconfiguration triggering threshold.

As in the previous section, the performance parameters of interest are the average frame delay and the average OLT energy savings. The simulation scenario is also the same as that used in the previous section.

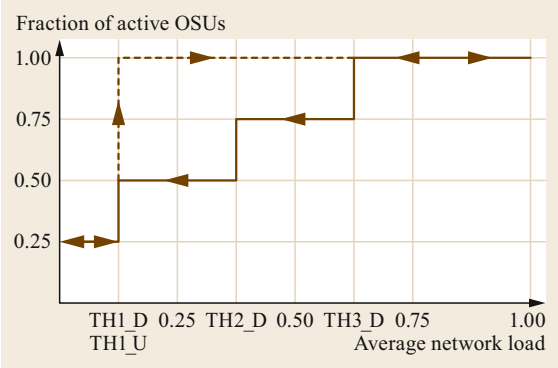


Fig. 28.12 Reconfiguration decision curve

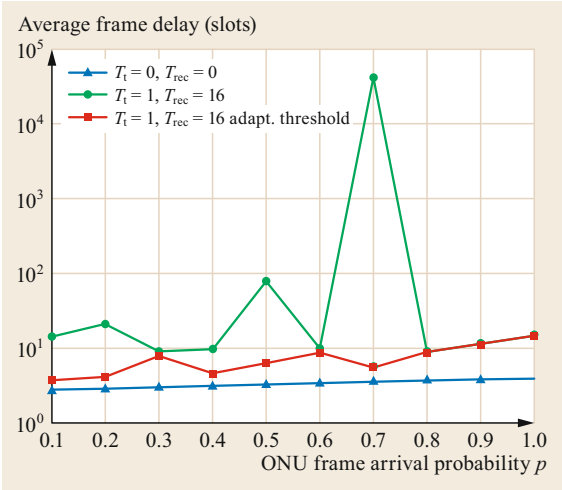


Fig. 28.13 Average frame delay

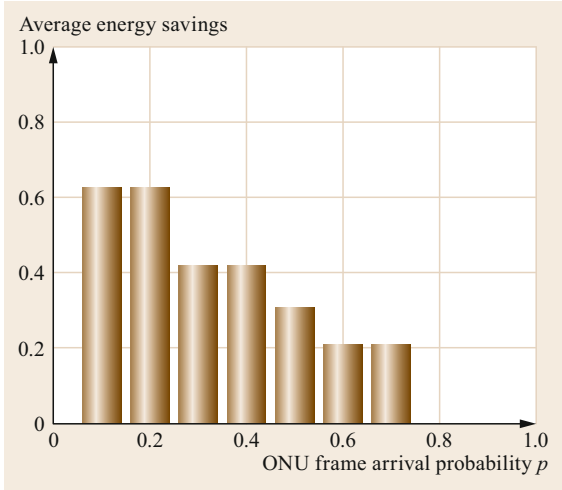


Fig. 28.14 Average energy savings with the fixed decision threshold

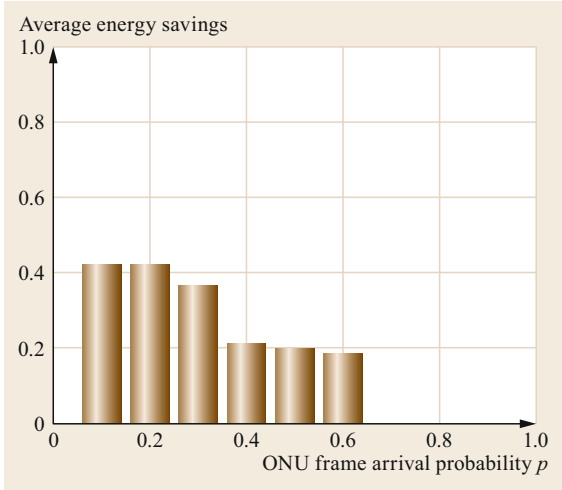


Fig. 28.15 Average energy savings with the adaptive decision threshold

All simulations last for one million slots and each value included in the plot is the average of twenty different traffic patterns generated utilizing twenty different random seeds. The resulting 95% confidence level is negligible, so it is not reported in the plot.

Figure 28.13 shows that if no reconfigurations are performed (i.e., all the OSUs are always active, $T_{rec} = 0$), the average frame delay is almost constant because the service rate is always less than or equal to the arrival rate. If reconfiguration thresholds are fixed at the maximum average network load supported by the active OSUs and the reconfiguration period $T_{rec} = 16T_s$, high delay peaks occur. If the proposed adaptive threshold policy is utilized (with TH1_D = 0.125, TH2_D = 0.375, TH3_D = 0.625, and TH1U = 0.125), the delay peaks are reduced with respect to those that occur when a fixed threshold is used. However, the energy savings achieved by using the adaptive decision threshold method are less than those achieved if the fixed threshold method is applied, as shown in Figs. 28.14 and 28.15.

Thus, by adopting an adaptive threshold policy, it is possible to reduce the average delay penalties when TWDM PONs are reconfigured. This reduction is achieved at the expense of a slight reduction in the saved energy.

28.2.3 How Much Energy Can Be Saved in a Commercial OLT?

All of the previously discussed methods provide interesting solutions to the problem of minimizing the OLT energy consumption. However, they assume that OLT energy consumption is directly proportional to the energy consumed by transceivers (i.e., OSUs) only. In real systems, the bulk of the energy used at the OLT is utilized to power shelves and racks. The energy consumption is independent of the number of transceivers that are active. This section aims to shed some light on this aspect by evaluating the real energy savings that can be achieved utilizing DWBA. Moreover, the impact of the tuning time on the average frame delay is evaluated.

The TWDM PON architecture considered is that depicted in Fig. 28.2, and the OLT model is depicted in Fig. 28.16. OLT line cards (LCs) are inserted into slots of a shelf installed in a rack. The following elements contribute to the overall OLT power consumption: the transceiver (TRX) array, the layer 2 (L2) switch for

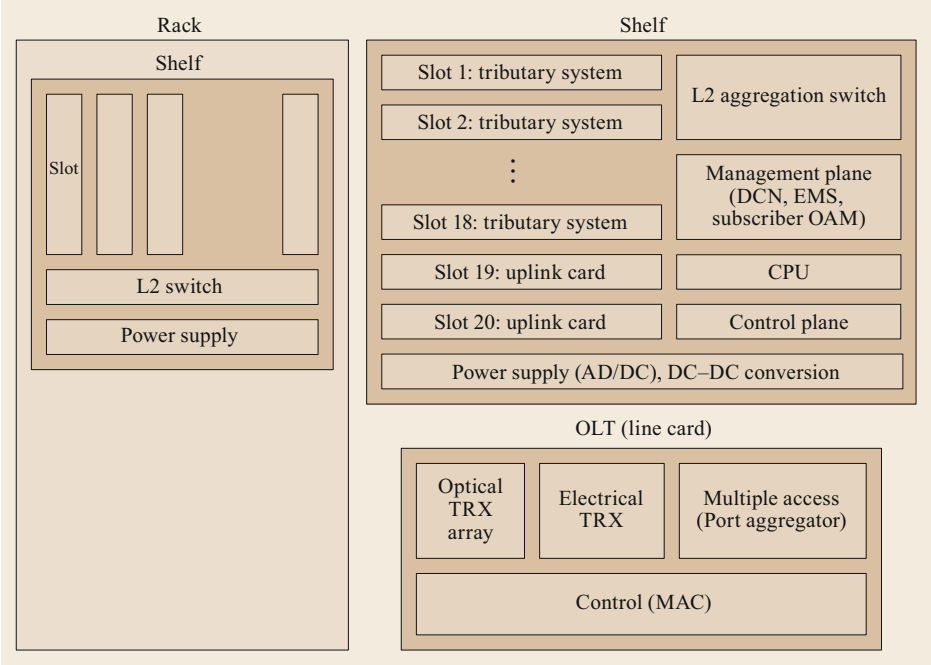


Fig. 28.16 TWDM-PON rack, shelf, and card mode

traffic aggregation, the port aggregator, the shelf power supply, and the amplifiers (if there are any). The power consumption of each component or subsystem utilized in the considered OLT is summarized in Table 28.8 as derived from what was reported in [28.29].

The performance evaluation utilizes a custom-built time-driven simulator coded in C. The time is slotted and each ONU transmits and receives upstream (US) and downstream (DS) data during time slots T_i^j of a fixed size, where i is the slot index and j is the cycle index. It is assumed that each frame transmission requires one slot and that the number of available slots in a cycle at one wavelength is equal to the number of ONUs. A frame that arrives at an ONU during a cycle can only be transmitted in the next cycle, even if slots are available in the current one, because wavelength and bandwidth allocation (WBA) is performed at the beginning of each cycle. The network control considered here is slightly different from the one described in the previous sections, but the WBA utilized is the same. The network control protocol is depicted in Fig. 28.17a and the WBA is depicted in Fig. 28.17b.

Table 28.8 OLT components and subsystem power consumption values

Component or subsystem	Power consumption (W)
OLT (line card)	
10×10 Gb/s TRX array TDMA colored line card (burst mode, including FEC)	20 (per line card)
Port aggregator	0.5 (per port, per 1 Gb/s)
Control (MAC)	1 (per port, per line card)
Shelf	
Basic shelf power	90 (per 20 slots; 18 slots for tributary)
L2 aggregation switch	1 (per 1 Gb/s, per card, per shelf)

The network control protocol is based on REPORT and GATE messages, as in EPON. A REPORT message containing the ONU queue status is transmitted by the ONU to the OLT in its slot. The OLT performs WBA at the end of each cycle T^{cj}. It then sends GATES to the ONUs carrying the wavelength(s) and the slot(s) in which they must transmit during the next cycle. GATEs are sent to the ONUs at the wavelength utilized in their latest scheduled slot. This signaling requires a time that is dependent on the WBA computation time T_{sch} and the signaling time T_{gate}, which, in turn, is proportional to the RTT. For simplicity, but without any loss of generality, the propagation delay is assumed to be negligible (i.e., round trip time = 0). Upon receiving the GATEs, the ONUs tune to the wavelength of their first transmission slot. To maintain synchronization among the ONUs, the scheduling is delayed by an amount equal to the time required by ONUs transmitting in the first slot of the cycle to tune to the assigned wavelength, if needed.

Slots are then scheduled and assigned a wavelength, which is when the WBA is performed. A longest-first (in terms of slots assigned to one ONU) and first-fit WBA is implemented, as depicted in Fig. 28.17b.

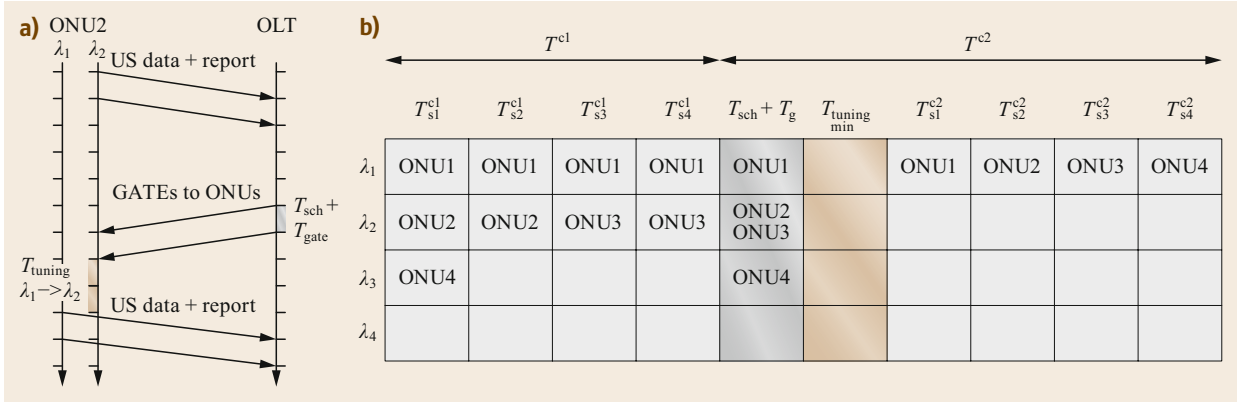


Fig. 28.17a,b Control protocol time chart **(a)** and example of wavelength and bandwidth (i.e., slot) assignment **(b)**

Symmetric US and DS traffic is generated based on a Bernoulli arrival process: in each slot i, each ONU n has a probability $P_i(n)$ of generating or receiving a frame. Since US and DS data transmissions are locked, the performance evaluation is valid for both US and DS data. In all simulations, the number of ONUs $N = 4$, the number of slots per cycle $S = 4$, and the number of OSUs (TRXs) at the OLT $L = 4$.

The power consumption values during ON and OFF periods are computed by assuming that the values reported in Table 28.8 are directly proportional to the OLT configuration considered, similar to the approach used for the FP7 OASE project (https://cordis.europa.eu/project/rcn/93075_en.html). The values reported in Table 28.9 are based on the following assumptions:

- The 4×10 Gb/s TRX array is assumed to occupy two shelf slots (one for TX and one for RX), and the power it consumes is linearly proportional to the power consumed by the 10×10 Gb/s TRX array.
- The power consumed by the TRX array is assumed to be proportional to the ports (i.e., TRXs) that are ON. Therefore, if all of them are OFF, the TRX array does not consume power.
- The port aggregator presents a bulk power consumption of 20 W and the rest of the power consumed by the port aggregator is assumed to be proportional to the aggregated bit rate.

The average OLT energy savings η and the average frame delay D are evaluated. The OLT average energy savings are defined as the energy saved when the TRXs are dynamically turned ON and OFF as a percentage of the energy consumed when all the TRXs are always ON. The average frame delay is defined as the average delay experienced by a frame between its arrival and its transmission.

Figure 28.18a shows that the bulk power consumption of the OLT allows maximum energy savings of only about 20%. Figure 28.18b confirms that the frame delay (shown on a natural logarithmic scale) increases if the average network load is close to the maximum load that can be supported by the active OSUs. This is mainly due to the impact of the tuning overhead T_{OH} and the suboptimality of the WBA. In particular, the frame delay increase is greater for higher average network loads.

In conclusion, the achievable OLT energy savings are limited by the bulk OLT energy consumption, and

Table 28.9 OLT power consumption during ON and OFF periods in the considered rack, shelf, and card model

Component or subsystem	Amount	Power consumption ON (W)	Power consumption OFF (W)
OLT			
4×10 G TRX array TDMA colored line card	2 line cards	16	0
Port aggregator	80 Gb/s	40	20
4×XG-PON control MAC	8 ports	8	8
Shelf			
Basic shelf power	3 slots (2 TRX array, 1 MUX/DEMUX)	10	10
L2 aggregation switch	80 Gb/s	80	80
Total		154	118

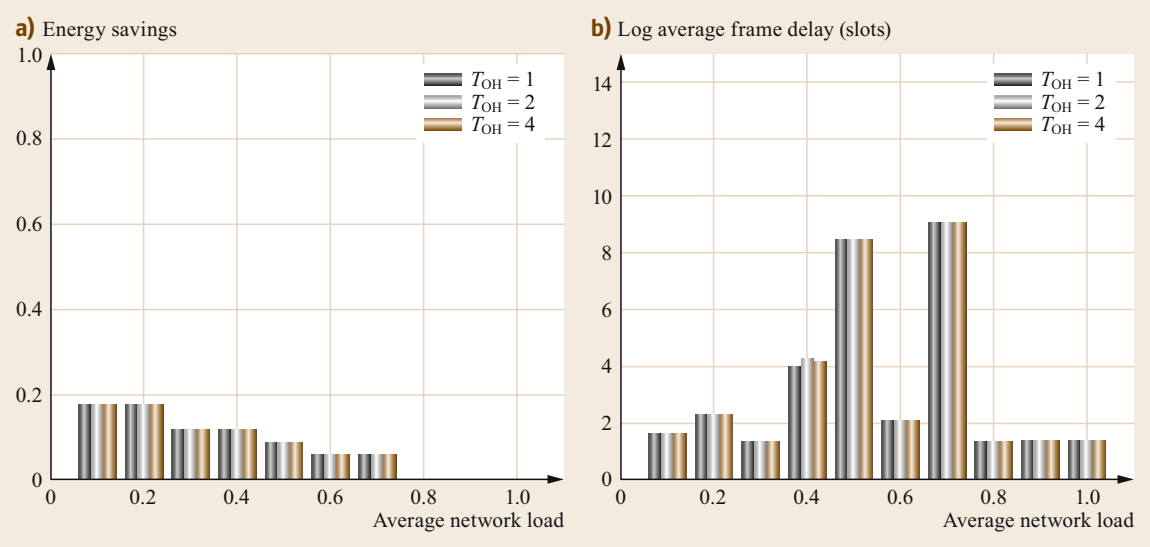


Fig. 28.18a,b Energy savings **(a)** and average network load **(b)**

the savings are achieved at the expense of an increase in the average frame delay. If the average network load is close to the load that can be supported by the active OLT transceivers, the increase in the average frame delay is larger still. These results confirm that the threshold for transceiver (de)activation must be chosen carefully, thus further limiting the achievable energy savings.

28.3 WDM-PONs

Although PtP WDM communication is included in the NG-PON2 standard, other solutions are also proposed in the literature for WDM-PONs. The main goal of each approach is to provide the connection between ONUs and OLT with a dedicated wavelength. For example, in [28.30], an architecture that uses an optical filter (e.g., an arrayed waveguide grating, AWG) at the remote node (RN) to route a single wavelength to an ONU is considered.

One of the key requirements for WDM PONs is a low-cost tunable transmitter. Important contributions to the overall cost of the tunable laser include the cost of the embedded wavelength locker/etalon, the material costs, and the assembly costs in particular. In [28.30] and [28.31], it is proposed that the individual wavelength lockers in each laser should be replaced with a centralized implementation in which only one locker in the OLT is shared by all attached ONU devices. Distinct continuous wave (CW) pilot tones (i.e., channel labels) are embedded into the US wavelengths to enable the calculation of individual feedback signals on the signal power and the relative wavelength deviation in the OLT. The feedback signals are then transferred back to the ONUs by an auxiliary management and control channel (AMCC). Eliminating the thermoelectric cooler (TEC) is another way of saving costs and reducing power consumption. Uncooled operation of both DS-DBR (digital super-mode DBR) and SG-DBR (sampled grating DBR) lasers has recently been demonstrated, with operating temperatures of up to 90 °C [28.30]. Finally, significant savings can be achieved at the OLT side of a WDM-PON system by integrating a multitude of transmitters and/or receivers into a photonic integrated circuit (PIC).

An overview of the possible approaches to implementing a WDM PON, which partially overlap with the NG-PON2 standard, is presented in [28.32]. In particular, different ways of generating the upstream signal are reported.

The simplest solution is to utilize fixed-wavelength transmitters. This solution can potentially achieve long transmission distances and high speeds, but the deployment of such a network would be cost prohibitive. Alternatively, tunable lasers can be utilized in all ONUs, with each laser tuned to the preassigned transmission wavelength.

Other possible solutions reported in [28.32] are laserless solutions at the ONU: wavelength reuse, injection locking/wavelength seeding, and self-seeding.

In *wavelength reuse* schemes, downstream wavelength channels are remodulated with upstream data and then sent upstream towards the CO, thus eliminating the optical source. In general, a reflective semiconductor optical amplifier (RSOA) is utilized at the ONU to implement this approach. By intentionally operating the RSOA in the gain saturation region, the amplitude squeezing effect is used to erase the downstream modulation of the seeding wavelength, and the resulting amplified RSOA output can be directly modulated with upstream data. Wavelength reuse presents the following benefits: remodulation of the downstream wavelength channel (eliminating the need for seeding sources) is less costly than using tunable lasers, and the RSOA modulation is direct. Note that the resulting RSOA output has a wavelength that is identical to the downstream wavelength, so upstream performance can be severely degraded by the interference between the residual downstream and upstream data at the CO. However, the residual downstream modulation can be minimized by ensuring that the upstream and downstream modulation formats are orthogonal.

In *injection-locking* schemes, optical light originating from the CO is fed into injection-locked Fabry–Pérot laser diodes (F–P LDs) or into wavelength-seeded reflective semiconductor optical amplifiers (RSOAs) at the ONUs. The injection-locking or wavelength-seeding light can be spectrally sliced light from a centralized broadband light source located at the CO. The wavelength-seeding scheme and the injection-locking scheme differ in that the former uses an RSOA that amplifies and modulates the incoming continuous-wave (CW) light. One advantage of injection-locking schemes is that all ONUs can be implemented with identical FP-LDs or RSOAs because the transmitting wavelength of a colorless ONU is determined externally by the wavelength of the incoming light. The main drawbacks are that an additional broadband light source (or sources) is needed and, most importantly, the transmission performance is impaired by factors such as fiber dispersion, Rayleigh backscattering noise, and broadband amplified spontaneous emission (ASE) noise from the broadband light source and the colorless source.

In some *self-seeding* schemes, the RSOA in each ONU is self-seeded by its own spectrally sliced CW light. The AWG located at the remote node spectrally slices the ASE light emitted from each RSOA. Self-seeding presents the following advantages: it removes the need for active temperature tracking between the optical components within the remote node and between the remote node and each RSOA, and identical RSOAs can be placed at all ONUs, as the wavelength of the seeding light and hence the wavelength transmitted by each RSOA are solely determined by the spectral characteristics of the AWG and BPF. One drawback of this scheme is that upstream performance is dependent on the initial optical seeding power, which affects the level of ASE noise suppression of the self-seeded upstream signal. Moreover, because of the nonzero polarization-dependent gain of the RSOAs, the self-seeding scheme is polarization dependent.

All of the aforementioned schemes are, however, still at the research stage, so ONUs based on tunable lasers were selected for the NG-PON2 standard.

28.4 OFDMA-PONs

Scaling up TDM-PONs to several tens of Gb/s of aggregate capacity can be challenging due to the speed limits of optoelectronic components, especially during burst-mode operation, and there can be bandwidth inefficiency from an ONU processing perspective. On the other hand, wavelength division multiplexed PONs (WDM-PONs) have been the focus of increasing interest as a means to deliver ultra-high-speed services, but they are currently costly to implement and have limitations in terms of their scalability and backward compatibility.

In the search for valid alternatives, orthogonal frequency division multiple access (OFDMA) has attracted considerable research interest in relation to passive optical networks (PONs) as it is used highly successfully in other wireless and wired access systems such as Wi-Fi, WiMAX, LTE, and ADSL, where it has shown advantages such as high spectral efficiency with QAM, powerful Fourier processing, and extraordinary adaptability to the transmission medium. Section 28.4 analyzes the adaptation of OFDMA to optical access and the relevant experiments performed, exploring pros and cons.

Section 28.5 presents advanced WDM-PON access, which densely exploits the optical spectrum (mostly through the application of ultra-dense WDM and coherent detection), and discusses techniques for reducing the complexity involved in implementing the *wavelength-to-the-user* concept in an efficient way.

Finally, Sect. 28.6 discusses the potential of optical code division multiple access (OCDMA) for PONs.

OCDMA has long been studied as a potential all-optical form of multiplexing because of its inherent transparency and its asynchronous operation, but there are practical issues with implementing OCDMA in the fiber domain.

Orthogonal frequency division multiplexing (OFDM) was originally conceived as a modulation method to achieve better transmission properties for bit streams in media that exhibit bandwidth limitations and nonuniform channel performance (i.e., radio, copper). It involves the use of several low-bit-rate link subcarriers that carry different QAM symbols simultaneously. In order to be orthogonal, individual subcarriers are spaced at multiples of the reciprocal symbol duration. This allows them to partially overlap (thus increasing the spectral efficiency) without interfering with one another at sampling.

As it is a form of pure high-capacity transmission, it has been extensively tested for optical fiber transport and has been found to transmit over a hundred Gb/s per optical carrier in long-haul and metro networks. In optical access networks, other potential features of OFDM can also be exploited:

- Fine bandwidth (BW) granularity for distributing the high shared aggregated bandwidth to the PON users in an efficient way. Current PONs utilize TDM and offer fine BW granularity, but their ONUs operate at the full aggregate speed, leading to high bandwidth and power inefficiency as the split ratio grows. This issue can potentially be alleviated with OFDM.
- Elastic BW provisioning via OFDM multiple access (OFDMA). As user demands are highly variable statistically and over time, the possibility of flexibly distributing the aggregate bandwidth in accordance with the demand is highly appealing.
- Since hardware costs are a critical influence on access network viability, low-cost components are required in ONUs and OLTs, such as DML, VCSEL, and low-bandwidth electronics. OFDM has the unique potential to adapt to their limited response and maximize the capacity available in nonideal media, especially when the DMT (discrete multitone) variant is used.
- Generally, new PONs must be backward compatible, given that there has been over a decade of splitter-based PON deployment around the world. This adds extra requirements, such as a large power budget and differential link loss support. The adaptability of OFDM is an important advantage.
- In addition, it enables the convergence of the optical infrastructure with standard wireless solutions that already utilize OFDM (albeit with different parameters), thus offering a way to integrate the dominant wired and wireless technologies into a seamless hybrid access network that supports various broadband services.

These aspects have been examined in studies of OFDM for optical access networks over the last decade, as is discussed below. Research has proven the feasibility of using OFDM for optical access networks but has also revealed the difficulties involved in implementing those aspects in the optical domain and at very high bit rates in a shared optical medium. The upstream direction is especially problematic because of the asynchronous and incoherent origins of the light sources. Therefore, the studies we consider below tend to be those that have focused on the multipoint-to-point upstream direction or have led to specific advances in the techniques used in OFDMA-PONs.

28.4.1 Pioneering Works

In 2008, a pioneering work in OFDM optical access demonstrated the transmission of 4 Gb/s over a distance of up to 100 km using low-cost GPON optics with a BW of about 1 GHz [28.33]. The modulation was 16-QAM over 16 subcarriers. Soon afterwards, real-time FPGA and fast DAC-ADC become available and were widely used for high-bit-rate optical transmission. 11 Gb/s were transmitted and processed in real time, with added functionality such as online performance monitoring and channel estimation [28.34]. A key advance was the incorporation of variable power loading, such that the power of each subcarrier was optimized according to the frequency response of the transmission medium in order to maximize the overall capacity. A spectral efficiency of 5.625 (bit/s)/Hz over 25 km for single-mode fibers was achieved with simple IM/DD downstream as part of the European project ALPHA.

More sophisticated systems were also developed in order to achieve higher aggregated rates and to effectively combine upstream signals. A potential PON capacity of 1.2 Tb/s for up to 90 km was experimentally validated in [28.35] by combining WDM with OFDM using a wavelength multiplexer at a local exchange and including power splitters in the distribution network. At every wavelength, the electrical spectrum was FDM sliced into individual OFDM subbands that were optically multiplexed at different RF frequencies over the same optical carrier sent from the OLT. In this form, the ONUs were colorless, and they adaptively selected the RF transmission channel and bandwidth and processed only their individual data using lower-speed processors. At the OLT RX, the upstream was detected coherently with an optical heterodyne front end and

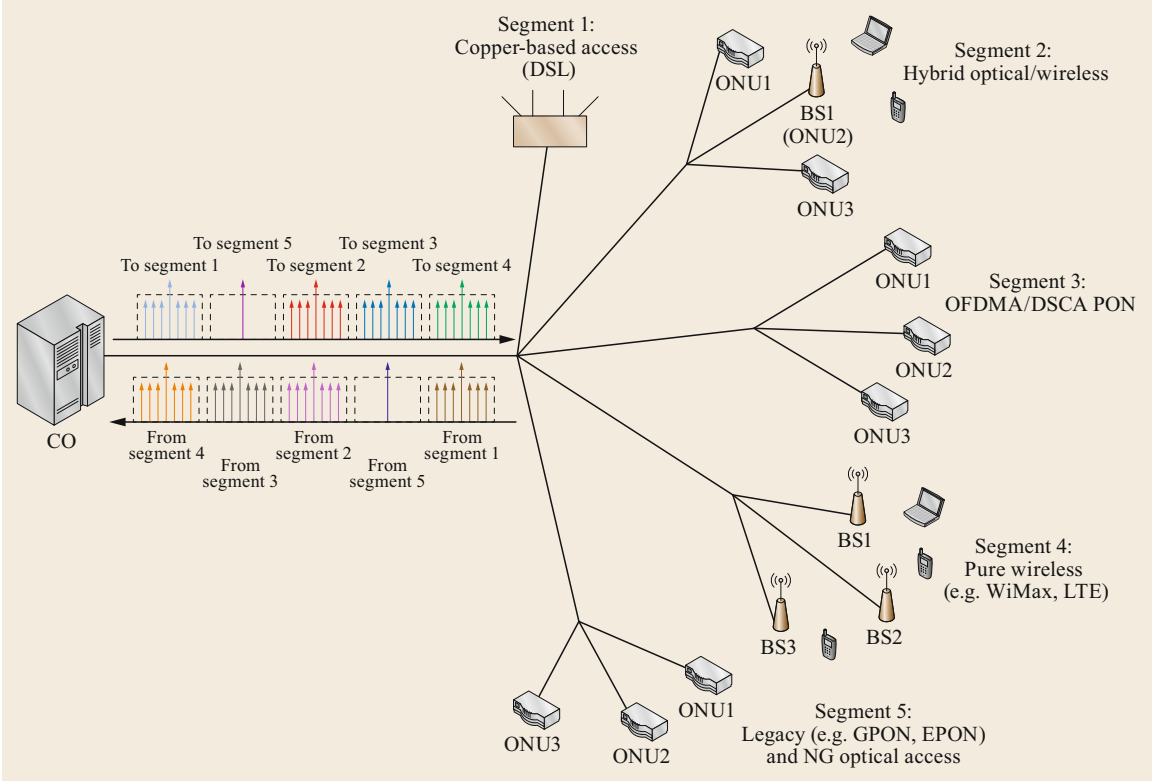


Fig. 28.19 The ACCORDANCE network architecture and the various types of connectivity involved

a multiband OFDM receiver. Later on, the same group incorporated a Meta-MAC and software-defined network (SDN) tools for hardware control.

In [28.36], the multiband OFDM approach with different levels of ONU bandwidths was employed in order to reduce ONU complexity and lower their energy requirements in IM-DD OFDM systems. In this technique, the ONUs included tunable radiofrequency (RF) oscillators to place/retrieve the assigned subcarriers (SC) in-/from the multiuser OFDM signal, thus reducing the bandwidth specifications and the power consumption.

ACCORDANCE OFDMA-PON Project

The European ACCORDANCE project [28.37] was a major consortium project that focused on OFDMA access. It aimed to realize the key application of OFDM, orthogonal frequency division multiple access (OFDMA), whereby different users are assigned to different OFDM subcarriers. OFDMA-based technology and protocols (physical and medium access control layers) were introduced to provide a variety of desirable characteristics. The introduction of OFDMA into PONs was implemented using a holistic approach, i.e., both the physical and the medium access control (MAC) layers were addressed in concert. Dynamic bandwidth allocation was employed, with straightforward subcarrier mapping to users and services, wired and wireless. Special emphasis was placed on the convergence of the optical infrastructure with standard wireless solutions, as well as migration from existing access technologies such as xDSL and xPON in order to enable a seamless OFDMA-based access network where virtually different telco services could be consolidated.

Its unified topology, as depicted in Fig. 28.19, consists of an optical line termination (OLT) located at the central office (CO), several optical network units (ONUs), and the fiber distribution network with power splitting and optional wavelength multiplexers for WDM overlay. The OLT distributes spectral segments of variable width among different services or parts of the network in a dynamic fashion. The MAC protocol and scheduling were implemented on FPGA for real-time transmission together with the physical layer OFDM processing.

MAC Layer

The tailored MAC protocol operates on two levels:

- *Intersegment*. Here, the spectral segments are provisioned to different service clusters (the aggregate

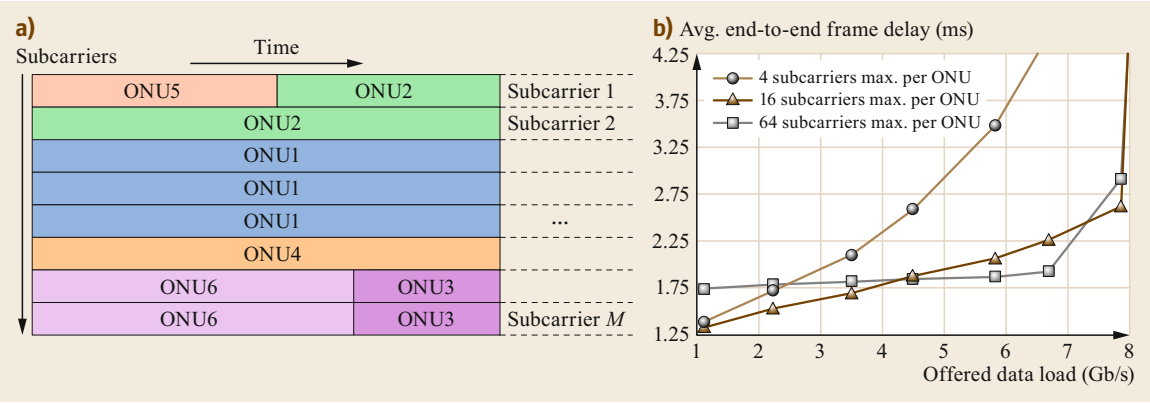


Fig. 28.20a,b Hybrid OFDMA/TDMA (intrasegment) MAC operation in ACCORDANCE **(a)**, and end-to-end delay versus offered load for different numbers of ONU subcarriers in a single segment of the architecture **(b)**

capacity in each segment depends on the exact service types) in a fair way.

- *Intrasegment*. At this level, the ONUs belonging to each service cluster share the subcarriers within the corresponding segment. This bandwidth assignment is more dynamic and thus more challenging from a MAC point of view, which was why it has received greater attention.

Various modes of DBA are supported, depending on the ONU type and requirements:

- Fixed subcarrier assignment, where the ONU uses the indicated subcarrier range in a continuous manner until it is instructed otherwise. This mode is suitable for less dynamic traffic since it leads to virtual circuits, making it particularly useful for mobile/xDSL/xPON backhauling scenarios.
- On the other hand, there is the hybrid OFDMA/TDMA mode, involving the formation of transmission pipes or two-dimensional *rectangles* updated once every scheduling window, as shown in Fig. 28.20. This is expected to prevail for bursty traffic patterns, as it provides the finest granularity during network bandwidth assignment.
- A separate channel consisting of a number of subcarriers is reserved for exchanging control information. The most important additional control fields in terms of bandwidth assignment are the assigned low/high subcarrier indices (which, in conjunction with the start/stop times, define the assigned bandwidth rectangles). In contrast to a typical PON MAC, this information is also conveyed for downstream transmission, since ONUs should not process the entire subcarrier range.

In addition, the MAC layer handles the procedures relating to the adaptive modulation concept. This involves the following additional functionality:

- The introduction of a new physical layer operation and maintenance (PLOAM) downstream message for the OLT to indicate the exact m-QAM format to be used by the ONU.
- Transmission performance (i.e., BER) monitoring by the OLT through appropriate MAC PLOAM message exchanges during the registration process (a series of them are needed per ONU to find the optimal format for a given BER threshold).

Figure 28.20 also presents results obtained assuming an EPON-based online scheduling framework, whereby each ONU can be dynamically allocated a variable number of subcarriers up to a certain value, as described above (the rectangle offering the minimum delay is selected each time). A segment of 10 Gb/s is considered, which is split among 64 data subcarriers and serves 32 ONUs located at distances from the OLT of between 0 and 100 km. As shown in the figure, given the significant processing effort (at the ONU side) as well as the increase in scheduling complexity (at the OLT side) as the subcarrier range increases, the optimal setting is a maximum of 16 subcarriers per ONU.

In the OFDMA-PON system (single λ), both downstream and upstream transmission were implemented in specially designed FPGA modules for the OLT and, albeit with a lower capacity, at the ONU. The respective sampling rates were 25 and 3.125 GS/s, which were mostly dictated by the affordability of high-speed DACs/ADCs. Data was organized into synchronized OFDM frames that had a fixed duration of

a)

Solution #	Downstream				Upstream				
	OLT TX		ONU RX		ONU TX			OLT RX	
	oIQ	IM	DD	CO	Non-seeded		Seeded	DD	CO
					oIQ	IM	IM		
1	✓		✓				✓		✓
2	✓		✓			✓			✓
3	✓		✓		✓				✓
4	✓			✓	✓				✓
5		✓	✓			✓		✓	
6	✓		✓			✓		✓	

b)

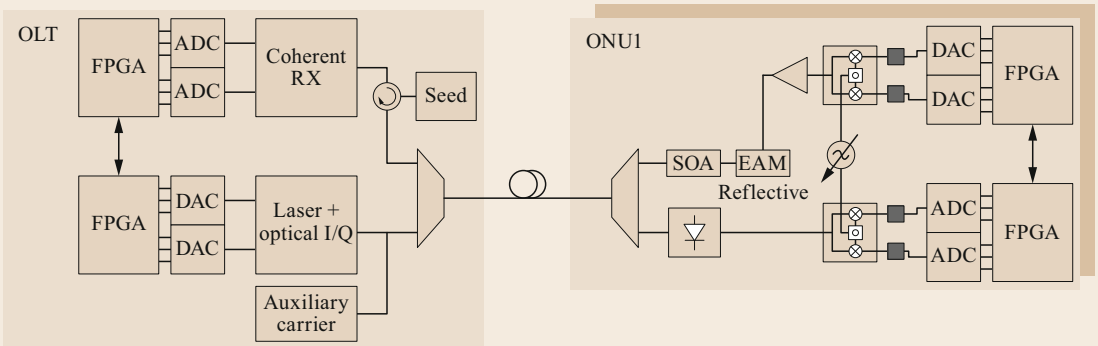


Fig. 28.21a,b Potential PHY technological solutions for realizing an OFDMA-PON (**a**), and a schematic of the OLT and ONU in solution #1 (**b**)

105.6 µs and included 8250 OFDM symbols of duration 12.8 ns [28.38]. A cyclic suffix lasting 25% of the symbol duration was used for dispersion management and simple FFT over 256 subcarriers. These were distributed into 16 spectral groups, with two subcarriers operating as control pilots. The MAC layer controlled the bandwidth allocation as explained. At each frame, after 32 training symbols for time and frequency synchronization, a common control block was distributed to all ONUs in all spectral groups. A functional MAC protocol using TDMA was implemented on Linux-based embedded processors that performed protocol functions such as automatic ONU registration and dynamic bandwidth allocation. Then, according to the MAC, the payload data assigned to each ONU was modulated using a specific *m*-QAM format in the defined virtual channel (VC) as *rectangles* in consecutive frames in a spectral group.

Physical Layer

The transmission and cost asymmetries of an OFDM-PON lead to various possible TX and RX architectures downstream and upstream. In the literature, six implementable TX and RX combinations (as compared in Fig. 28.21) have been proposed that vary in the type of modulation used (IM or optical IQ), whether the ONU is wavelength seeded from the OLT, and the form of detection (direct or coherent), with a range of tradeoffs between complexity and performance. Comparative analysis of the various solutions led to the following conclusions: lower-cost IM options (#5, #6) are limited to scenarios below 40 km by the PAPR, the sensitivity, and the chromatic dispersion, even when optical amplification is used; coherent detection at the OLT (#1–4) is required and can be justified because its cost is shared; optical amplification is required to reach 100 km in all cases; and unseeded solutions (#2–4) with coherent detection at the OLT would require the ONU lasers to be extremely fine tuned to maintain the orthogonality. Solution #1 solves this with a seeded ONU that uses an EAM reflective modulator and coherent detection at the OLT, and was thus chosen as the most cost-effective solution for the intended performance. Downstream, optical IQ (oIQ) modulation is used at the OLT, which preserves the OFDM signal linearity, and simple direct intensity detection at the ONU. To minimize the DSP and DAC requirements at the ONU, the OFDM signal is upconverted to its corresponding electrical spectral group, modulating the seeded optical carrier at the EAM. A guard band is included at low frequencies to minimize the effects of Rayleigh backscattering and intermodulation beating. The modulated upstreams from the ONUs are combined in the feeder fiber and detected at the OLT coherent RX. All

of them share the same seeded carrier, so coherent reception and OFDM group processing are practicable.

In downstream tests, a real-time OFDM transmitter achieved line rates of 101.5 Gb/s [28.39]. The 64-point IFFT was computed and 58 subcarriers were modulated with 16QAM. Four pilot tones aided phase recovery at the receiver. All but the outermost subcarriers showed an EVM well below the FEC limits. Upstream, heterodyne OFDM and a novel multicarrier Nyquist FDM (NFDM) PON access network compatible with the WDM overlay were implemented [28.40]. The ONUs remodulated the US seed from the OLT with its specific assigned subcarrier/s by means of the EAM. Interestingly, the optical carrier seed was filtered out at the OLT, together with Rayleigh backscattering, and the signals were heterodyne detected using the same original seed within the OLT. Thus, using commercially available equipment, a symmetric network demonstrator with a capacity of 31 Gb/s was realized, with a power budget of 23–31 dB, depending on the accepted launch power.

In the same project, a lower-cost solution was also developed, avoiding coherent detection, costly lasers, and high-speed DSP. It used conventional, directly modulated DFB lasers with no preselection of the nominal wavelength. Statistical spectrum allocation for nonpreselected light sources and relative wavelength control was proposed. Since upstream signals are optically coupled in the PON splitter, the issue of optical beat interference (OBI) was addressed. Even though they are generated at different RF frequencies, the coupling in the PON splitter of simultaneous upstream signals that are close to each other in the optical spectrum can lead to beating between them, generating uncontrolled critical crosstalk at the OLT DD receiver. This requires the use of tunable lasers to separate their respective wavelengths or the use of nontunable lasers (such as DFB lasers) with different wavelength channels. However, the latter option is not acceptable in PONs where ONUs must be identical or colorless. To solve this, the proposed method employs independent nonpreselected ONUs with lasers operating at random wavelengths that are statistically combined in the PON, and slight thermal tuning when there is a risk of OBI [28.41]. Other works have also proposed forms of statistical WDM to reduce the splitting losses; these use a cyclic AWG [28.42] or adapt the laser fabrication yield [28.43].

Due to its variable parameters and DSP, OFDM inherently encompasses a high degree of flexibility and adaptability to the transmission medium and bandwidth demands. In access networks, users in the same PON are widely spread geographically and afford different splitting ratios [28.44]. PON standards define a minimum differential link loss among the users of 15 dB, which places a severe limitation on TDM-PONs. OFDM can, however, modify its dynamic parameters to overcome this. In [28.45], subcarriers (SC) are assigned flexibly and modulation levels are dynamically varied according to the user's power budget availability. This multiband technique reduces the processing effort in the ONU by setting its allocated BW into nonoverlapping orthogonal frequency bands via a RF mixing stage and reducing the fast Fourier transform (FFT) size, incurring only a minimal penalty with respect to the digital subcarrier allocation technique. In QPSK modulation, a spectral guard band of 6% is included upstream to detect the different ONUs, compensating for a differential link loss of 20 dB among the users.

28.4.2 Recent Works

Another European consortium project, OTONES [28.46], developed a novel topology that combines OFDM with very dense WDM using a new optical remote hub (RH), with interleavers and offset AWGs feeding multiple passive split ODNs with specific spectral structure. In this approach, 12.5 GHz spectral slices are subdivided into four 3.125 GHz slots with different functions (downstream, upstream, pilot seeding, and guard band). Each slot is subdivided into ten streams of BW 312 MHz for each ONU transmitting at 1 Gb/s with a polarization diversity of 16-QAM OFDM. The resulting ONUs are self-coherent, laserless, colorless, and do not have tunable filters. They comprise a standard SOA, a silicon-based photonic integrated circuit (PIC), and mixed-signal electronic integrated circuits (ICs) that perform signal processing at a relatively slow rate. The OLT uses a homodyne receiver. Transmitting 40G/40G in total over an optical bandwidth of 50 GHz, the spectral efficiency is 1.6 (bit/s)/Hz.

A pure OFDMA approach, as opposed to previous lab experiments where each ONU generated its own signal in a predefined subband, was used in field experiments in [28.47]. This approach was employed in both upstream and downstream (US and DS) transmission, with flexible bandwidth allocation per ONU. Using this scheme, demodulation of the signals at the coherent OLT is realized by using a single FFT for each polarization for all ONUs. This makes guard bands unnecessary, thereby increasing the spectral efficiency and allowing arbitrary subcarrier allocation for the ONUs. In the US, four colorless ONUs modulate a remotely seeded laser, and the accumulated data rate of the ONUs is 6.5 Gb/s with QPSK. Dynamic bandwidth allocation and power loading were implemented. In the DS, a di-

rect detection offset single-side band (SSB) approach is used to transmit 20 Gb/s with 8-QAM. In order to further increase the power budget, trellis coded modulation (TCM) was applied to improve the BER performance without expanding the signal bandwidth, so that a total of 32 ONUs were supported in both up- and downstream transmission.

Another interesting work on coherent OFDM demonstrated a simpler polarization-insensitive receiver that uses an Alamouti-coded OFDM signal combined with heterodyne reception [28.48]. This technique removes the requirement for a PBS, reducing the receiver optical components to a coupler, a single balanced photodiode, and a local oscillator (LO) laser; in addition, only a single high-speed ADC is required. Bidirectional symmetric transmission of 8×10.7 Gb/s WDM signals over an installed fiber PON and a power budget of 44 dB were demonstrated.

In a different approach, the whole OFDM FFT was performed optically at the remote node with a AWG to demultiplex the OFDM subcarriers without using high-speed electronics at the ONU [28.49]. The OFDM-modulated subcarriers were generated at the centralized OLT by electronic DSP. Using on-off keying (OOK) modulation for each subcarrier enabled the use of DD receivers, guaranteeing simplicity at the ONUs. Experiments demonstrated that the hybrid OFDM system permitted the transmission of 4×10 Gb/s over a SSMF more than 40 km long as a proof of concept.

Recently, [28.50] demonstrated practical real-time OFDM transmission with a BW of 16 GHz using real-time low-speed ONUs with 1 GS/s DACs-ADCs of BW 250 MHz and electrical IQ modulators. The OLT operated at 16 GS/s with 16-QAM, for a total aggregated capacity of 64 Gb/s. A relevant aspect here was the processing algorithm used to precisely eliminate imbalances due to carrier and sampling frequency offsets (CFO, SFO). Another recent work [28.51] applied generalized FDMA with circular pulse shaping of waveforms to enhance MAI (multiple access interference) tolerance and improve the budget by 3 dB with respect to OFDMA, along with 5 Gb/s ONUs. A different approach was taken in [28.52], which used delta-sigma modulation to convert the OFDM signal into a digital form to enable the use of standard digital transceivers. A sensitivity improvement of 4 dB was obtained for a capacity of 3 Gb/s.

An optically configurable OFDM RF signal transmitter is demonstrated in [28.53]. This is based on a sliced ASE source for multiband OOFDM-WDM networks, which is selected by adjusting the delay of a Mach–Zehnder interferometer (MZI) at the ONU. Experiments demonstrated the viable implementation of OFDM access and multiband transmission upon the appropriate reconfiguration of the electrical transfer function of the global optical link. Testing was performed with QPSK and 16QAM OFDM signals at a data rate of several Gb/s over frequencies of 3.5 and 5 GHz in a 10 km WDM access network.

FDM-PONs

Other works have dealt with nonorthogonal FDM for multiplexing many low-bit-rate signals in the RF domain at the optical feeder. This represents a simplification with respect to OFDM, but has a more practical granular implementation upstream, and can deal with a wide range of ONU rates.

For example, in the European project FABULOUS [28.54], a silicon Mach–Zehnder modulator (MZM) was used as a reflective ONU to show that a FDMA PON provides different classes of customer service. A more recent work used very low cost optical devices (VCSEL and RSOA) to transmit 16 FDM $\times$ 500 MBd with 16 QAM over 20 km, providing 23 Gb/s in a bandwidth of 9 GHz, at 2 Gb/s per ONU, with a power budget of 31 dB [28.55]. A capacity improvement was achieved by discrete multitone modulation (DMT). The system was also overlaid with mode division multiplexing (MDM).

The set of results discussed in this section, and other published results, demonstrate that the orthogonal FDM format makes it possible to reuse the robustness and the bandwidth efficiency concepts of modern wireless communications for optical access if properly adapted, with dynamic bandwidth allocation by subcarriers mapping to users and services (wired and wireless). The future success of this approach appears to depend on the cost of fast digital signal processors and on sensitivity improvements, which are limited by the high PAPR.

28.5 Ultra-Dense WDM-PONs

Currently deployed FTTH PONs, based on TDM equipment, use only one or two wavelengths and may not cope with expected bandwidth demands in the future, while OFDMA systems are limited by the bandwidths of analog and digital electronic as well as the poor sensitivity caused by the PAPR. An alternative to extending the electrical bandwidths of TDMA and OFDMA involves exploiting the optical spectrum. Multiplexing

the optical signals of the ONUs in the optical domain implies a reduction in the electronic requirements (bandwidth and power consumption) of the ONU and OLT terminals for the same aggregate capacity. Considering the required spectral efficiency and the tuning limitations of optical devices, this can be achieved by ultra-dense WDM (UD-WDM) multiplexing, by using very narrow optical filtering techniques, or by coherent homodyne detection, as will be described in this section.

Here, we use the term ultra-dense WDM to refer to a system with a channel spacing of less than 25 GHz, in accordance with the DWDM frequency grid standardized by the ITU-T in G.694.1. If we assume that user bandwidth demands will increase from 100 Mb/s to 1 Gb/s, a channel spacing of considerably less than 50 GHz must be pursued to achieve a substantial improvement in the spectral efficiency over a wide optical band.

28.5.1 Direct-Detection UD-WDM-PONs

A few works have considered ultra-narrow WDM that uses direct-detection receivers, which is limited by the optical filtering required. Tunable filters have not yet achieved narrow selectivity, a wide tuning range, and the required stability; cascading them will result in substantial optical losses.

A representative work [28.56] experimentally demonstrated a UD-WDM-PON system with 8 wavelengths at 10 Gb/s in a 12.5 GHz grid, with digital pre-equalization adopted at the transmitter. The pre-equalization considered the transfer function of the narrow-band optical filter. PON transmission, for up to 45 km, used a tunable filter, an interleaver, and a wavelength-selective switch (WSS) to select the channels. A less complex solution is presented in [28.57], where high spectral density was achieved through the combination of an ultra-dense AWG (channel spacing of 12.5 GHz), a comb generator at the OLT to provide the seed light, and a RSOA that is directly QPSK modulated at the ONU.

28.5.2 Coherent UD-WDM-PONs

Coherent transmission systems are currently deployed in long-haul WDM networks, and are not considered for use in PONs. However, because bandwidth demands are set to increase, optical coherent detection with a local laser has emerged as a future-proof alternative way to limitlessly increase the number of users who can be transparently multiplexed in the optical domain. This is enabled by achieving narrow selectivity of electrical filtering at the coherent receiver, rather than optical filtering, and by improving the sensitivity.

In the late 1980s and early 1990s, pioneering works on coherent systems were carried out. However, in practice, the systems demonstrated had only a limited impact due to several constraints of the state-of-the-art technology at the time, and because of the subsequent invention of EDFA-supported direct-detection long-haul systems, which meant that coherent detection was largely overlooked. The modern evolution of photonic and electronic integration technologies and the advanced digital processing capabilities that are now available have reignited research into many of the initial concepts and ideas associated with coherent systems, and have led to attempts to apply those concepts to the low-cost access arena.

The recent development of high-speed DSP ASICs has opened the door to coherent systems, as it removes the need for an optical phase locked loop (OPLL). The detection of the electromagnetic field, rather than the intensity, permits impressive DSP-based compensation for transmission impairments such as fiber chromatic dispersion and PMD, the use of higher-order modulation formats, and the ability to effectively receive polarization multiplexed signals, thus increasing the spectral efficiency. However, ad hoc ASICs for DSP and coherent optics are very expensive. DAC/ADCA and DSPs for FTTH can operate at relatively low speeds and can reuse or adapt common technology developed for the mobile and optical transport sectors.

It should be noted that the commercial development of UD-WDM-PON coherent transceivers for use in access networks has been slow, in part due to stringent requirements in terms of laser spectral width (laser phase noise) and polarization mismatch handling. Other issues relating to transceiver bidirectional transmission over a single fiber are also relevant, as Rayleigh backscattering, reflection, and modulation formats are being researched. If the technical hurdles can be solved, coherent transmission has the potential to be highly advantageous to PONs. Benefits could include:

- Highly improved selectivity due to the use of electrical instead of optical filtering
- Greatly increased spectral efficiency through the use of higher-order modulation formats
- Major improvements in the sensitivity and power budget, enabling a high splitting ratio and longer distances
- Transparency and scalability in occupying the fiber bandwidth due to the independent operation of ONUs
- Backward compatibility with deployed splitter-based TDM-PONs.

Initial approaches focused on heterodyne optical coherent receivers, but they have an image frequency problem and require high electrical bandwidth, so more advanced systems have made use of homodyne or intradyne reception.

Phase Noise Handling

Laser phase noise is one of the major limitations on the performance of optical coherent receivers. It is related to the random Langevin forces that occur inside the laser cavity. In order to solve this problem, optical TX-LO synchronization or a compensation system is needed at the receiver. This is especially important at the lower bit rates of the PON user, and with the requirement for low costs.

In order to properly synchronize local laser and received signals, early homodyne coherent systems used an optical phase-locked loop (oPLL) module. Several architectures were proposed and analyzed: decision-driven [28.58], Costas [28.59], balanced [28.60], subcarrier-modulated [28.61], and heterodyne [28.62] loops. However, all of these have the same problem: the optical path between the local laser and optical mixer (e.g., the optical hybrid and photodetection stages) introduces a substantial loop delay [28.63], resulting in a significant phase tracking penalty. In order to avoid this, it is necessary to use extremely low linewidth tunable external cavity lasers (ECL), which are bulky and complex.

Another approach to homodyne reception came later, with the concept of zero-IF/intradyne diversity receivers. The main goal of these receivers was to replace the feedback loop (oPLL) with feedforward postprocessing. For full phase diversity schemes (using a 90° hybrid), the architecture uses a free-running local oscillator, which is nominally at the same frequency as the incoming signal, but with a tolerance. Several relevant postprocessing approaches have been utilized, depending on the estimation/decoding: digital Wiener filtering [28.64], regenerative frequency dividers [28.65], the Viterbi–Viterbi algorithm [28.66], fuzzy logic data estimation [28.67], and differential detection [28.62].

In these systems, the core component is the 90° hybrid, which, for low-cost FTTH, could be integrated as a polymer waveguide device [28.68], as this requires fairly simple and inexpensive fabrication involving low-temperature processes and low-cost packaging based on passive alignment.

A third approach is time-switching phase diversity, where the receiver has two main parts: an intradyne coherent receiver with an added phase modulator at the local laser output, and electronic postprocessing at two taps per bit. The optical phase modulator at the local laser output is controlled by a data clock, which produces periodic 0–90° phase modulation to yield the I and Q signal components at the first and second halves of each bit time, respectively, after the optical homodynation [28.67, 69, 70]. A further simplification consists of directly driving the local laser with a sinusoidal wave at the bit frequency [28.71]. With the laser working in saturation mode, this takes advantage of its adiabatic chirp and leads to sinusoidal frequency modulation that results in phase modulation (0–127° peak-to-peak), yielding further linewidth tolerance enhancement of the time-switching diversity. This constitutes a low-cost, real-time ONU with a conventional DFB laser and not a 90° hybrid at UD-WDM spacing.

Table 28.10 compares the main techniques available in terms of the linewidth tolerance for a 10^{-3} BER and the baseline sensitivity penalty. The complexity of each technique is also highlighted, as well as the key optical component.

Polarization Mismatch Handling

Polarization is another major issue in coherent systems, as the optical mixing of the LO and the received signal fields occurs per electromagnetic field component and thus needs to be aligned in polarization. There are four different approaches to dealing with the polarization mismatch after link propagation: full polarization

Table 28.10 Summary of phase noise cancellation techniques. The linewidth tolerance values are for a 10^{-3} BER, and the sensitivity penalty values shown were obtained with respect to an ideal system

Technique	Linewidth tolerance	Sensitivity penalty (dB)	Key optical component	Complexity
Decision-driven loop	5 MHz	0	90° hybrid	High
Costas loop	4.9 MHz	0	90° hybrid	Medium/high
Subcarrier loop	5.1 MHz	0	90° hybrid	High
Balanced loop	2.4 MHz	2	Optical coupler	Low
Heterodyne loop (hom. det)	6.4 MHz	1	Optical coupler	Low
Heterodyne loop (het. det)	6.4 MHz	4	Optical coupler	Low
Full phase diversity	5% bitrate	0	90° hybrid	Medium
Time switch (scrambler)	1.8% bitrate	4	Phase modulator	Medium
Time switch (direct-drive)	5.4% bitrate	4	High-chirp laser	Low

diversity, time-switching polarization diversity, polarization scrambling, and local polarization control:

- Full polarization diversity schemes using a polarization beam splitter (PBS) and double balanced receivers have been extensively developed and integrated into transport networks. The implementation of such schemes doubles the number of optical components and necessitates DSP processing for polarization-insensitive and MIMO operation if the capacity is doubled by polarization multiplexing.
- In time-switching polarization diversity schemes, each bit time is split into two, with the first part containing horizontal (H) polarization information for example, and the other containing vertical (V) polarization information [28.72, 73]. This diversity is achieved by driving the clock signal to a polarization switch/scrambler. Afterwards, at the electronic postprocessing block, the H and V components are appropriately combined. This approach allows simplified optics and avoids the need for a 90° optical hybrid and half the photodetectors, but it also comes with drawbacks: a 3 dB loss of sensitivity and increased bandwidth.
- Polarization scrambling can be implemented using a highly birefringent modulator operating at the baud rate clock or higher. It can be performed either at the customer's premises or at the central office (a better approach, as the cost of the device can then be shared).
- Local polarization tracking requires the inclusion of a polarization actuator (a liquid crystal, Faraday rotator, or fiber squeezer to mention a few) in the customer's equipment, which is relatively expensive. It also means that an additional electronic control (at low frequency) must be placed at the electronic part of the receiver.

Handling Rayleigh Backscattering

One of the key issues in PONs is full transmission bidirectionality over a single access fiber. If downstream and upstream spectra overlap, Rayleigh backscattering may become a substantial impairment to the received signal for longer fibers [28.74]. This is especially harmful in coherent systems, as they are enable to operate at very low received powers and are easily overwhelmed by the Rayleigh backscattering. A double-fiber scenario is not acceptable for modern PONs, so there are the following options:

- A dual laser configuration at the ONU, with one laser used as a LO for detection and the other used for upstream transmission in a different optical band (as in commercial TDM-PONs)
- Heterodyne detection with one ONU laser that is simultaneously used as a RX local oscillator and as a TX with an external modulator
- The inclusion of a wavelength-shifting device at the ONU [28.75].

28.5.3 COCONUT UD-WDM-PON Project

Following the promising pioneering experiments described above, a major international project aimed to fully demonstrate a UD-WDM-PON, addressing all of its key aspects. The goal for the envisioned access network solution was to establish a path to a scalable *wavelength-to-the-user* (WTTU) concept. The key enabling technology was a new and simple coherent detection scheme with flexible wavelength allocation. The resulting UD-WDM optical access network was developed as the EU COCONUT project (COst-effective COhereNt Ultra-dense-WDM-PON for λ-To-the-user access, 2013–2016) [28.76–78]. The project developed coherent transmitters/receivers that utilize low-cost components and simple electronics, ensuring that typical line terminals are affordable to end users.

Architecture

The reference distribution architecture (Fig. 28.22) is compatible with the standard passive optical networks used by current PONs with power-splitting optical distribution networks (ODNs). Various application scenarios were considered: residential, business, remote fronthaul, and small and macro cell backhauling [28.77], with dedicated wavelengths or with shared wavelengths facilitated by applying a TDMA protocol at the MAC layer [28.79].

To cope with future demands in these scenarios, the optical line terminals of the central office (CO OLT) can be dimensioned to support 256 ONUs by using 64–256 wavelengths at 6.25 or 12.5/25 GHz spacing for data rates of 1.25 or 10 Gb/s, respectively. Thus, the terminals can provide appropriate wavelength selectivity over the UD-WDM flexible grid. For viable mass production and provisioning, the terminals should be inexpensive and identical without using preselected wavelengths.

The basic hardware architecture of the ONU and OLT at the CO consists of components such as UD-WDM coherent transmitters (TX) and receivers (RX), downstream/upstream combiners, TX and RX local oscillator (LO) laser automatic wavelength controllers (AWC), a CO digital switch, and a CO wavelength manager module. Among these components, a CO digital switch can map external data channels to OLT coherent transceivers. A CO wavelength manager module includes a high-resolution spectrum monitor

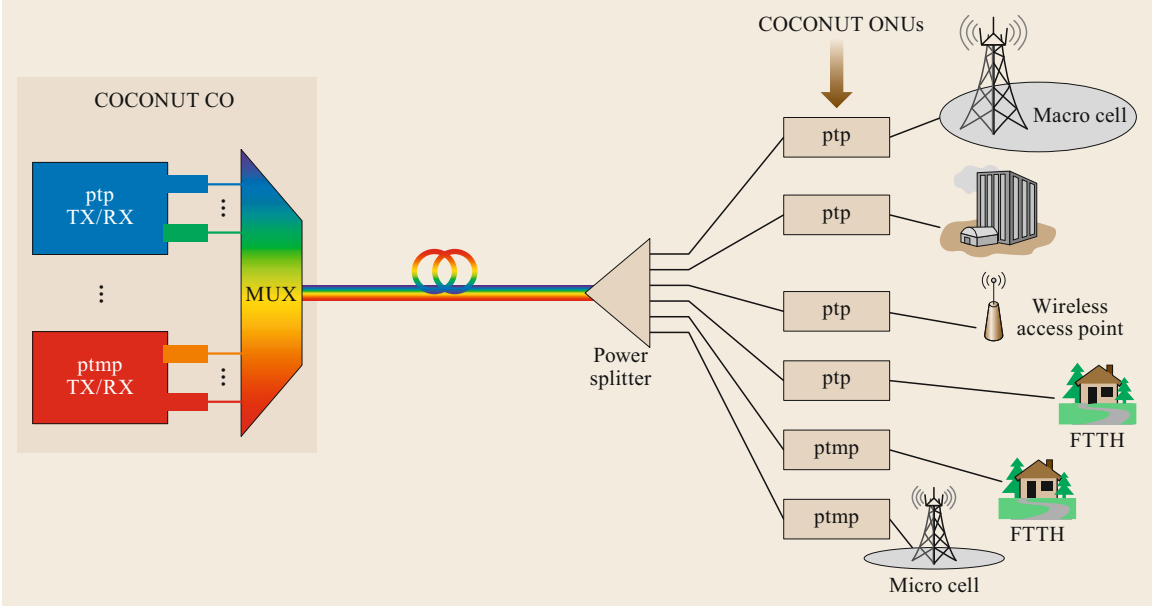


Fig. 28.22 UD-WDM-PON application scenario (*ptp* point to point, *ptmp* point to multipoint, *CO* central office, *TX/RX* transceivers)

and the OLT-ONU control communications signaling system.

Spectrum Management

The most critical elements of the UD-WDM PON are the TX and LO lasers in the ONU and OLT transceivers. Given the cost and simplicity requirements of FTTH, widely tunable lasers (e.g., external cavity or multielectrode DBR lasers) are not considered. On the contrary, common single-electrode single-mode lasers, such as distributed feedback (DFB) lasers, are used. These lasers have very limited, albeit continuous and repeatable, thermal tuning. The main problem to address is the narrow tunability of the ONUs with respect to the defined wide PON spectral band. To address this issue, the optical spectrum organization and transceiver operation can be adapted, as explained below. Given the ultra-dense channel spacing, normal temperature-based tuning (typically between 200 and 400 GHz) permits many, though not all, grid positions to be covered with 6.25 GHz spacing. The fact that users can acquire any ONU at any initial nonpreselected wavelength in the band has been termed *statistical WDM multiplexing* [28.41, 80]. This is where random wavelengths enter the network, and contention may occur if they overlap. In this case, there will be serious interference, which must be avoided by applying anticollision algorithms. Next, the wavelengths of the CO laser adapt to those of the ONUs, which is the reverse of the usual WDM-PON policy for the PON multiplexer and the OLT to set the light waves. Thus, the overall cost of the access network (dominated by the user's equipment) can be minimized. This novel concept implies a radical new management procedure for the optical spectrum and terminals, as well as for activating additional ONUs in the spectrum in a hitless form by combining thermal (slow) and laser current (fast) control [28.79, 81, 82].

The statistical WDM multiplexing concept has been analyzed in terms of the number of wavelength slots required by performing Monte Carlo simulations with 256 active ONUs located in a portion of the C-band with a spacing of 6.25 GHz. Although the minimum wavelength separation between two adjacent channels can be as low as 3 GHz [28.82], it is convenient to leave a few GHz as a guard band. Here, the laser tunability is assumed to be 0.1 nm/°C (typical of DFBs), with a tuning range of 2 nm (250 GHz) upon varying the temperature ±10 °C, which can be easily achieved with a thermoelectric cell or just by heating.

Three heuristic strategies for ONU wavelength assignment were analyzed: first fit (FF), which was used as a basis for comparison; maximum scattering (MS); and maximum admittance (MA) for future requests [28.83]. The FF, in which the closest free channel is selected, is the simplest assignment scheme and requires less initial temperature tuning. The ability to relocate an existing channel when including a new ONU was implemented in a dynamic algorithm (dynamic wavelength allocation (DWA)). As noted from Fig. 28.23, MS exceeds the performance of FF during

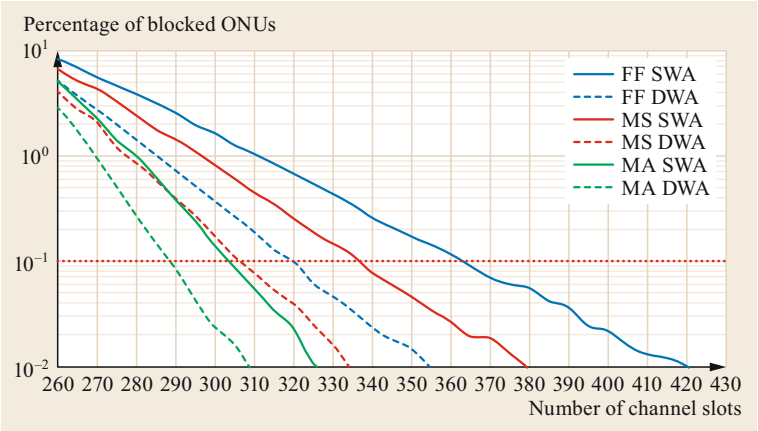


Fig. 28.23 Rejection probability (%) versus optical slots required when using FF, MS, and MA heuristic assignment schemes and static and dynamic algorithms for 256 allocated ONUs

the activation process, but MA provides the best results, with the lowest rejection probability of 10^{-3} (one ONU replacement every 1000 ONUs installed): the 256 active ONUs are successfully allocated over the 288 slots (about 15 nm in the C-band). A channel efficiency of 89% is therefore achieved under the worst conditions, where all ONUs are operating simultaneously. Using the same acceptance ratio, 307 slots are required when MS is employed, and 320 slots are needed when FF is used. Variable environmental conditions and demand evolution were analyzed to ensure that new and existing connections are set up efficiently and flexibly and to minimize the probability of possible conflicts. Outdoor and indoor environmental conditions were considered in the design of activation and reconfiguration algorithms in the TX and LO boards in [28.84], which yielded similar results to when dynamic adaptation was implemented: the channel efficiency decreased to 83% under the worst conditions. This implies that the UD-WDM-PON spectrum with 256 ONUs requires a band of 15.5 nm. In practice, the operational statistics would be even more favorable because the number of connected and active users at a particular time would be much lower than the split ratio. Occasional ONU replacement is best done at installation to minimize any disturbance to the user and operator. OLT laser array dimensioning would not require oversubscription because lasers can be distributed along the spectrum band and operated under controlled thermal conditions. In order to multiplex the 256 channels, a low-loss OLT architecture with two interleaved AWGs was implemented.

The implementation of this strategy in the PON requires a centralized spectrum monitor and manager at the OLT for initiation and dynamic operation under varying conditions, along with an OLT-ONU communication channel for TX/LO control in order to achieve minimum delay and crosstalk between wavelength channels.

Transceiver Developments

Several novel transmission techniques have been proposed and tested. Compared with traditional coherent transceivers, these novel techniques exhibit advanced performance and reduced complexity. Two main aims were considered: the use of direct laser modulation and the development of new receivers to achieve polarization-independent coherent detection in smarter ways.

Direct laser modulation is a convenient solution to attain the target of low-cost access networks. The novel ability to generate a PSK signal with direct DFB laser modulation and coherent detection has been developed. This takes advantage of the natural adiabatic frequency chirp of the laser. The measured phase response (PM) of a DFB laser with a linewidth of 4 MHz ($\Delta\nu$) is shown in Fig. 28.24. From Fig. 28.24, we can see that the PM response closely approximates the expected $1/f$ characteristic of the adiabatic chirp. This is derived from the physical laser chirp model [28.85] (which is dominated by the adiabatic chirp from a few MHz to several GHz),

$$\Delta\Phi(t) = 2\pi \int_{-\infty}^{t} \nu(t)\mathrm{d}\nu \,;$$

$$-\frac{\alpha\kappa}{2} \int_{-\infty}^{t} P(t)\mathrm{d}\nu = -\frac{\alpha\kappa\varepsilon}{2} \int_{-\infty}^{t} I(t)\mathrm{d}\nu \,,$$

where $\nu(t)$ is the instantaneous laser frequency, α is the chirp parameter, κ is the adiabatic chirp constant, and ε is the current-to-power laser efficiency. Using a simple passive derivative pre-equalizer, this frequency chirp can be converted to flat phase chirp and adjusted to 180° phase shifts for PSK. The equalized PM response presents an almost flat phase shift characteristic below the range of interest (approximately < 2.5 GHz). At 1.25 Gb/s (1G ethernet), the residual IM is about

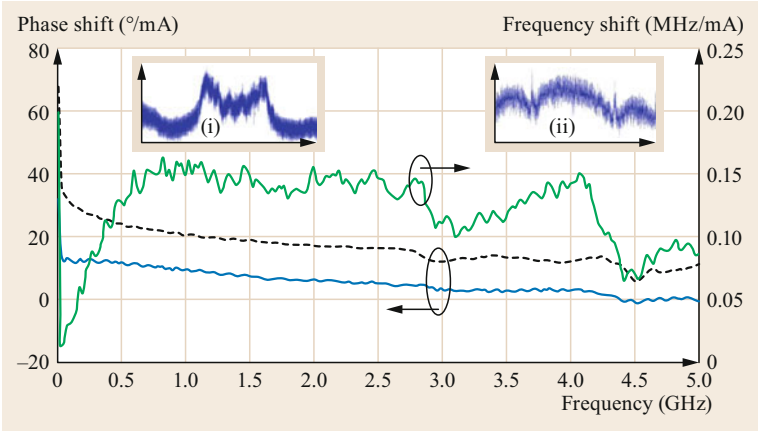


Fig. 28.24 Original (*dashed line*) and equalized PM (*lower solid line*) and FM (*upper solid line*) responses. The *insets* correspond to the optical signal spectrum (*i*) unequalized at 1.25 Gb/s and (*ii*) equalized at 1.25 Gb/s

1.5 dB [28.86], leading to a penalty of only 2 dB with respect to external modulation.

A RX sensitivity of below −50 dBm was obtained with intradyne detection based on a 3×3 coupler at BER $= 10^{-3}$ with offline processing. Recall that the I and Q field components can be obtained from the photocurrents of the three photodiodes using $I = 2I_2 - I_1 - I_3$ and $Q = 3^{1/2}(I_3 - I_1)$. The system attained a channel spacing of only 3 GHz (24 pm) and a negligible phase noise effect [28.82, 87]. When the technique was extended to QPSK with four current levels, 5 Gb/s was reached [28.88]. The transmitter coder and equalizer were recently made digitally configurable [28.89].

The chirp of the DML can also be used to transmit a FSK signal and to convert it to ASK in a coherent receiver; this was done in a 6.25 GHz grid UD-WDM experiment [28.90] at 1.25 Gb/s, yielding a sensitivity of −48 dBm with a 3×3 coupler.

In conventional coherent receivers, polarization diversity is used to address the variation in the polarization state and to double the capacity through polarization multiplexing. However, the cost of the 90° hybrids and eight photodetectors needed makes it difficult to justify this complexity for access networks. A novel highly simplified receiver was therefore proposed [28.91] in which the 3×3 symmetrical coupler configuration was modified to include only three photodiodes, and an intermediate frequency of about half the bit rate was set. An experimental demonstration (Fig. 28.25) indicated that the power penalty with varying polarization was below 1 dB for 1.25 Gb/s [28.92, 93].

An even simpler polarization-independent PON system was also developed that used only one photodiode per receiver. At the OLT side, a common polarization scrambler was utilized that was shared by all ONU wavelengths and switched the polarization state synchronously with the downstream data clock at the half-period of every bit [28.94]. At the RXs, the PSK signals differentially detected from the orthogonal polarizations were electrically combined, as depicted in Fig. 28.26.

To make it more competitive with conventional systems, it would be desirable for the coherent ONU to have only one laser instead of two (TX and LO). This would also make wavelength control of the ultra-dense WDM spectrum easier. Several ways to achieve this have been proposed. The first used heterodyne detection, and the same local laser was modulated with the upstream data, exploiting the chirp of a reflective semiconductor optical amplifier (RSOA). The bidirectional experiment shown in Fig. 28.27 used a length-optimized RSOA to produce PSK data at 1.25 Gb/s, which was received at an intermediate frequency of 2.5 GHz. It achieved a sensitivity of −46 dBm and a channel spacing of less than 12.5 GHz, including downstream and upstream optical signals [28.95]. In [28.96], the intermediate frequency was reduced to only 1 GHz with ASK signaling, resulting in bidirectional interference such as Rayleigh scattering and ODN losses of up to 35 dB. This impairment can be avoided by means of half-duplex transmission [28.97].

An interesting transmitter chip device (denoted DEML) monolithically integrated and synchronously modulated a DFB section and an EA section to produce an ASK-modulated signal for the TX and a CW signal for the RX LO from the front and rear facets of the DEML [28.98]. Other integrated devices were adopted, such as dual-section DFBs, which were employed as pure frequency-modulated lasers (FML) for fast tuning and modulation.

With such narrow channel spacings, an UD-WDM-PON must handle the wavelength drifts of its ONU and OLT lasers, as well as their frequency noise, with great care. Although guard bands of one or a few GHz are left between wavelength channels, the thermal

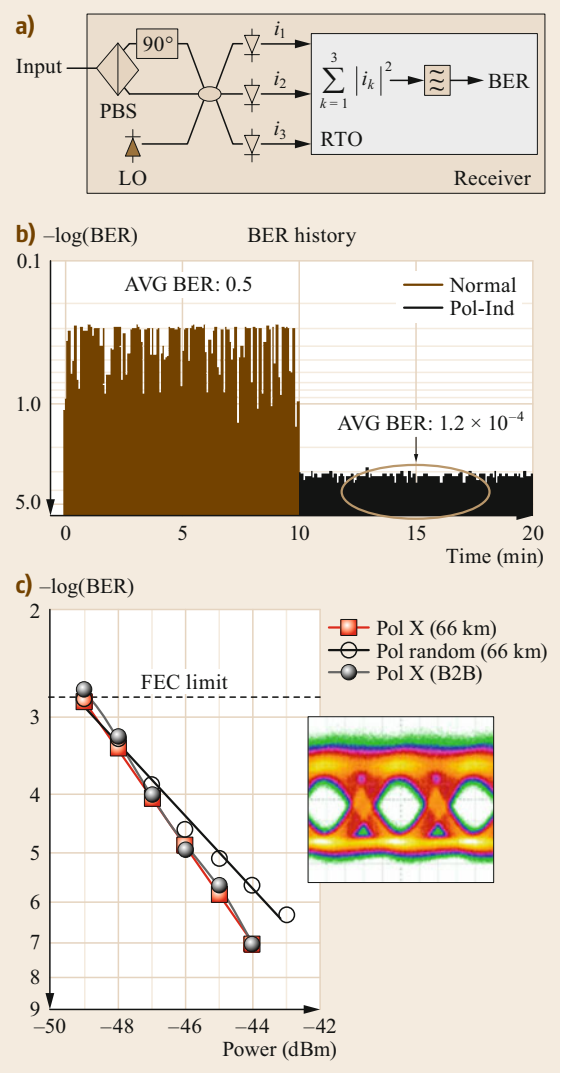


Fig. 28.25 **(a)** Schematic of a polarization-independent ASK receiver. **(b)** BER evolution in a polarization-scrambled signal for a conventional RX and for this RX. **(c)** BER curves for single-polarization and scrambled signals and an eye diagram obtained with real-time implementation

and intrinsic variations of TXs lead to a risk of adjacent channel crosstalk or incorrect coherent operation. The long-term and short-term wavelength variations of closely spaced DFB lasers have been tested. For example, frequency offset measurements of two physically separated free-running lasers, as shown in Fig. 28.27, highlight variations over 30 min of less than 200 MHz when measuring at 1 min intervals, and these variations can slowly increase over longer periods of time. This sets the speed requirements for the automatic frequency control of the LO laser, and for a spectrum monitor in the PON at the OLT.

In the absence of optical frequency compensation, the coherent homodyne detection of DPSK can tolerate a maximum deviation of ±100 MHz, so this detection is rapidly lost. To address this, an automatic wavelength controller (AWC) was embedded that simultaneously controlled the LO temperature and bias current with feedback control. At the same time, a feedforward fast compensation was carried out via real-time DSP processing in a FPGA. With both of these operating simultaneously, a tolerance of ±500 MHz was achieved in 600 ms [28.99] using correlation and Viterbi algorithms (Fig. 28.28). A simplified carrier recovery algorithm used fewer FPGA resources by simultaneously compensating for the detuning and recovering the phase [28.100].

The RX optical hardware was reduced using IQ phase time scrambling [28.101], and the data rates of individual wavelength channels were increased by multilevel modulation and duobinary phase modulation [28.102].

Testbed

The COCONUT UD-WDM-PON was demonstrated in a field trial that took place in the city of Pisa in February 2016. Both analogue and digital real-time processing were performed [28.103]. Several prototypes based on ASK and DPSK modulation formats and heterodyne and intradyne RX were tested. The systems were evaluated in the fiber network deployed in Pisa. ODN losses of up to 40 dB were noted. A HR-OSA (high-resolution optical spectrum analyzer) in the OLT monitored the signals US and DS. With ASK data, the RX sensitivity achieved was −47 dBm at a FEC-level BER. When the ASK signal was scaled to 10 Gb/s and offline post-processing was used, the RX sensitivity was around −36 dBm. For DPSK, the RX sensitivity was −37 and −51 dBm for heterodyne and intradyne RX, respectively. In all of the systems, the TX was based on the direct modulation of intensity or phase, and the RX had simple processing. Multiple videos and internet data (encapsulated in a standard GbE) were also transmitted simultaneously, validating real traffic operation through the network while also showing EPON coexistence.

The project demonstrated that ultra-dense wavelength division multiplexing distributed among distant users could be a feasible alternative to high-bit-rate TDM-PONs for future access networks, as it yields aggregate bandwidths of hundreds of Gb/s and limited energy consumption of the ONU. The feasibility of an aggregate bandwidth of 320 Gb/s (1.25 Gb/s per ONU; 256 ONUs) in only 14.1 nm was demonstrated. The UD-WDM approach solves inherent issues

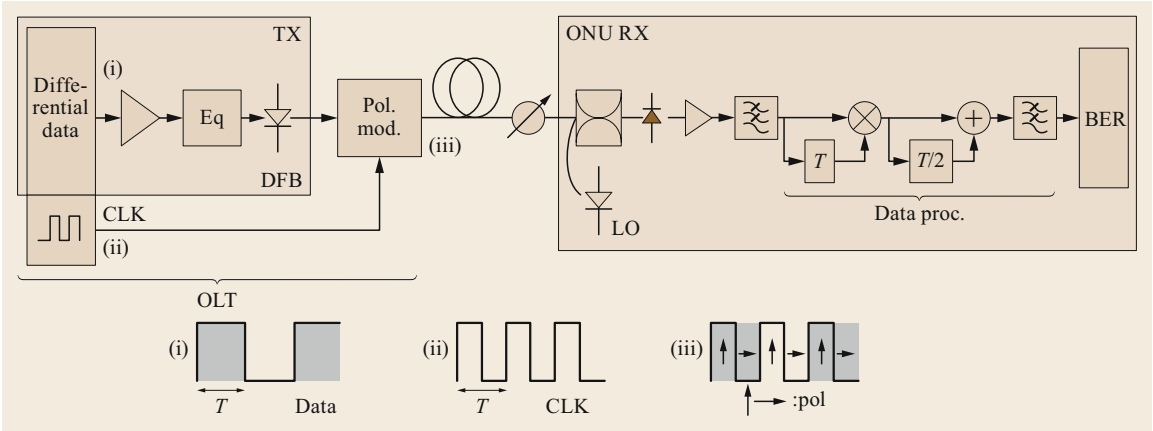


Fig. 28.26 Experimental setup; the *insets* present the signal modulation in a qualitative manner (*Eq* electrical equalizer)

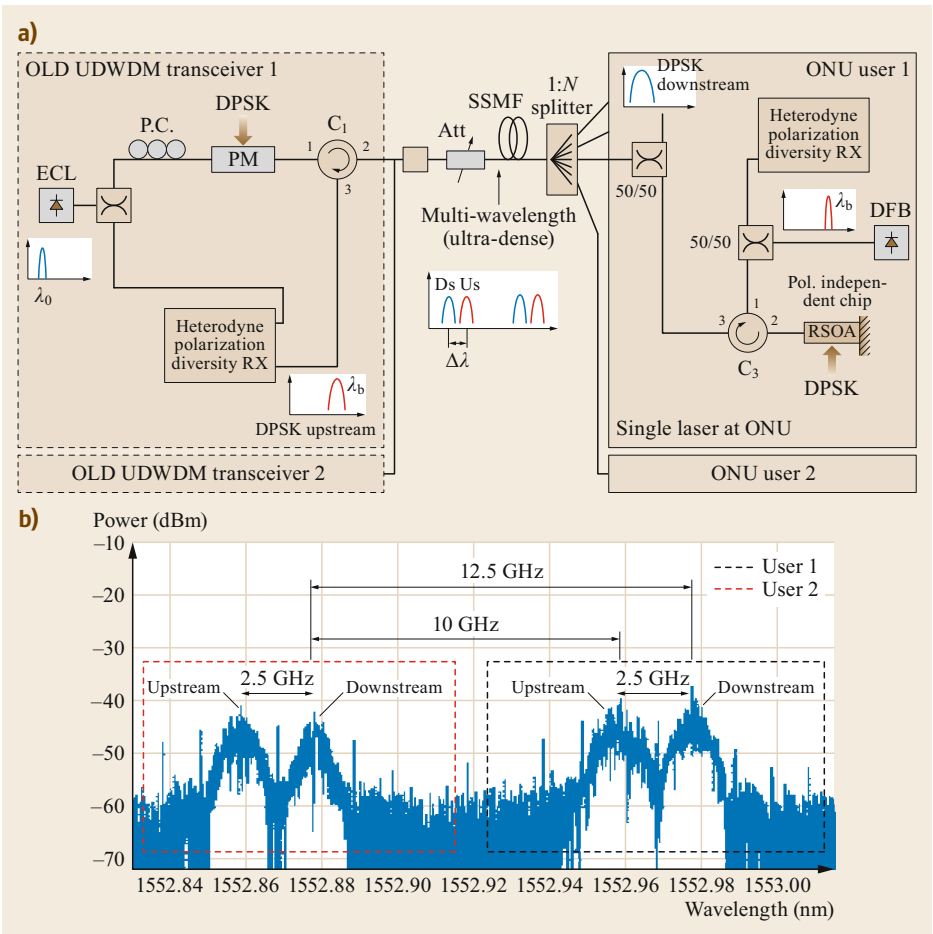


Fig. 28.27 **(a)** Bidirectional subsystem based on heterodyne detection with a single laser at the ONU for UD-WDM-PON using a RSOA chip, and **(b)** the spectrum for two DPSK-DPSK ONUs with bidirectional transmission (*ECL* external cavity laser, *P.C.* polarization controller, *C* circulator)

with NG-PON2, i.e., tunable filtering and laser transient crosstalk. The proposed solution reduces costs by around 80% when compared to a conventional coherent receiver as it eliminates the need for external modulators, external cavity or fully tunable lasers, and 90° hybrids (for example) without incurring a substantial drop in performance. In terms of the migration path, the UD-WDM PON proposal offers an enhanced loss

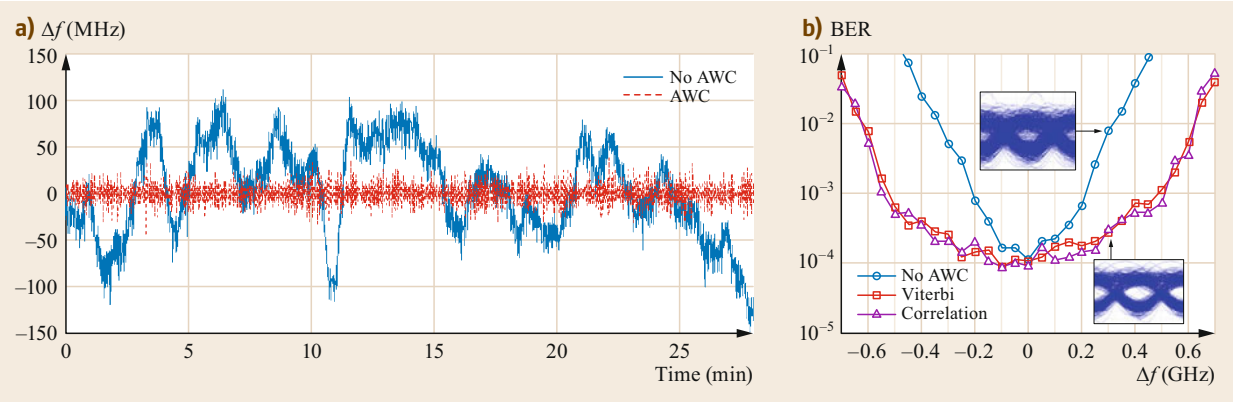


Fig. 28.28a,b Frequency offset measurements performed over 30 min with and without AWC **(a)**, and the dependence of BER on frequency drift for an intradyne receiver with/out data offset compensation, with *inset* eye diagrams at 300 MHz drift **(b)**

budget (allowing for node consolidation), an increase in the number of homes passed, greater reach, and—by using a narrow band in the same fiber—increased spectral efficiency. For this to become a deployable solution, further investigations must be carried out on, for instance, flexible and fast sharing and tuning of the wavelength channels, simplified real-time processing, software-defined transceivers, and integrated optics.

28.5.4 Recent Works

In another project [28.104], an UD-WDM system capable of providing up to 1000 wavelengths and data rates ranging from 150 Mb/s to 10 Gb/s for a reach of up to 100 km in metro and access networks was developed. A *paired channel technology* (*PCT*) generated ultra-dense wavelengths spaced ≈ 3 GHz apart using coherent reception and a tunable laser. A demonstrator with 30 wavelengths (in groups of ten) was built and used as the basis for different experimental setups. A comb generator was used to produce the full capacity of the downstream transmission, 1.194 Tb/s (achieved with 960 wavelengths, each carrying 1.244 Gb/s). In the OLT, a single laser source generated ten independently modulated DQPSK signals at 622 MBd through the use of DSP. In the ONU, a tunable ECL was locked onto the downstream wavelength and served as the upstream signal, via an external modulator, as well as LO for downstream coherent reception. An analysis of the effect of fiber nonlinearity on system performance showed that the main limiting factor was FWM, which limits the downstream power to about −2 dBm per wavelength. When the wavelength count was higher than 32, the impact of FWM remained unchanged. Experimental results on the coexistence of UD-WDM with deployed systems confirmed that the application of DSP-based ONUs in UD-WDM for transmission over 20 km of SSMF is feasible with guard bands of around 1 nm. A more recent work [28.105] reported the demonstration of a dynamic bandwidth and power range optical metro and access network. Transmission scenarios over 80 km of fiber were investigated, as were FSO links, and a field trial was performed. Hybrid DP-QPSK and DP-16QAM modulation was enabled by the format-transparent Stokes space PolDemux at a channel spacing of 6.25 GHz. The performance achieved in the field trial confirmed its value in reconfigurable DSP for flexible optical metro-access networks, as well as optical wireless scenarios.

The authors of [28.106] have made interesting contributions to the development of DSP for UD-WDM. An adaptive equalizer with two reduced-complexity tap weight update algorithms, one of which (sign–sign) was multiplier-free, was identified. These algorithms were tested with a PDM-QPSK signal and found to operate without a sensitivity penalty in simulations and experimentally in loss-limited transmission over 100 km of standard SMF. The channel-selective filter was also optimized, given its importance for maximizing the transmission capacity of the PON and for minimizing crosstalk due to backreflections. A further simplification to the DSP was a blind-adaptive equalizer to simultaneously track the signal's polarization state and apply the matched filter response. Provided a 0.5 dB sensitivity penalty was acceptable, it was found that the matched filter could be omitted when the signal was shaped using a RRC filter with a roll-off factor of 0.1 and a five-tap adaptive equalizer at the receiver. The rate at which polarization rotations could be tracked was determined for each of the tap weight update algorithms considered, allowing a 1 dB sensitivity penalty. Experimentally, for 3 GBd (12 Gb/s) PDM-QPSK signals, the sensitivity of the less complex receiver was found to be −46 dBm.

The practical work reported in [28.107] investigated the possibility of meeting the requirements of a NG-PON2 by employing a simplified coherent detection scheme based on directly modulated (DM) low-cost lasers. In a series of experiments, it was found that it was possible to employ DM-VCSELs as transmitters in a coherent PON, and even VCSELs as LOs. The proposed envelope detection scheme was demonstrated in a digital implementation using offline processing, as well as in an analogue implementation where the nonlinear transfer function of an XOR gate was utilized. Experiments were carried out in the C-band around 1550 nm as well as in the O-band around 1300 nm, yielding a receiver sensitivity of -37.3 dBm at 5 Gb/s. The chirp-enabled envelope detector eliminated the need for phase tracking.

The work described in [28.108], which (as mentioned in the previous section) used OFDM modulation, also aimed to reduce the hardware used in the coherent ONU, as this would bypass the need to use PBS and halve the number of photodiodes at the RX. The works [28.35, 40, 46] discussed in the section on OFDM also involved the use of ultra-dense coherent detection, and can therefore be referenced here too.

A low-cost electronic hardware enabler consisting of an integrated all-analog equalizer for DP-QPSK was recently presented [28.109] as a low-energy alternative to DSP. It was successfully demonstrated at 8 Gb/s.

All these and other relevant works show that there is a practical development path to high-density, high-performance FTTH PONs and xHaul networks in which capacity and transparency are maximized at affordable cost.

Lately, with the commercial advent of ultra-fast coherent systems, and with the expected huge demand in wireless backhaul access, the use of 100 Gb/s coherent transceivers has been proposed and tested by multiplexing eight wavelength channels at a spacing of 50 GHz in a PON with a total capacity of 800 Gb/s [28.110]. The transceivers were standard CPF-DCO (centum form-factor pluggable digital coherent optics) from long-haul networks, but the DSP was simplified for an access domain of 80 km, so its power consumption was reduced.

28.6 OCDMA-PONs

Optical code multiplexing (CDM), or, more specifically, optical code division multiple access (OCDMA), is an attractive form of all-optical multiplexing that has been the focus of ongoing research due to its inherent transparency and asynchronous operation. It utilizes the optical generation, combination, and recognition of optical codes, which can be defined in the time or in the spectral domain. Because of its asynchronous operation, CDM can be applied in heterogeneous network architectures to handle different types of data channels. Moreover, statistical multiplexing can be applied in a straightforward manner to increase the number of channels with bursty traffic. CDM can coexist with TDMA and WDM in sections of the network.

These spectrally encoded CDM approaches are advantageous because the different channels do not have to be synchronized to each other to achieve full orthogonality between codes, in contrast to most time domain encoded CDM approaches. This guarantees the independence of the optical multiplexing scheme from the network architecture. In time-domain CDM, more complex optical codes can be processed but the optical pulses have to be much shorter than the bit period, leading to the need for the components to have greater transmission bandwidth. There have been different proposals on how to implement optical CDM by applying time-domain coding [28.111, 112] or wavelength-domain coding [28.113, 114]. Also, combined WDM and CDM techniques have been proposed in some studies [28.112, 113, 115, 116].

A relevant approach to CDM in the wavelength domain applies periodic spectral encoding of broadband optical sources. This concept, as proposed by *Möller* [28.117], is a generalization of coherence multiplexing in the sense that any type of optical filter with periodic power transmission characteristics can be considered for encoding and decoding. The optical output power of a thermal light-emitting source, such as an LED, is intensity modulated by the electrical data. After the thermal light has been emitted, the broadband spectrum is optically shaped by applying passive filters with periodic transmission functions, such as Mach–Zehnder (MZ) or Fabry–Pérot (FP) filters. The periodicity of the transmission function for these filters is given in terms of the free spectral range (FSR) as defined by the filter round-trip time, which is determined by the filter geometry. This means that the system performance is stable and insensitive to component drift and specifications, making it attractive for MAN and access applications. On the networking side, the benefits are as follows: asynchronous operation of different optical channels; support for different signal formats; and

facile redistribution of electrical bandwidth among optical channels.

A representative setup for a spectrally encoded optical CDM system enhanced by WDM was developed [28.116]. Twenty time-division multiplexed electrical channels with a peak rate of 2.5 Gb/s were collected into five optical CDM channels and multiplexed into four WDM slices 10 nm wide with 3 nm guard bands. Different FSRs were allocated to different optical transmitters to define the codes in the system. At the receiver, an optical filter with a periodic transmission function was tuned by matching its FSR to the desired channel. Taking advantage of the fact that optical CDM can support more codes than can be used simultaneously, the proposed optical packet transmission system can support more than 400 optical channels using statistical multiplexing gain. Using a novel optical TDM-to-CDM conversion technique, optical packet transmission was realized sequentially within each CDM channel, with the average channel bit rate reaching 100 Mb/s. In the latter approach, the optical CDM technology was shown to be insensitive to poor component specifications and drift, thus keeping component and network supervision costs low in the physical layer.

Important research has also been conducted recently on time-domain OCM to exploit its ability to provide multiple access at high rates. In [28.118], an asynchronous full-duplex OCDMA-PON prototype was built as a candidate for NG-PON3. Four-user $\times$40 Gb/s was achieved over 50 km. Code generation and recognition relied on a new superstructured fiber Bragg grating (SSFBG) at the ONU side and on a multiport encoder/decoder at the OLT side. Eight-chip (320 Gchip/s) optical codes were used, with encoding performed in phase at eight levels. In that test, multiple access interference (MAI) and beat noise limited the number of users to four. The use of WDM and improvements to the code efficiency were proposed as solutions to this problem.

Based on these experiences, it can be concluded that optical code multiplexing may provide another route to increasing the number of connected PON users in the future, although it is currently constrained by unsolved practical issues such as high insertion losses and phase errors of the optical CDM elements.

References

28.1 ITU-T Rec. G.989.1: *40 Gigabit-capable passive optical networks (NG-PON2): General requirements* (ITU-T, Geneva 2013)

28.2 Y. Luo, X. Zhou, F. Effenberger, X. Yan, G. Peng, Y. Qian, Y. Ma: Time- and wavelength-division multiplexed passive optical network (TWDM-PON) for next-generation PON stage 2 (NG-PON2), J. Lightwave Technol. **31**(4), 587–593 (2013)

28.3 D. Nesset: NG-PON2 technology and standards, J. Lightwave Technol. **33**(5), 1136–1143 (2015)

28.4 Y. Luo, H. Roberts, K. Grobe, M. Valvo, D. Nesset, K. Asaka, H. Rohde, J. Smith, J.S. Wey, F. Effenberger: Physical layer aspects of NG-PON2 standards – Part 2: System design and technology feasibility, J. Opt. Commun. Netw. **8**(1), 43–52 (2016)

28.5 ITU-T Rec. G.989.2: *40-Gigabit-capable passive optical networks 2 (NG-PON2): Physical media dependent (PMD) layer specification* (ITU-T, Geneva 2014)

28.6 ITU-T Rec. G.989.3: *40-Gigabit-capable passive optical networks (NG-PON2): Transmission convergence layer specification* (ITU-T, Geneva 2015)

28.7 J.S. Wey, D. Nesset, M. Valvo, K. Grobe, H. Roberts, Y. Luo, J. Smith: Physical layer aspects of NG-PON2 standards – Part 1: Optical link design, J. Opt. Commun. Netw. **8**(1), 33–42 (2016)

28.8 K. Asaka: Consideration of tunable components for next-generation passive optical network stage 2, J. Lightwave Technol. **33**(5), 1072–1076 (2015)

28.9 Y. Rao: *InP-based long wavelength VCSEL using high contrast grating*, Ph.D. Thesis (University of California, Berkeley 2012)

28.10 C. Gierl, K. Zogal, S. Paul, F. Küppers: Tunable MEMS-VCSEL with 140-nm tuning range using SiO_2/SiC-based MEMS-DBR, Proc. SPIE (2014), https://doi.org/10.1117/12.2041280

28.11 D.A. Khotimsky: NG-PON2 transmission convergence layer: A tutorial, J. Lightwave Technol. **34**(5), 1424–1432 (2016)

28.12 ITU-T Rec. G.987.3: *10-Gigabit-capable passive optical networks (XG-PON): Transmission convergence (TC) layer specification* (ITU-T, Geneva 2014)

28.13 J. Kani: Power saving techniques and mechanisms for optical access networks systems, J. Lightwave Technol. **31**(4), 563–570 (2013)

28.14 A. Dixit, B. Lannoo, D. Colle, M. Pickavet, P. Demeester: Novel DBA algorithm for energy efficiency in TWDM-PONs. In: *39th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2013), https://doi.org/10.1049/cp.2013.1666

28.15 M.P.I. Dias, D.P. Van, L. Valcarenghi, E. Wong: Energy-efficient framework for time and wavelength division multiplexed passive optical networks (TWDM-PONs), J. Opt. Commun. Netw. **7**(6), 496–504 (2015)

28.16 L. Valcarenghi, Y. Yoshida, A. Maruta, P. Castoldi, K. Kitayama: Impact of energy efficient schemes on virtualized TWDM PONs. In: *Asia Commun. Photonics Conf. (ACP)* (2013), https://doi.org/10.1364/ACPC.2013.AF3G.1

28.17 K. Taguchi, H. Nakamura, K. Asaka, S. Nakano, S. Kimura, N. Yoshimoto: 100-ns λ-selective burst-mode transceiver for 40-km reach symmetric 40-Gbit/s WDM/TDM-PON. In: *39th Eur.*

Conf. Exhib. Opt. Commun. (2013), https://doi.org/10.1049/cp.2013.1328

28.18 K. Asaka: What will be killer devices and components for NG-PON2? In: *40th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2014), https://doi.org/10.1109/ECOC.2014.6964211

28.19 T. Yoshida, S. Kaneko, S. Kimura, N. Yoshimoto: An automatic load-balancing DWBA algorithm considering long-time tuning devices for λ-tunable WDM/TDM-PON. In: *39th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2013), https://doi.org/10.1049/cp.2013.1434

28.20 S. Kaneko, T. Yoshida, S. Furusawa, M. Sarashina, H. Tamai, A. Suzuki, T. Mukojima, S. Kimura, N. Yoshimoto: First system demonstration of hitless λ-tuning sequence for dynamic wavelength allocation in WDM/TDM-PON. In: *Opt. Fiber Commun. Conf. (OFC)* (2014), https://doi.org/10.1364/OFC.2014.W3G.6

28.21 Y. Senoo, S. Kaneko, T. Yoshida, J. Sugawa, T. Odaka, S. Kimura, N. Yoshimoto, H. Kimura: Dynamic-load-balancing algorithm suppressing the number of wavelength reallocations for λ-tunable WDM/TDM-PON. In: *40th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2014), https://doi.org/10.1109/ECOC.2014.6963963

28.22 L. Valcarenghi, P. Castoldi: TWDM PON: How much energy can we really save? In: *16th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2014), https://doi.org/10.1109/ICTON.2015.7193735

28.23 L. Valcarenghi: Energy efficient optical access netwoks and subsystems. In: *OptoElectron. Commun. Conf./Austral. Conf. Opt. Fibre Technol. (OECC/ACOFT)* (2014) pp. 1004–1006

28.24 A. Dixit, B. Lannoo, G. Das, D. Colle, M. Pickavet, P. Demeester: Flexible TDMA/WDMA passive optical network: Energy efficient next-generation optical access solution, Opt. Switch. Netw. **10**(4), 491–506 (2013)

28.25 K. Kondepu, L. Valcarenghi, P. Castoldi: Reconfiguration triggering optimization in TWDM PONs with fluctuating load. In: *Opt. Fiber Commun. Conf. Exhib. (OFC)* (2016)

28.26 M. McGarry, M. Reisslein, M. Maier: WDM ethernet passive optical networks, IEEE Commun. Mag. **44**, 15–22 (2006)

28.27 K. Kondepu, L. Valcarenghi, D.P. Van, P. Castoldi: Trading energy savings and network performance in reconfigurable TWDM-PONs, J. Opt. Commun. Netw. **7**(5), 470–471 (2015)

28.28 L. Valcarenghi, K. Kondepu, P. Castoldi: Reducing delay penalties in energy-efficient TWDM PON through reconfiguration threshold adaptation. In: *Asia Commun. Photonics Conf. (ACP)* (2015), https://doi.org/10.1364/ACPC.2015.ASu1G.3

28.29 Oase: Technical assessment and comparison of next-generation optical access system concepts, Deliverable 4.2.2, https://cordis.europa.eu/docs/projects/cnect/5/249025/080/deliverables/001-OASEWP4D422EAB30032012V20.pdf (2013)

28.30 S. Pachnicke, J. Zhu, M. Lawin, M.H. Eiselt, S. Mayne, B. Quemeneur, D. Sayles, H. Schwuchow, A. Wonfor, P. Marx, M. Fellhofer, P. Neuber, M. Dietrich, M.J. Wale, R.V. Penty, I.H. White, J.-P. Elbers: Tunable WDM-PON system with centralized wavelength control, J. Lightwave Technol. **34**(2), 812–818 (2016)

28.31 J. Zhu, A. Wonfor, S.H. Lee, S. Pachnicke, M. Lawin, R.V. Penty, J.-P. Elbers, R. Cush, M.J. Wale, I.H. White: Athermal colourless C-band optical transmitter system for passive optical networks, J. Lightwave Technol. **32**(22), 3651–3658 (2014)

28.32 E. Wong: Next-generation broadband access networks and technologies, J. Lightwave Technol. **30**(4), 597–608 (2012)

28.33 C.W. Chow, C.H. Yeh, C.H. Wang, F.Y. Shih, Y.M. Lin, S. Chi: Demonstration of high spectral efficient OFDM-QAM long reach passive optical network. In: *34th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2008), https://doi.org/10.1109/ECOC.2008.4729420

28.34 R. Giddings, X.Q. Jin, E. Hugues-Salas, E. Giacoumidis, J.L. Wei, J.M. Tang: Experimental demonstration of a record high 11.25 Gb/s real-time optical OFDM transceiver supporting 25 km SMF end-to-end transmission in simple IMDD systems, Opt. Express **18**(6), 5541–5555 (2010)

28.35 N. Cvijetic, M.F. Huang, E. Ip, Y.K. Huang, D. Qian, T. Wang: 1.2 Tb/s symmetric WDM-OFDMA-PON over 90 km straight SSMF and 1:32 passive split with digitally-selective ONUs and coherent receiver OLT. In: *Opt. Fiber Commun. Conf./Nat. Fiber Optic Eng. Conf. (OFC/NFOEC)* (2011), https://doi.org/10.1364/NFOEC.2011.PDPD7

28.36 B. Charbonnier, A. Lebreton, P. Chanclou, G. Beninca, S. Mexezo, R. Dong, J. LeMasson: Low complexity FDM/FDMA approach for future PON. In: *Opt. Fiber Commun. Conf./Nat. Fiber Optic Eng. Conf. (OFC/NFOEC)* (2013), https://doi.org/10.1364/OFC.2013.OTh3A.7

28.37 European 7th framework programme project ACCORDANCE, 2011–2014. A converged copper-optical-radio OFDMA-based access network with high capacity and flexibility (ACCORDANCE, ICT-248654)

28.38 K. Kanonakis, I. Tomkos, H.-G. Krimmel, F. Schaich, C. Lange, E. Weis, M. Dreschmann, R. Schmogrow, P. Kourtessis, M. Milosavljevic, I. Cano, J. Prat, J.A. Torrijos Gijón: Results from the EU project ACCORDANCE on converged OFDMA-PON networks. In: *15th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2013), https://doi.org/10.1109/ICTON.2013.6602855

28.39 R. Schmogrow, M. Winter, B. Nebendahl, D. Hillerkuss, J. Meyer, M. Dreschmann, M. Huebner, J. Becker, C. Koos, W. Freude, J. Leuthold: 101.5 Gbit/s real-time OFDM transmitter with 16QAM modulated subcarriers. In: *Opt. Fiber Commun. Conf./Nat. Fiber Optic Eng. Conf. (OFC/NFOEC)* (2011), https://doi.org/10.1364/OFC.2011.OWE5

28.40 P.C. Schindler, R. Schmogrow, M. Dreschmann, J. Meyer, I. Tomkos, J. Prat, H.-G. Krimmel, T. Pfeiffer, P. Kourtessis, A. Ludwig, D. Karnick, D. Hillerkuss, J. Becker, C. Koos, W. Freude, J. Leuthold: Colorless FDMA-PON with flexible

bandwidth allocation and colorless, low-speed ONUs, J. Opt. Commun. Netw. **5**(10), A204–A212 (2013)

28.41 I. Cano, M.C. Santos, V. Polo, J. Prat: Dimensioning of OFDMA PON with non-preselected independent ONUs sources and wavelength-control, Opt. Express **20**, 607–613 (2012)

28.42 W. Poehlmann, T. Pfeiffer: Demonstration of wavelength-set division multiplexing for a cost effective PON with up to 80 Gbit/s upstream bandwidth. In: *37th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2011), https://doi.org/10.1364/ECOC.2011.We.9.C.1

28.43 N. Cheng, G. Wei, F. Effenberger: Dynamic spectrum managed passive optical network, IEEE Commun. Mag. **49**, 86–93 (2011)

28.44 P. Chanclou, A. Cui, F. Geilhardt, H. Nakamura, D. Nesset: Network operator requirements for the next generation of optical access networks, IEEE Network **26**, 8–14 (2012)

28.45 I.N. Cano, X. Escayola, P.C. Schindler, M.C. Santos, V. Polo, J. Leuthold, I. Tomkos, J. Prat: Experimental demonstration of a statistical OFDM-PON with multiband ONUs and elastic bandwidth allocation, J. Opt. Commun. Netw. **7**(1), A73–A79 (2015)

28.46 A. Agmon, M. Nazarathy, D.M. Marom, S. Ben-Ezra, A. Tolmachev, R. Killey, P. Bayvel, L. Meder, M. Hübner, W. Mer: OFDM/WDM PON with laserless, colorless 1 Gb/s ONUs based on Si-PIC and slow IC, J. Opt. Commun. **6**(3), 225–237 (2014)

28.47 C. Ruprecht, Y. Chen, D. Fritzsche, J. von Hoyningen-Huene, N. Hanik, E. Weis, D. Breuer, W. Rosenkranz: 37.5-km urban field trial of OFDMA-PON using colorless ONUs with dynamic bandwidth allocation and TCM. In: *Opt. Fiber Commun. Conf. (OFC)* (2014), https://doi.org/10.1364/OFC.2014.Th3G.5

28.48 M.S. Erkilinc, D. Lavery, K. Shi, B.C. Thomsen, P. Bayvel, R.I. Killey, S.J. Savory: Bidirectional symmetric 8x10.7 Gb/s WDM-PON over 108 km installed fiber using low complexity polarization-insensitive coherent ONUs. In: *42th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2016), Paper M1.E.1

28.49 P. Boffi, P. Martelli, P. Parolari, L. Blundetto, J. Morosi, G. Cincotti: Demonstration and performance investigation of hybrid all-optical OFDM systems for optical access network applications, IEEE Photonics J. **7**(1), 7900309 (2015)

28.50 K. Habel, M. Koepp, S. Weide, L. Fernandez, C. Kottke, V. Jungnickel: 100G OFDM-PON for converged 5G networks: From concept to real-time prototype. In: *Opt. Fiber Commun. Conf. (OFC)* (2017), https://doi.org/10.1364/OFC.2017.W1K.4

28.51 A. Saljoghei, A. Farhang, C. Browning, N. Marchetti, L.E. Doyle, L. Barry: Investigation of the performance of GFDMA and OFDMA for spectrally efficient broadband PONs. In: *Opt. Fiber Commun. Conf. Exhib. (OFC)* (2017), https://doi.org/10.1364/OFC.2017.Th2A.25

28.52 R. Hu, C. Li, H. Li, Q. Yang, M. Luo, S. Yu, W. Shieh: Digital OFDM-PON employing binary intensity modulation and direct detection channels. In: *Opt. Fiber Commun. Conf. (OFC)* (2017), https://doi.org/10.1364/OFC.2017.W1K.5

28.53 F.I. Chicharro, B. Ortega, M. de Diego, J. Mora: Reconfigurable optical OFDM signal transmitter based on sliced ASE source for DD MB-OFDM next generation WDM access networks. In: *19th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2017), https://doi.org/10.1109/ICTON.2017.8024887

28.54 A. Lebreton, B. Charbonnier, G. Beninca de Farias, P. Chanclou, R. Dong, J. Le Masson, S. Menezo: Low complexity FDM/FDMA approach for future PON. In: *Opt. Fiber Commun. Conf./Nat. Fiber Optic Eng. Conf. (OFC/NFOEC)* (2013), https://doi.org/10.1364/OFC.2013.OTh3A.7

28.55 A. Gatto, P. Parolari, P. Boffi: Frequency division multiplexing for very high capacity transmissionin bandwidth-limited systems. In: *Opt. Fiber Commun. Conf. (OFC)* (2017), https://doi.org/10.1364/OFC.2017.W1K.1

28.56 Z. Dong, X. Li, J. Yu, Z. Cao, N. Chi: 8 × 9.95-Gb/s ultra-dense WDM-PON on a 12.5-GHz grid with digital pre-equalization, IEEE Photonics Technol. Lett. **25**(2), 194–197 (2013)

28.57 H.K. Shim, H. Mu, U.H. Hong, Y.C. Chung: A practical 10-Gb/s ultra-dense WDM PON. In: *Opto-Electron. Commun. Conf./Aust. Conf. Opt. Fibre Technol. (OECC/ACOFT)* (2014)

28.58 L.G. Kazovsky: Decision-driven phase-locked loop for optical homodyne receivers, J. Lightwave Technol. **3**(6), 1238–1247 (1985)

28.59 S. Norimatsu, K. Iwashita, K. Sato: PSK optical homodyne detection using external cavity laser diodes in costas loop, IEEE Photonics Technol. Lett. **2**(5), 374–376 (1990)

28.60 L.G. Kazovsky: Balanced phase-locked loops for optical homodyne receivers, J. Lightwave Technol. **4**(2), 182–195 (1986)

28.61 S. Camatel, V. Ferrero, P. Poggiolini: 2-PSK homodyne receiver based on a decision driven architecture and a sub-carrier optical PLL. In: *Proc. Opt. Fiber Commun. Conf./Nat. Fiber Optic Eng. Conf. (OFC/NFOEC)* (2006), https://doi.org/10.1109/OFC.2006.215396

28.62 J.M. Fabrega, L. Vilabru, J. Prat: Experimental demonstration of heterodyne phase-locked loop for optical homodyne PSK receivers in PONs. In: *10th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2008), https://doi.org/10.1109/ICTON.2008.4598412

28.63 S. Norimatsu, K. Iwashita: PLL propagation delay-time influence on linewidth requirements of optical PSK homodyne detection, J. Lightwave Technol. **9**(10), 1367–1375 (1991)

28.64 M.G. Taylor: Phase estimation methods for optical coherent detection using digital signal processing, J. Lightwave Technol. **27**, 901–913 (2009)

28.65 R. Noe: Phase noise-tolerant synchronous QPSK/BPSK baseband-type intradyne receiver concept with feedforward carrier recovery, J. Lightwave Technol. **23**(2), 802–808 (2005)

28.66 D. Borne, C.R.S. Fludger, T. Duthel, T. Wuth, E.-D. Schmidt, C. Schulien, E. Gottwald, G.D. Khoe, H. Waardt: Carrier phase esti-

mation for coherent equalization of 43-Gb/s POLMUXNRZ-DQPSK transmission with 10.7-Gb/s NRZ neighbours. In: *33th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2007), https://doi.org/10.1049/ic:20070258

28.67 J.M. Fabrega, J. Prat: Fuzzy logic data estimation based PSK receiver with time-switched phase diversity, Electron. Lett. **42**(16), 940–941 (2006)

28.68 N. Keil, H. Yao, C. Zawadzki, W. Döldissen, W. Schlaak, M. Möhrle, D. Schmidt: Polymer as integration platform for low-cost devices in future optical networks. In: *11th Eur. Conf. Netw. Opt. Commun. (NOC)* (2006)

28.69 J. Prat, J.M. Fabrega: New homodyne receiver with electronic I&Q differential demodulation. In: *31st Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2005), https://doi.org/10.1049/cp:20050683

28.70 J.M. Fabrega, J. Prat: Homodyne receiver prototype with time-switching phase diversity and feedforward analog processing, Opt. Lett. **32**(5), 463–465 (2007)

28.71 J.M. Fabrega, J. Prat: Simple low-cost homodyne PSK receiver. In: *33rd Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2007), https://doi.org/10.1049/ic:20070260

28.72 R. Noe, H.J. Rodler, A. Ebberg, G. Gaukel, B. Noll, J. Wittmann, F. Auracher: Comparison of polarization handling methods in coherent optical systems, J. Lightwave Technol. **9**(10), 1353–1366 (1991)

28.73 J.M. Fabrega, J. Prat: New intradyne receiver with electronic-driven phase and polarization diversity. In: *Opt. Fiber Commun. Conf./Nat. Fiber Optic Eng. Conf. (OFC/NFOEC)* (2006), https://doi.org/10.1109/OFC.2006.215826

28.74 M.O. van Deventer: *Fundamentals of Bidirectional Transmission over a Single Optical Fibre* (Kluwer, Dordrecht 1996)

28.75 M. Omella, J. Lazaro, V. Polo, J. Prat: Driving requirements for wavelength shifting in colorless onu with dual-arm modulator, J. Lightwave Technol. **27**(17), 3912–3918 (2009)

28.76 J. Prat, I. Cano, M. Presi, I. Tomkos, D. Klonidis, G. Vall-llosera, R. Brenot, R. Pous, G. Papastergiou, A. Rafel, E. Ciaramella: Technologies for cost effective udWDM-PONs, J. Lightwave Technol. **34**, 783 (2016)

28.77 G. Vall-llosera: COCONUT requirements for residential, business and outdoor scenarios. In: *15th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2013), https://doi.org/10.1109/ICTON.2013.6602827

28.78 European 7th Framework Programme Project COCONUT: http://www.ict-coconut.eu

28.79 C.N. Ververidis, I. Tomkos, D. Klonidis, A. Rafel, N. Parkin, P. Urban, J. Prat, J. Segarra: Control and management requirements for a coherent ultra-dense WDM PON for lambda to the user access networks. In: *16th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2014), https://doi.org/10.1109/ICTON.2014.6876284

28.80 D. Piehler: PICs in PONs. In: *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC)* (2012), Paper NTu1J.6

28.81 J. Segarra, V. Sales, J. Prat, R. Pous: A new flexible ONU design for UDWDM-PON with coherent transceivers and smart activation process. In: *16th Int. Telecommun. Netw. Strategy Plan. Symp.* (2014), https://doi.org/10.1109/NETWKS.2014.6959241

28.82 V. Polo, P. Borotau, A. Lerín, J. Prat: DFB laser reallocation by thermal wavelength control for statistical udWDM in PONs. In: *Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2014), https://doi.org/10.1109/ECOC.2014.6964080

28.83 V. Sales, J. Segarra, V. Polo, J. Prat: Statistical UDWDM-PONs operating with ONU lasers under limited tunability, IEEE Photonics Technol. Lett. **27**(3), 257–260 (2015)

28.84 V. Sales, J. Segarra, V. Polo, J.C. Velásquez, J. Prat: UDWDM-PON using low-cost coherent transceivers with limited tunability and heuristic DWA, J. Opt. Commun. Netw. **8**(8), 582–599 (2016)

28.85 G.P. Agrawal, N.K. Dutta: *Long-Wavelength Semiconductor Lasers* (Van Nostrand Reinhold, New York 1993)

28.86 I. Cano, A. Lerín, V. Polo, J.A. Tabares, J. Prat: Simple ONU transmitter based on direct-phase modulated DFB laser with heterodyne detection for udWDM-PON. In: *39th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2013), https://doi.org/10.1049/cp.2013.1433

28.87 I. Cano, A. Lerín, V. Polo, J. Prat: Direct phase modulation DFBs for cost-effective ONU transmitter in udWDM PONs, IEEE Photonics Technol. Lett. **26**(10), 973–975 (2014)

28.88 I. Cano, A. Lerín, J. Prat: DQPSK directly phase modulated DFB for flexible coherent udWDM-PONs, IEEE Photonics Technol. Lett. **28**(1), 35–38 (2016)

28.89 J.C. Velásquez, I. Cano, J. Prat: Direct beat phase modulated DFB for flexible 1.25–5 Gb/s coherent UDWDM-PON. In: *Opt. Fiber Commun. Conf. Exhib. (OFC)* (2017), https://doi.org/10.1364/OFC.2017.Th2A.32

28.90 M. Presi, R. Corsini, M. Artiglia, F. Bottoni, G. Cossu, E. Ciaramella: Low-cost 6.25 GHz UDWDM-PON based on direct intensity-modulated transmitters. In: *Opt. Fiber Commun. Conf. Exhib. (OFC)* (2015), https://doi.org/10.1364/OFC.2015.Th3I.1

28.91 E. Ciaramella: Polarization-independent receivers for low-cost coherent OOK systems, IEEE Photonics Technol. Lett. **26**(6), 548–551 (2014)

28.92 M. Presi, R. Corsini, E. Ciaramella: Experimental demonstration of a novel polarization independent coherent receiver for PONs. In: *Opt. Fiber Commun. Conf. (OFC)* (2014), https://doi.org/10.1364/OFC.2014.M2I.2

28.93 F. Bottoni, M. Presi, M. Artiglia, J. Prat, E. Ciaramella: Real-time coherent ONU for λ-to-the-user based on fully analogue processing of OOK signals. In: *Opt. Fiber Commun. Conf. Exhib. (OFC)* (2015), https://doi.org/10.1364/OFC.2015.Th2A.59

28.94 A. Lerín, I. Cano, V. Polo, J. Prat: Polarization independent single-PD coherent ONU receiver

with centralized scrambling in udWDM-PONs. In: *Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2014), https://doi.org/10.1109/ECOC.2014.6963989

28.95 G.Y. Chu, A. Lerín, I. Cano, V. Polo, J.A. Tabares, J. Prat: Exploiting RSOA for uplink transmission with coherent detection for low cost UDWDM-PON. In: *Asia Commun. Photonics Conf. (ACP)* (2014), https://doi.org/10.1364/ACPC.2014.AF2B.12014

28.96 M. Presi, E. Ciaramella: Full-duplex, λ-to-the-user bidirectional PON supporting up to 35 dB optical distribution networks loss. In: *Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2014), https://doi.org/10.1109/ECOC.2014.6963943

28.97 J. Segarra, V. Sales, V. Polo, J. Prat: Half-duplex transmission avoiding Rayleigh backscattering crosstalk in UDWDM-PON with coherent receivers. In: *16th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2014), https://doi.org/10.1109/ICTON.2014.6876283

28.98 G.Y. Chu, I. Cano, V. Polo, C. Kazmierski, R. Brenot, J. Prat: Monolithically integrated dual-output DEML for full duplex DPSK-ASK and DPSK-SSB ONU in ultra-dense channel spaced access network, J. Lightwave Technol. **34**(8), 2042–2048 (2016)

28.99 J. Tabares, V. Polo, I. Cano, J. Prat: Automatic λ-control with offset compensation in DFB intradyne receiver for udWDM-PON, IEEE Photonics Technol. Lett. **27**(4), 443–446 (2015)

28.100 J. Tabares, S. Ghasemi, V. Polo, J. Prat: Simplified carrier recovery for intradyne optical PSK receivers in udWDM-PON, J. Lightwave Technol. **36**(14), 2941–2947 (2018)

28.101 J. Prat, V. Polo, P. Zakynthinos, I. Cano, J.A. Tabares, J.M. Fabrega, D. Klonidis, I. Tomkos: Simple intradyne PSK system for UDWDM-PON, Opt. Express **20**(27), 28758–28763 (2012)

28.102 I. Cano, A. Lerín, M. Presi, V. Polo, E. Ciaramella, J. Prat: 6.25 Gb/s differential duobinary transmission in 2 GHz BW limited direct phase modulated DFB for udWDM-PONs. In: *Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2014), https://doi.org/10.1109/ECOC.2014.6963822

28.103 M. Presi, M. Artiglia, F. Bottoni, M. Ranello, I.N. Cano, J. Tabares, J.C. Velásquez, S. Ghasemi, V. Polo, G.Y. Chu, J. Prat, G. Azcarate, R. Pous, C. Vilá, H. Debregeas, A. Rafel, E. Ciaramella: Field-trial of a high-budget, filterless, λ-to-the-user, UDWDM-PON enabled by an innovative class of low-cost coherent transceivers, J. Lightwave Technol. **35**(23), 5250–5259 (2017)

28.104 H. Rohde, E. Gottwald, A. Teixeira, J. Dias Reis, A. Shahpari, K. Pulverer, J.S. Wey: Coherent ultra dense WDM technology for next generation optical metro and access networks, J. Lightwave Technol. **32**(10), 2041–2052 (2014)

28.105 S. Ziaie, N.J. Muga, R.M. Ferreira, F.P. Guiomar, A. Shahpari, A.L. Teixeira, A.N. Pinto: Adaptive Stokes space based polarization demultiplexing for flexible UDWDM metro-access networks. In: *Opt. Fiber Commun. Conf. (OFC)* (2017), https://doi.org/10.1364/OFC.2017.Th1K.6

28.106 D. Lavery, B.C. Thomsen, P. Bayvel, S.J. Savory: Reduced complexity equalization for coherent long-reach passive optical networks, J. Opt. Commun. Netw. **7**(1), A16–A27 (2015)

28.107 J.B. Jensen, R. Rodes, A. Caballero, N. Cheng, D. Zibar, I. Tafur Monroy: VCSEL based coherent PONs, J. Lightwave Technol. **32**(8), 1423–1433 (2014)

28.108 M.S. Erkılınç, D. Lavery, K. Shi, B.C. Thomsen, P. Bayvel, R.I. Killey, S.J. Savory: Polarization-insensitive single-balanced photodiode coherent receiver for long-reach WDM-PONs, J. Lightwave Technol. **34**(8), 2034–2041 (2016)

28.109 N. Nambath, M. Anghan, N. Thaker, R. Ashok, R. Kamran, A. Kumar Mishra, S. Gupta: First demonstration of an all analog adaptive equalizer for coherent DP-QPSK links. In: *Opt. Fiber Commun. Conf. (OFC)* (2017), https://doi.org/10.1364/OFC.2017.M3D.5

28.110 N. Suzuki, H. Miura, K. Matsuda, R. Matsumoto, K. Motoshima: 100G to 1T based coherent PON technology. In: *Eur. Conf. Exhib. Opt. Commun. (ECOC)* (2017), https://doi.org/10.1109/ECOC.2017.8346032

28.111 A. Grunnet-Jepsen, A.E. Johnson, E.S. Maniloff, T.W. Mossberg, M.J. Munroe, J.N. Sweetser: Demonstration of all-fiber sparse lightwave CDMA based on temporal phase encoding, IEEE Photonics Technol. Lett. **11**, 1283–1285 (1999)

28.112 W. Huang, I. Andonovic: OCDMA/WDMA networks based on OCDMA systems with interference cancellation. In: *25th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (1999) pp. I-192–I-193

28.113 M.J.L. Cahill, G.J. Pendock, D.D. Sampson: Hybrid coherence multiplexing/coarse wavelength-division multiplexing passive optical network for customer access, IEEE Photonics Technol. Lett. **9**, 1032–1034 (1997)

28.114 M. Kavehrad, D. Zaccarin: Optical code-division multiplexed systems based on spectral encoding of noncoherent sources, J. Lightwave Technol. **13**, 534–545 (1995)

28.115 K. Kitayama: OCDM/WDM networks for gigabit access: 1.24 Gbit/s, 2 OCDM by 2 WDM experiment. In: *25th Eur. Conf. Exhib. Opt. Commun. (ECOC)* (1999) pp. I-194–I-195

28.116 T. Pfeiffer, J. Kissing, J.P. Elbers, B. Deppisch, M. Witte, H. Schmuck, E. Voges: Coarse WDM/CDM/TDM concept for optical packet transmission in metropolitan and access networks supporting 400 channels at 2.5 Gb/s peak rate, J. Lightwave Technol. **18**(12), 1928 (2000)

28.117 L. Möller: An optical CDMA method based on periodic spectrum encoding. In: *Proc. 13th Annu. Conf. Eur. Fibre Opt. Commun. Netw. (EFOC)* (1995) pp. 178–181

28.118 R. Matsumoto, T. Kodama, S. Shimizu, R. Nomura, K. Omichi, N. Wada, K.I. Kitayama: 40G-OCDMA-PON system with an asymmetric structure using a single multi-port and sampled SSFBG encoder/decoders, J. Lightwave Technol. **32**(6), 1132–1143 (2014)

Josep Prat

Signal Theory and Communications Dept.
Universitat Politecnica de Catalunya (UPC)
Barcelona, Spain
jprat@tsc.upc.edu

Professor Josep Prat received his PhD from Universitat Politècnica de Catalunya (UPC), Barcelona, in 1995. He is Full Professor in the Optical Communications Group (www.tsc.upc.edu/gco). His research focuses on optical broadband transport and next-generation access networks. He has led the FP7 European project SARDANA (Scalable Advanced Ring-based Passive Dense Access Network Architecture). He has been a guest scientist at UCL and Stanford Universities.

Luca Valcarenghi

Istituto TeCIP
Scuola Superiore Sant'Anna
Pisa, Italy
luca.valcarenghi@sssup.it

Dr Luca Valcarenghi has been an associate professor at the Scuola Superiore Sant'Anna of Pisa, Italy, since 2014. He has published more than 150 papers in international journals and conference proceedings. His main research interests are optical networks design, analysis, and optimization; communication networks reliability; energy efficiency in communications networks; fixed and mobile network integration; and 5G transport. He is an IEEE Senior Member.

29. PON Architecture Enhancements

Thomas Pfeiffer

In this chapter, possible architecture enhancements will be discussed, that allow for applying passive optical networks (PONs) to a wider range of network scenarios than previously considered. The ongoing specification and early deployment of 5G wireless networks and services turns out to be an important driver for the evolution of future optical access technologies and network architectures. The interworking of wireless networks with PONs, which provide for fixed transport, will hence be the starting point for introducing advanced architectures, especially those that can support low-latency requirements as imposed by 5G radio technologies and services (Sect. 29.2). These services are provided over a flexible and versatile (long-reach) PON infrastructure on a metro scale, together with other service types over the same common network. Most hardware functions of such networks will be virtualized on data center platforms, supported by centralized resource orchestration across multiple network segments and technologies (Sect. 29.3). In some scenarios, direct optical links between the end nodes of a PON segment enable lowest transmission latency or offloading high traffic volumes from the main PON link. Sample use cases, as found in wireless and intradata center networks, are discussed in Sect. 29.4. Finally, in Sect. 29.5, optical solutions are introduced that can help in remotely supervising and managing the passive fiber infrastructure, as well as in reconfiguring the connectivity map of complex PON-based metro-access networks, while respecting access operational models and cost targets.

The architectures presented are based on current PON technologies and deployment practices. Most of the modifications described that are required for accommodating advanced functionalities, such as those mentioned before, are either under investigation in research or even under development already. A few are still considered only on the conceptual level.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_29

Unlike in the early days, PONs are nowadays being considered for more than just residential services. More demanding applications, such as wireless backhaul/fronthaul (x-haul), enterprise or intradata center networks benefit from increased capacities and from improved networking capabilities offered by state-of-the-art PON systems and their envisioned successors. Powerful optics allow PONs to transmit over long distances, connecting many nodes in remote places over a single fiber. Certain user plane services as well as wireless fronthaul applications require low-latency transport, frequently at very high data rates. Appropriately designed metro-access networks employing advanced PON technologies on the physical and logical layers, together with distributed computing entities and coordinated resource scheduling across network segments, can meet these requirements at low cost. Aside from providing for pure data transmission, the optical layer in such complex networks also requires simple means for reconfiguration and management that can support the low-cost business cases that PONs were always devised for.

29.1 Background

TDM-PONs (time division multiplexing PON) were initially introduced for providing low-cost broadband optical access for residential users and for small businesses. Accordingly, both the physical layer (PHY) and the media access control layer (MAC) were specified for accommodating these service types under widely varying network conditions, allowing for high or low split factors, as well as for transmission over long or short distances at moderate line rates below 10 Gb/s (broadband PON (BPON) [29.1], gigabit-capable PON (GPON) [29.2], Ethernet PON (EPON) [29.3]). With more recent generations, PONs have evolved into versatile high-capacity systems (XG-PON [29.4], XGS-PON [29.5], 10G-EPON [29.3]) that can be used also for more demanding services up to 10 Gb/s. Other system variants have been specified for operation on multiple wavelength channels (NG-PON2 [29.6], MW-PON [29.7]), and for further increased line rates of 25 and 50 Gb/s [29.3], all of these also supporting longer distances (up to 40 km) than with early generations (up to 20 km). On the networking side, resilience is receiving more attention, e.g., fiber protection [29.8] and network operation [29.9]. Also active optical repeaters in the otherwise passive fiber plant are now in scope [29.2, 10, 11].

These evolutions have turned PONs into attractive low-cost platforms for a large variety of services and applications, such as x-hauling (backhaul or fronthaul) in metro scale wireless or coaxial networks, for residential and business services, or for campus, enterprise, or intradata center networks. TDM-PONs are essentially operating as a distributed switch. So, whenever the logical network architecture is based on a tree topology, as is the case in many Ethernet switched local area networks (LAN), a TDM-PON can be a cost-efficient alternative to an actively switched LAN (passive optical LAN: POL). Wavelength division multiplexing (WDM) offers more degrees of freedom for PON design. The simplified implementation and management of WDM technologies as specified for NG-PON2 networks helps save fibers and offers optical routing to different end nodes at affordable cost [29.6]. Strict network synchronization and timely resource allocation for time-sensitive applications can be accommodated by TDM- and WDM-based PONs. With increasing network complexity and with demanding challenges imposed by services and network operations, the fiber plant needs to be evolved from today's model of simply providing fixed pipes for optical channels towards a flexible and manageable infrastructure that supports the provision of a mix of heterogeneous services over a common cable network.

29.2 Wireless x-Haul over PON

The optical access architectures discussed in this section have been designed to comply with the service and transport requirements of future wireless network architectures based on 4G and 5G technologies. PON systems and architectures, both TDM-PON and WDM-PON, will be considered for backhaul and fronthaul transport links (here generically termed x-haul); the reader is referred to Chap. 31 by Cavaliere and Larrabeiti.

29.2.1 Wireless Services and Radio Access Network Requirements on x-Haul Transport

5G radio networks are set to provide a large variety of user plane services with widely varying requirements on bandwidth and latency. (In the discussion of suitable x-haul solutions, 4G radio systems and services

can in most cases be considered a subset of 5G. The following discussions will hence be focused on 5G radio networks, unless noted otherwise.)

5G User Plane Service Requirements

Usually the services in 5G are coarsely classified as enhanced mobile broadband services (eMBB), and Internet-of-Things (IoT), the latter being further divided into and massive or critical machine-type communication (mMTC, cMTC) and ultra-reliable low-latency communication (URLLC) [29.12].

For eMBB services the main target is to provide very high capacities over the air, while the end-to-end latencies are rather moderate: peak rates per user over the air are envisioned to reach up to 20 Gb/s, and average rates up to multiple 100 Mb/s. End-to-end latencies on the service level will be in the order of 10 ms or more.

For IoT services the main emphasis is on low latency (cMTC) and huge numbers of devices (mMTC), while the transmission capacities can be rather low: peak and average rates per device in the range of multiple Mb/s, even kbit/s. End-to-end latencies are envisioned to be in the range of 1 ms or even less for URLLC, whereas for mMTC up to multiple tens of milliseconds can be tolerated, each depending on the use case.

eMBB services will exhibit sustained but dynamic traffic characteristics at high bit rates, whereas some services of the IoT class may exhibit scarce and bursty traffic profiles at moderate bit rates per device. However, with many devices aggregated on an antenna port, the x-haul link will likely carry a smoothened traffic profile. These applications are currently not widely established, hence the traffic details are yet unknown. It must also be noted that the above service types will not necessarily all be present in the same radio network. Particularly, for operational and security reasons, URLLC services will benefit from being provisioned in dedicated networks separate from eMBB (URLLC services are for industrial applications, autonomous driving, and more, whereas eMBB services are mostly for private user applications).

Radio Access Network Requirements

In 5G wireless networks the formerly compact radio base station will be split into functional entities, some of which can be pooled in a central location, and others will be located closer to or at the antenna site. An initial variant of this centralized radio access network (C-RAN) architecture is adopted already now in some 4G network deployments for reducing implementation costs and for ease of operation [29.13]. In the following, a brief introduction is given to splitting the base station architecture in C-RAN, explaining details as much as needed for an understanding of the x-haul solutions described in subsequent sections. For a more thorough discussion of C-RAN architectures, the reader is referred to Chap. 31.

The functional splits of the base station architecture have generically been defined by 3GPP (Fig. 29.1a) [29.14]. The 3GPP group considered a split into two groups of processing functions, with a fronthaul interface between them as further detailed below. Other groups (e.g., [29.15]) later extended this architecture to functional groups hosted in up to three processing entities: CU, DU, and RU (central, distributed and remote unit or radio unit, respectively) which together form the radio access network (RAN). This C-RAN model will be adopted throughout this chapter.

The processing entities CU, DU, and RU can be all collocated (Fig. 29.1b, row 4G), pairwise collocated (rows 5G (a) and 5G (b)), or all separate (row 5G (c)). They are interconnected via external interfaces F1 and Fx, depending on which split option is considered. 3GPP selected the Option 2 split for high-layer splitting (with the F1 interface for fronthaul) and the Option 6 or Option 7 split for low-layer splitting (with the Fx interface for fronthaul [29.16]). The backhaul link connects the radio core (5GC: 5G core) to the RAN. Finally, the low-layer split between digital processing functionalities and analog RF (radio frequency) antenna electronics is located at the Option 8 split. This split is only considered for 4G networks, with the architecture elements EPC (evolved packet core), BBU (baseband unit), and RRH (remote radio head).

The data transmission for backhaul is typically accomplished over Layer 3 networks, whereas at the F1 interface it will be on Layer 3 or Layer 2, employing IP (Internet Protocol) or Ethernet transport. The transmission data rates at both interfaces are given by the data rates of the user plane services. They change with the amount of traffic for those services and their dynamics over time. The data rate at the F1 interface is higher than for backhaul by at most 1% due to the additional small header inserted for Option 2 splitting [29.14, 17]. The transport latencies required at both interfaces are typically in the range of milliseconds, as given by the user plane services. For more detailed bandwidth and latency considerations, see Sect. 29.2.1.

The data transmission at the low-layer functional splits is accomplished on Layer 2, employing Ethernet transport for Options 6, 7, and 8 (after encapsulation into radio specific transport frames, see Fig. 29.2). In the case of Option 8, also native CPRI transport (common public radio interface [29.18]) can optionally be employed. At Option 6, the data are user plane data with some additional headers, their rates and dynamics being approximately equal to those at backhaul or

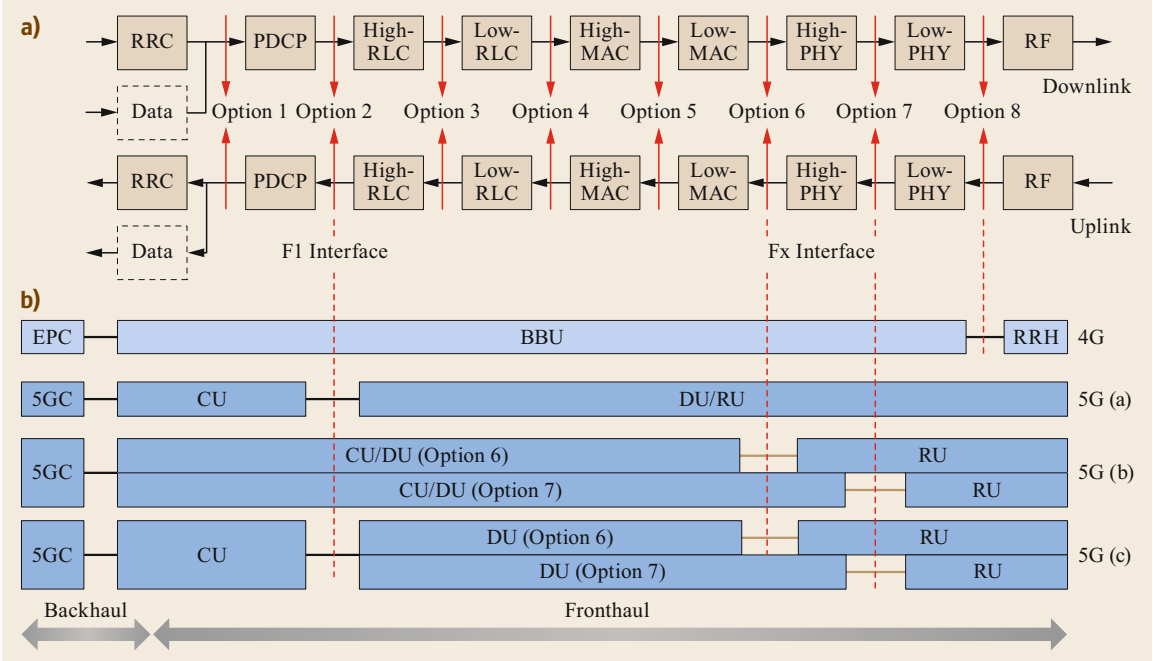


Fig. 29.1 **(a)** Functional splits in the 3GPP base station architecture [29.14]. **(b)** How they map to the BBU/RRH and CU/DU/RU architectures of 4G and 5G, respectively. For 5G, only those functional split options have been used that are specified in detail by 3GPP (after [29.16])

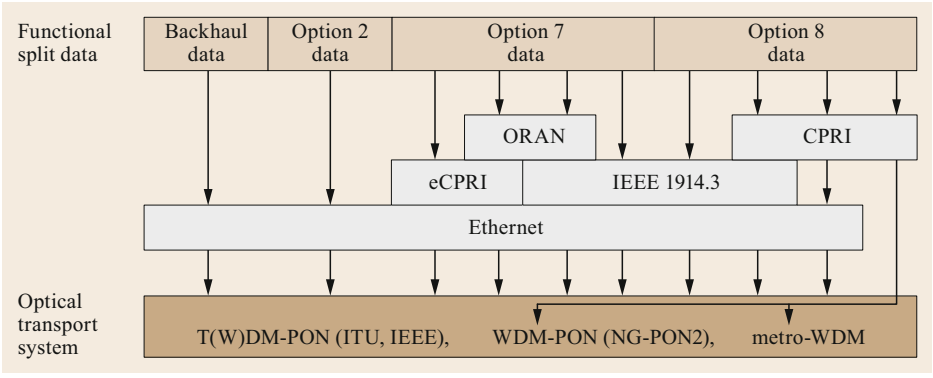


Fig. 29.2 Encapsulation of functional split data for x-haul transport over passive optical systems (Layer 3 and 4 encapsulations are optional, they are not shown in the figure)

at split Option 2. However, the tolerated transport latencies are now given by the requirements of radio technology, requesting them to be in the range of hundreds of microseconds [29.16]. The Option 7 split is the most popular low-layer split in 5G. At this split point the data are digital frequency domain representations of the antenna signals. Their data rate can be higher by up to about an order of magnitude compared to the user plane data rates (i.e., at backhaul or at split Option 2). The exact ratio between the rates depends on the chosen subsplit within Option 7 and on the detailed radio link settings [29.16]. The Option 7 data rate typically varies over time with the user plane data rate. The transport latencies are required to be in the hundreds of microseconds range [29.19]. Finally, the data at Option 8 are time domain data representing the digitized antenna signals. They are constant over time, i.e., they do not vary with the user plane data, but they scale with the bandwidth of the antenna RF spectrum and with the number of antenna elements [29.13]. The data rate is typically in the range of many tens to hundreds of Gb/s, for big antenna arrays they would even reach into the Tbit/s range. The associated implementation effort and equipment cost for optical transport systems supporting such high bit rates is the main reason why this split option is no longer considered for new system deploy-

ments in 5G networks. The latencies are again in the subtilisation range, just like for Option 6 and 7 splits.

Aside from the above bandwidth requirements and latency limitations, also timing accuracy and jitter of the radio signals are important parameters for the network design, hence also for the design of the x-haul transport. They vary from few ten nanoseconds up to a microsecond depending on how the RAN is operated, e.g., when employing one of the different CoMP schemes (cooperative multipoint) [29.20].

29.2.2 Optical System Technologies for x-Haul Transport over Passive Networks

In this section, those optical system technologies will be introduced that have been designed for operation over passive fiber plants in access and metro-access networks. As such, they are possible candidates for cost-efficient x-haul solutions in, e.g., ultra-dense networks (UDN) comprising large numbers of small cells. Focusing in this section only on the optical layer parameters of the optical systems, the interworking between wireless and optical systems will then be discussed in the next section.

Two substantially different system concepts are available for PONs: TDM-PON and WDM-PON. The dynamic bandwidth assignment (DBA) makes TDM-PON particularly useful for networks in which the ONU (optical network unit) traffic load (downstream and/or upstream: US and/or DS) dynamically varies on short time scales down to the order of milliseconds or even below. This allows for leveraging statistical multiplexing gains in networks with many ONUs, thus reducing the required total system capacity below the sum of the individual ONU peak rates. This is the typical mode of operation that TDM-PONs have been designed for. However, there is also another mode of operation, in which each ONU is assigned an individual fixed bit rate, which is independent of varying traffic loads. In this mode, the PON operates on the logical layer like a bundle of dedicated fibers or wavelength channels at fixed capacity. This solution typically constrains the number of connected ONUs as compared to the case of dynamic bandwidth assignment. For moderate rates per ONU, this point-to-multipoint architecture (ptmp) can still be more cost-efficient than a system operating on physical point-to-point channels (ptp) per ONU. For high aggregated transport capacities, however, that would exceed the available TDM-PON line rate, and WDM-PONs or separate ptp-fibers, offering dedicated high-capacity physical channels per ONU, will be needed. Depending on the scenario considered, one or the other solution will be more appropriate. Multiple factors can play a decisive role in selecting the right technology and designing the network architecture: achievable optical line rates, user traffic characteristics, operational considerations such as service or user group segregation, anticipating system migration, and more – all to be considered under constraints on the total cost-of-ownership.

In the context of wireless x-haul, or in the case of heterogeneous networks serving wireless along with other services, both TDM-PON and WDM-PON can be beneficially combined in a common network architecture. NG-PON2 is the first specification that includes both system variants operating on a common optical distribution network (ODN) as part of one combined transmission system [29.6]. The TWDM-PON subnetwork (time/wavelength division multiplexing) of NG-PON2 comprises multiple TDM-PONs, each operating on a separate wavelength channel that is shared among a group of ONUs. Multiple such wavelength channels, currently up to 4 or 8, operate on a 50, 100, or 200 GHz optical frequency grid. The WDM-PON subnetwork comprises multiple densely spaced wavelength channels on a 50 or 100 GHz grid in a separate band. Each channel is optically connected to only one ONU, so that the channel capacity is unshared.

Table 29.1 shows the wavelength ranges specified for the currently available ITU-PON and IEEE-PON (PON systems specified by ITU-T and IEEE, respectively, cf. Table 29.1) generations and their expected successors. Also shown are the frequency ranges specified for G.698.4 compliant passive metro-WDM systems (channel spacing 50 or 100 GHz) [29.21]. In contrast to NG-PON2 networks, the ODN for this system does not contain power splitters but only wavelength multiplexing/demultiplexing devices. The total loss budget is thus constrained to 8 up to 14 dB, i.e., max. 6 dB differential loss. This is to be compared with the low-loss ODN classes L1 and L2 for ptp WDM-PON in NG-PON2, which are specified as 8 to 17 dB (L1) and 16 to 25 dB (L2), i.e., max. 9 dB differential loss. These loss classes were introduced for easing implementation of networks with only short distances and small numbers of ONUs. For regular ODNs in NG-PON2, the loss classes range from max. 29 (N1) to max. 35 dB (E2) with max. 15 dB differential loss.

Several modifications of T(W)DM-PONs are currently under discussion. The serial line rates of state-of-the-art TDM-PON variants are specified up to 10 Gb/s (neglecting line code overhead in IEEE-PONs). The specifications of new generations are being prepared for 25 or 50 Gb/s line rates and above (Chap. 27 by van Veen and Kani). The number of wavelength channels for TWDM-PON is currently limited to max. 8, mostly for reasons of tunable optics cost and of crosstalk issues

Table 29.1 Spectral ranges for ITU-PON (upper table), IEEE-PON (middle table), passive metro-WDM (lower table)

PON generation (DS/US bit rate)	ITU-T document	Wavelength range downstream (DS)	Wavelength range upstream (US)
GPON (2.5G/1.25G)	G.984.5 [29.2]	1480–1500 nm	1260–1360 nm (regular) 1290–1330 nm (reduced) 1300–1320 nm (narrow)
XG-PON (10G/2.5G) XGS-PON (10G/10G)	G.987.2 [29.4] G.9807 [29.5]	1575–1580 nm	1260–1280 nm
25G/50GPON	G.hsp[a]	O-band	O-band
NG-PON2/TWDM (4–8 channels, 10G/10G, 10G/2.5G, 2.5G/2.5G)	G.989.2 [29.6]	1596–1603 nm	1524–1625 nm (wide) 1528–1540 nm (reduced) 1532–1540 nm (narrow)
NG-PON2/WDM (bit rate not specified)	G.989.2 [29.6]	1524–1625 nm (if only WDM ptp is used) 1603–1625 nm (if shared with TWDM ptmp)	
Video distribution	G.983.3 [29.1]	1550–1560 nm	
Maintenance	L.66 [29.22]	1625–1675 nm	
PON generation (DS/US bit rate)	**IEEE document**	**Wavelength range downstream (DS)**	**Wavelength range upstream (US)**
1G EPON (1G/1G)	802.3 [29.3]	1480–1500 nm (PX10, PX20, PX30, PX40)	1260–1360 nm (PX10, PX20, PX30) 1290–1330 nm (PX40)
10G EPON (10G/1G, 10G/10G)	802.3 [29.3]	1575–1580 nm (10G DS)	1260–1280 nm (for 10G US; for 1G US same as 1G EPON)
25G/50G EPON	802.3ca[a]	O-band	O-band
	ITU-T document	**Frequency range upstream (DS)**	**Frequency range downstream (US)**
Passive metro-WDM (10G/10G, 25G/25G[a])	G.698.4 [29.21]	194.05–196.00 THz (lower C-band wavelengths)	191.45–193.40 THz (upper C-band wavelengths)

[a] IEEE 802.3ca will be published in 2020; G.hsp and 25G/25G passive metro-WDM are currently under discussion

in upstream. The latter are caused by the high-power dynamic range due to high differential losses that must be accommodated in power splitter-based ODNs. This, in combination with the spectral characteristics of ONU lasers in burst mode operation, results in challenging component requirements that currently do not allow for high channel counts in such networks under realistic conditions and cost targets [29.23]. Research efforts are ongoing to reduce the impact from crosstalk by, e.g., improving the laser spectrum in burst mode [29.24], eventually allowing for increased channel counts in TWDM-PON.

29.2.3 x-Haul Transport over Passive Optical Networks

The transport of x-haul data over passive optical systems in all cases, except for native CPRI transport [29.18], involves the Ethernet layer as the interface to the transmission system protocol (Fig. 29.2). Backhaul data and Option 2 data are available as IP datagrams and packet data convergence protocol (PDCP) datagrams, respectively. They are encapsulated into Ethernet frames and then forwarded to the transport system. Option 7 data are physical layer data of the radio system. There are multiple ways to forward them to the transport system, involving enhanced CPRI (eCPRI) [29.25] or IEEE 1914.3 [29.26], and/or ORAN [29.27] encapsulation, as shown in the figure. Similar forwarding of data to the transport system is provided for Option 8 data. In addition, in the case of WDM-PON or passive metro-WDM for transport, CPRI frames can also be transported in native format. The choice of the particular protocol sequence depends on the equipment used on the wireless and on the optical side, and on their respective interface specifications. In most practical cases, the connection to the optical transmission system is accomplished via an Ethernet port.

In the following, some use cases for transport over single-channel TDM-PONs will be discussed in more detail. It will be assumed that the connected radio sites are small cells providing for moderate air interface capacities, so that a single optical wavelength channel can aggregate the x-haul data for multiple cells. For the high-layer split architecture, the discussion will be focused on transport capacities, whereas for the low-layer split architectures, the focus will be put on the realization of low-latency transport over TDM-PON.

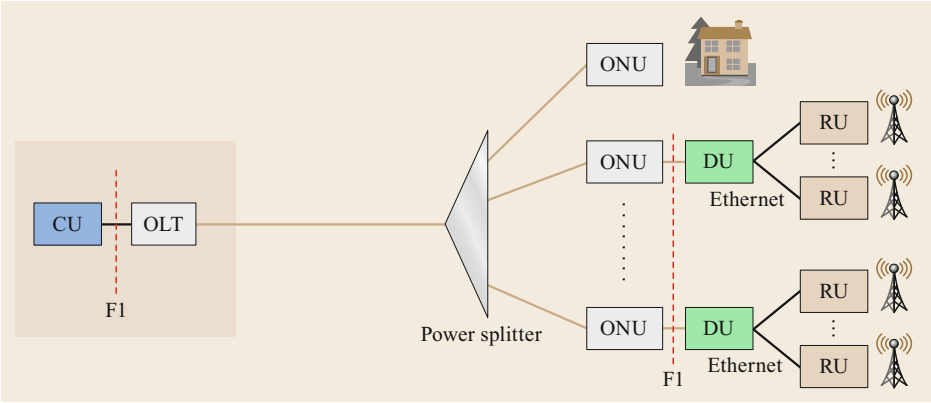


Fig. 29.3 Reference architecture for calculating F1 fronthaul capacities over PON

Nonrealtime Transport at the F1 Interface

The bandwidth and latency requirements of the transport at the F1 interface (Option 2) are similar to those for the backhaul link. The bandwidth is approximately equal to the capacity of the air interface, varying according to the dynamics of the user plane data. The transport is tolerant to latencies in the millisecond range (so-called *nonrealtime transport*). For small cells, with moderate air interface capacities, the traffic dynamics at the F1 interface are, hence, similar to FTTx services (fiber-to-the-home/building etc.). It can be transported by conventional TDM-PON without requiring system modifications.

The network shown in Fig. 29.3 is a reference scenario for analyzing the aggregated transport capacities at the F1 interfaces in a multicell architecture. One or multiple ONUs are each connected to one DU, which in turn serves multiple RUs. The aggregated F1 transport bandwidths have been calculated for 10 RUs being connected to one or multiple DUs (Table 29.2). Various cell configurations have been considered in the analysis: different numbers of MIMO layers (multiple-input-multiple-output), i.e., independent user plane data streams via multiple antenna elements, and various RF bandwidths. For RF bandwidths up to 100 MHz, a carrier frequency below 6 GHz was assumed with modulation at 64 QAM (quadrature amplitude modulation), whereas for RF bandwidths above 100 MHz, a carrier frequency in the mm-wave range (well above 6 GHz) was assumed, being modulated at 256 QAM [29.16].

The traffic variations at each RU are assumed to be smoothed out by aggregation of the traffic for multiple cells. A simple, commonly-used model suggests that the aggregate traffic capacity amounts to 20% of the sum of the individual air interface peak rates at quiet time under perfect radio link conditions. The overhead at the F1 interfaces was accounted for by adding 1% to the air interface rates. Finally, the additional protocol overhead introduced on the transport links was accounted for by adding 20% to the above rates [29.16].

The analysis shows that with 10 RUs, the aggregate transport capacities for almost all cell configurations investigated are within reach of current and emerging TDM-PON systems (10, 25, and 50 Gb/s) by taking advantage of statistical multiplexing. If only moderate configurations are considered, e.g., below 100 MHz RF bandwidth and 8 MIMO layers for carrier frequencies below 6 GHz, then even more RUs can be connected.

(The air interface peak data rates for the considered configurations can be derived from the data in the table as follows: the entries were calculated as $10 \times$ peak rate per RU $\times$ 20% $= 2 \times$ peak rate. So, the peak rate at the air interface per RU amounts to half of the respective entry in the table. For extreme configurations, such as with 800 (200) MHz and 4 (16) MIMO layers,

Table 29.2 Aggregate F1 transport data rates for 10 RU (Fig. 29.3)

MIMO layers	RF bandwidth						
	10 MHz (Mb/s)	20 MHz (Mb/s)	40 MHz (Mb/s)	100 MHz (Mb/s)	200 MHz (Mb/s)	400 MHz (Mb/s)	800 MHz (Mb/s)
1	92	183	366	916	2424	4848	9695
2	183	366	732	1831	4848	9695	19390
4	366	732	1465	3662	9695	19 390	38780
8	732	1465	2930	7324	19 390	38 780	77560
16	1465	2930	5860	14 649	38 780	77 560	155 120

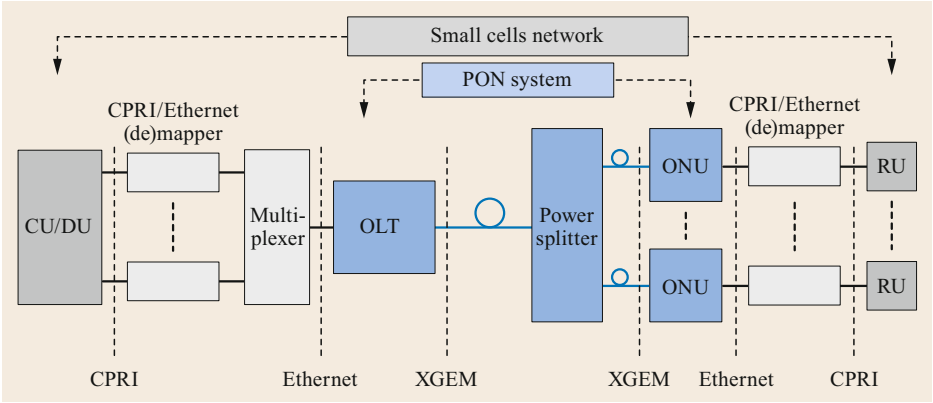


Fig. 29.4 Functional blocks of CPRI fronthaul over XGS-PON

the peak rate reaches almost 20 Gb/s, as envisioned by 3GPP to be the maximum rate in 5G networks.)

In contrast to the assumptions underlying the above calculations, not all RUs in a single PON will necessarily work with the same configuration. Moreover, as indicated in Fig. 29.3, there may be other services on the same PON, such as FTTH. As a result, the dimensioning of the network will be more involved. The above model can still give a realistic impression of which capacities are to be expected when providing F1 fronthaul services for a given number of radio cells.

Realtime Transport at Low-Layer Interfaces

Latency considerations do not play a dominant role in the architecture discussed in the section above. However, at low-layer split points, e.g., Option 7 or Option 8, latency is a critical design parameter of the fronthaul transport. In 4G RAN, the synchronous HARQ (hybrid automated repeat request) process leaves only a few 100 µs for transport (so-called *realtime transport*). In 5G networks both the short slot durations of the air interface data and the URLLC class of latency critical services call for submillisecond transport over fronthaul links.

Integrating small cells into a TDM-PON system is an appealing approach also for low-layer split fronthaul architectures. However, due to the operational principles of TDM-PONs, special precautions must be taken to meet the latency requirements of the radio systems without sacrificing transport capacity. In the following, it will be discussed how this can be accomplished by slightly modifying the way how the TDM-PON is operated in such applications.

Constant Bit Rate, Low-Latency Transmission for an Option 8 Split. In TDM-PON, a constant bit rate service can be set up for an ONU by appropriately setting the bandwidth parameters in the system configurations. A fixed bandwidth pipe alone, however, does not guarantee low-latency and low-jitter transport of the data. Ethernet switches and other logical entities in the PON system equipment (both on the OLT (optical line termination) and the ONU side) will introduce latencies in the microsecond range and jitter, which both strongly depend on the traffic load of the system. In addition, if the fronthaul data are encapsulated in CPRI frames, then additional CPRI-to-Ethernet mappers/demappers must be used to extract the CPRI payload data and encapsulate it into Ethernet frames as specified in IEEE 1914.3 [29.26]. Finally, the Ethernet frames are encapsulated into PON-specific frames, XGEM frames (10G encapsulation method) in the case of XGS-PON (Fig. 29.4). These elements, as well as buffers that are used for jitter reduction, add more latency to the traffic. However, a careful design of the interplay between CPRI, Ethernet, and XGEM streams with respect to their respective frame sizes and relative timing, along with strict mutual synchronization, ensures low latency and jitter for the end-to-end transmission. Using a commercial XGS-PON system, augmented by external CPRI-to-Ethernet (de)mappers, round-trip delays in the range of 110 µs (without fiber propagation delay) have been achieved, as will be detailed next.

Downstream, the data are transmitted over the PON as soon as they arrive from the mapper (assuming that there are no other additional services on the PON), hence the latency contributions are the ones discussed above. Upstream, an additional latency contribution is given by the slot assignments for the optical bursts on the PON. In conventional XGS-PON settings, each ONU is assigned one burst per 125 µs upstream frame, adding a correspondingly high latency. By increasing the scheduling frequency for each ONU to allow for multiple bursts per 125 µs frame, this latency is accordingly reduced. At the same time, the relative amount of protocol overhead will be increased, so that a compromise must be found.

Using commercial XGS-PON and 4G wireless equipment together with suitably adapted CPRI/

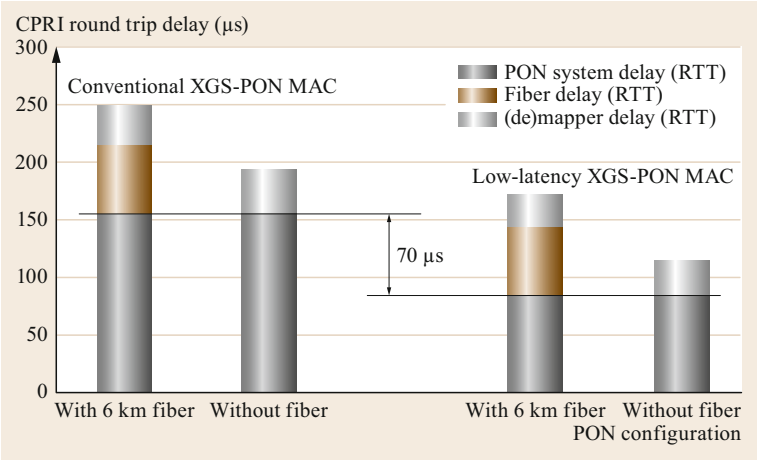


Fig. 29.5 Experimental latency results for XGS-PON with and without fiber link, employing one upstream burst (*left*) and 4 upstream bursts (*right*) per 125 µs frame [29.28]

Ethernet (de)mappers, the round-trip time (RTT) over the PON may be decreased by 70 µs (the difference between the dark gray bars in Fig. 29.5) when assigning 4 evenly spaced bursts to each ONU per 125 µs [29.28]. This leaves sufficient latency budget for transmission over 6 km fiber, which otherwise would not have been feasible within the latency limit of 200 µs set by the wireless system (see the measured round-trip delays in Fig. 29.5, decomposed into individual contributions). This distance matches well with the requirements encountered in UDN deployments in which each macro cell is surrounded by a few nearby small cells (Sect. 29.3.2). The experimental set-up was designed for multiplexing up to 4 CPRI streams at 2.5 Gb/s (2 Gb/s after line code removal) onto the 10 Gb/s downstream and upstream channels of the XGS-PON.

The TDM-PON allows for either combining multiple such CPRI streams for a group of small cells in a given distribution area, or for combining the CPRI stream for, e.g., only one cell together with residential traffic in that area. In the latter case, the small cell would be assigned a fixed fraction of the 125 µs frame at always the same (four evenly spaced) time-slot positions within the frame to minimize jitter, while the residential customers would be dynamically assigned bandwidth capacities in the remainder of the frame. In either case, the conventional process in TDM-PON for activating and ranging new ONUs must be modified to ensure uninterrupted and lowest-latency transmission for the CPRI streams at all times. Appropriate solutions for ONU activation and ranging in latency sensitive service environments will be described below.

Dynamic Bandwidth and Low-Latency Transmission for an Option 7 Split. In contrast to an Option 8 split, the transport data rates at an Option 7 split are significantly lower, and they dynamically change in proportion to the user plane traffic dynamics. However, the latency requirements are as challenging as at Option 8 (the transport requirements for, e.g., the eCPRI protocol are specified in [29.19]). To meet that requirement, the low-latency transport over TDM-PON as described above could be applied also in this case. However, reserving a constant bit-rate channel for this dynamic traffic would be very inefficient, as it would have to be dimensioned to the expected peak data rate, thus wasting transport bandwidth for most of the time.

Instead, a dynamic assignment of transport capacity that can follow the actual interface rate would allow for taking advantage of statistical multiplexing gains, thus improving the overall bandwidth efficiency of the PON channel. In the downstream direction, this does not require special precautions to be taken. However, in the upstream direction the conventional DBA scheme does not support sufficiently fast adaptation to dynamically changing traffic needs at the ONU. The processes of status reporting (of ONU buffer filling) and of traffic monitoring (of utilization of previously assigned ONU bandwidth) [29.6] lead to reaction times in the order of milliseconds.

A cooperative dynamic bandwidth assignment (Co-DBA) can help reduce the transport latency by replacing the reactive with a proactive assignment of ONU upstream bandwidth [29.29]. The CU/DU knows about the capacity that it had scheduled for the UEs (user equipment) served by a certain RU. Hence, it knows how much traffic will be generated at which time slot. In the process of UE scheduling, the respective bandwidth is simultaneously reported to the OLT, along with the associated timing information. The OLT, in turn, assigns to the ONU the appropriate amount of upstream bandwidth ahead of time, thus reducing the end-to-end latency for the wireless data in the uplink (Fig. 29.6).

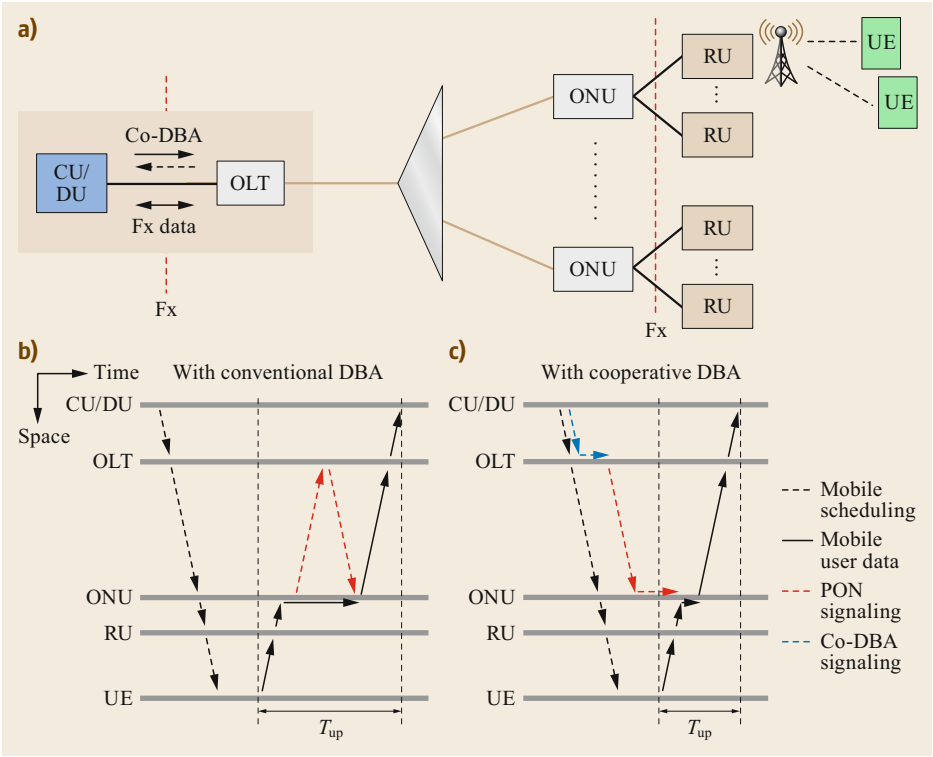


Fig. 29.6a–c Simplified depiction of Co-DBA enhanced Fx-fronthaul over TDM-PON: network architecture **(a)**, uplink time diagram for transport of Fx data when employing conventional DBA **(b)** versus cooperative DBA **(c)**, exemplified by a single UE scheduling; (T_{up} denotes the end-to-end uplink latency from UE to CU/DU)

In the discussion, it is implicitly assumed that the PON bandwidth required by the mobile system is available at any time as needed. In general, this requirement will not be met, especially if the PON system serves multiple RUs or other applications in addition. A mutual exchange of information between CU/DU and OLT about the respectively available system capacities will help optimize the overall performance in terms of achievable latency and bandwidth efficiency (indicated by the dashed Co-DBA arrow in Fig. 29.6).

The Co-DBA concept was proposed already a few years ago [29.29]. However, it is only now that the various scenarios are being explored in detail and that an appropriate interface that enables this cooperation is being specified by the O-RAN Alliance. While the description above focused on low-latency transport at the Fx interface, an equivalent scheme applied at the F1 interface can support end-to-end time sensitive user plane services, e.g., in IoT networks.

Activating and Ranging New ONUs in Low-latency TDM-PON. In the previous two paragraphs it was described how the transport of time-sensitive traffic over TDM-PON, specifically in the upstream direction, can be optimized such that there is minimal delay induced by waiting for the next burst opportunity in the case of fixed upstream bandwidth (cf. CPRI transport at Option 8 split), or by waiting for an update in the case of dynamic upstream bandwidth (cf. eCPRI transport at Option 7 split).

In both cases, there is one more source of delay that originates from a native TDM-PON process, dubbed ONU activation and ranging [29.6]. To let new ONUs join an operational TDM-PON, the OLT occasionally opens a *quiet window* during which no upstream bandwidth is assigned to any of the already activated ONUs. This allows new ONUs in a first (discovery) phase to make themselves known to the system in a random access mode, and then in a second (ranging) phase to measure their distance from the OLT in a coordinated mode (more precisely, the round-trip time between OLT and ONU). After that, they can join the operational system in regular transmission mode. The quiet windows for both phases typically last 250 or 500 µs for differential ODN distances of 20 or 40 km, respectively.

Whenever this process is invoked by the OLT, the upstream traffic for all ONUs is interrupted for the duration of several frames, thus adding an equivalent amount of latency to their budget. In PONs supporting low-latency user plane services or low-layer fronthaul, this process can hence only be applied during system initialization and start-up. During regular system operation, this process must be avoided.

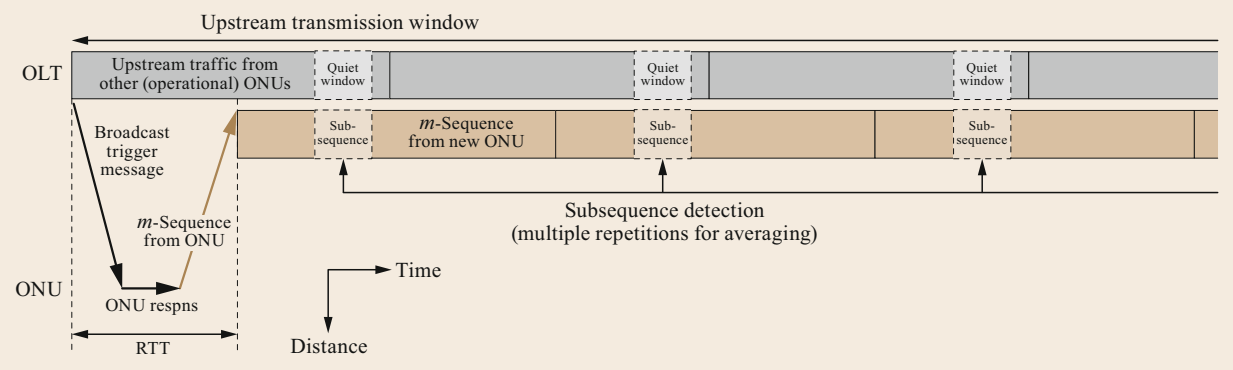


Fig. 29.7 Distance measurement by identifying a subsequence falling into a short quiet window when receiving the ONU's *m*-sequence at the OLT. The ONU starts repeated transmissions of the *m*-sequence after receiving a broadcast message from the OLT that triggers the process. The known position of the quiet window and the position of the identified subsequence within the *m*-sequence allows the OLT to calculate the round-trip time (RTT) between OLT and ONU to within $\pm 1\,\mu s$ precision, which includes the fiber propagation delays and the ONU response time *ONU respns*. The *upstream transmission window* can comprise a single or multiple PON frames, depending on the fiber distance to be measured

There are two classes of options for avoiding the detrimental impact of the ONU activation and ranging process on the latency budget of operational ONUs: use a separate wavelength channel for the above process (out-of-band) or limit the ranging phase to a much shorter quiet window (in-band).

Out-of-Band. The low-latency traffic is serviced on one particular channel without interrupts. The ONU activation and ranging is accomplished on a separate wavelength channel. This can be another wavelength of the same system (e.g., in the case of TWDM-PON) or a separate GPON channel (e.g., if XGS-PON is used for time-sensitive traffic). After completion, the information about the measured distance (round-trip time) is transferred to the OLT operating the time-sensitive channel. In either case, both the OLT and ONUs must be suitably equipped for supporting a dual channel operation: wavelength-tunable optics in the case of TWDM-PON or dual optics (and dual logic) in the case of XGS-PON/GPON. In a mixed-service PON, the ranging channel can also be used to serve other ONUs on the same PON with time-insensitive traffic.

In-Band. If a new ONU is known to the system by its serial number before it tries to join (by preregistration), then the discovery phase can be skipped. If, in addition, the approximate distance from the OLT is known to within a few hundred meters, the OLT requests the potential candidate ONUs to adjust their equalization delay to match the approximate distance before promoting them into the ranging phase, one at a time. This coarse adjustment of the equalization delay then allows for a much shorter quiet window, e.g., $3-5\,\mu s$ for a distance uncertainty of 300–500 m. The time-sensitive traffic is always assigned in the remainder of the frame, so that it will not be affected by the fine ranging process invoked for new ONUs. If this scheme is chosen, then it must be applied also to those ONUs on the same PON that serve time-insensitive traffic. The approximate distance from the OLT can be extracted from ODN deployment data showing cable routes and, e.g., the position of a street cabinet close to the new ONU location. Alternatively, it can be measured by an optical time-domain reflectometry (OTDR) measurement, possibly only at the time of the ONU implementation.

In a further in-band variant, the new ONUs emit a weak binary bit sequence that allows the OLT to measure the distance by pattern identification. Upon receipt of the downstream frame, a new ONU repeatedly emits a previously defined bit sequence to the OLT receiver (Fig. 29.7). In the case of *m*-sequences, the specific subsequence falling into the short quiet window enables determining the round-trip time between OLT and ONU. To avoid distortions of the ongoing transmission from the operational ONUs, the bit sequence is emitted at ultra-low optical powers (max. −50 dBm at the OLT receiver), which in turn requires transmission at low bit rate (10 Mb/s) and multiple averages (10^4) at the receiver for unambiguously identifying the weak signal during the quiet window. After another iteration of this process, now with random delays autonomously added by the ONUs, the OLT can single out one ONU (if multiple ONUs have tried to join) and promote it to the next phase, ranging it within the short quiet window, as described above. Unlike with the above solution, the ONU ID in this approach does not need to be known to

the OLT to enter the second phase; potential collisions can be resolved without knowing it. The serial number is communicated (discovery) only in the second (ranging) phase. This entire process is accomplished with the same optics used for transmission of the time-sensitive traffic after finishing activation and ranging [29.30].

29.3 Flexible Converged Metro-Access Networks Based on PON

In the past, the node equipment in telecommunication networks was specially designed to realize a certain network function, e.g., the OLT function in PON. It contained more than just the optical ports: data switching and routing, bandwidth scheduling, traffic management, and more functions are included in conventional OLT nodes by design. In most cases, this approach does not easily support system upgrades or rapid introduction of new services.

29.3.1 Reconfigurable Computing and Transport Resources in Metro-Access Networks

To allow for more flexibility during the life cycle of a network, some market players have kicked off initiatives to modify node designs that can leverage data center architectures and virtualization on general-purpose computing platforms, for example, the central office rearchitectured as data center (CORD) [29.31], the flexible access system architecture (FASA) [29.32], or the open network automation platform (ONAP) [29.33]. These network designs build on software-defined networking (SDN) concepts in which the control plane is separated from the forwarding plane. The utilization of so-called white box switches allows for lower cost implementation of network functions by virtualization (network function virtualization, NFV). They can be remotely reconfigured, thus easing modification of user services and the rapid introduction of new services during network operation. (It must be noted, however, that not all network functions can be virtualized. Time-sensitive physical layer functions, such as, e.g., the bandwidth mapper in PON-DBA, still benefit from implementation on dedicated hardware.) This SDN/NFV approach has been proposed for all network segments and services. It has meanwhile been adopted throughout the industry as the new paradigm for future network architectures [29.34].

The previously introduced CU and DU functionalities in wireless networks are supposed to work on general-purpose servers in data center architectures where processing capacities can be shared among multiple entities. The RU functionality, instead, includes analog radio hardware as well as digital functions (in the case of Option 6 or 7 split) and is located at a remote radio site. In large networks containing a few large and many small data center-like nodes, it can be beneficial to dynamically distribute the digital processing functions in the form of virtual containers among a group of computing nodes, in response to varying traffic loads in the network and depending on the availability of computing resources.

On the transport side, there are multiple dimensions to the optical resources in PONs, for example, the time domain (slots assigned in a TDM-PON channel), the wavelength domain (as a shared channel in a TWDM-PON, or as an unshared WDM channel in a WDM-PON), and the space domain (fiber). An end-to-end resource management on the system layer and on the fiber layer helps ensure availability and timely coordination of the capacities offered by the systems (via DBA and DWA processes (dynamic wavelength assignment)) and by the ODN (by reconfiguration of fiber routes in the branching nodes).

A properly coordinated management of computing and transport capacities finally allows for a most efficient exploitation of the network's capacities and guarantees an optimized end-to-end performance also for challenging services. A centralized orchestrator provides for the necessary coordination, while also enabling network slicing, i.e., establishing end-to-end virtual networks pertaining to different mutually independent operators, clients, or services.

29.3.2 PON-Based Metro-Access Network with Coordinated Resource Allocation

The network in Fig. 29.8 is an example of a converged metro-access network with data centers hosting virtualized network functions (VNF) and PON systems for optical transport over (nominally) passive ODNs.

The network contains a main data center (edge cloud, preferably duplicated for redundancy, not shown here) for providing services to residential and mobile users, to business and enterprise customers over fixed lines and over wireless connections. The fixed line connections are supposed to be fiber links all the way to the end users and the wireless connections to be based on 5G technologies.

In general, there can be also copper-based fixed line services, either via DSL (digital subscriber line) over

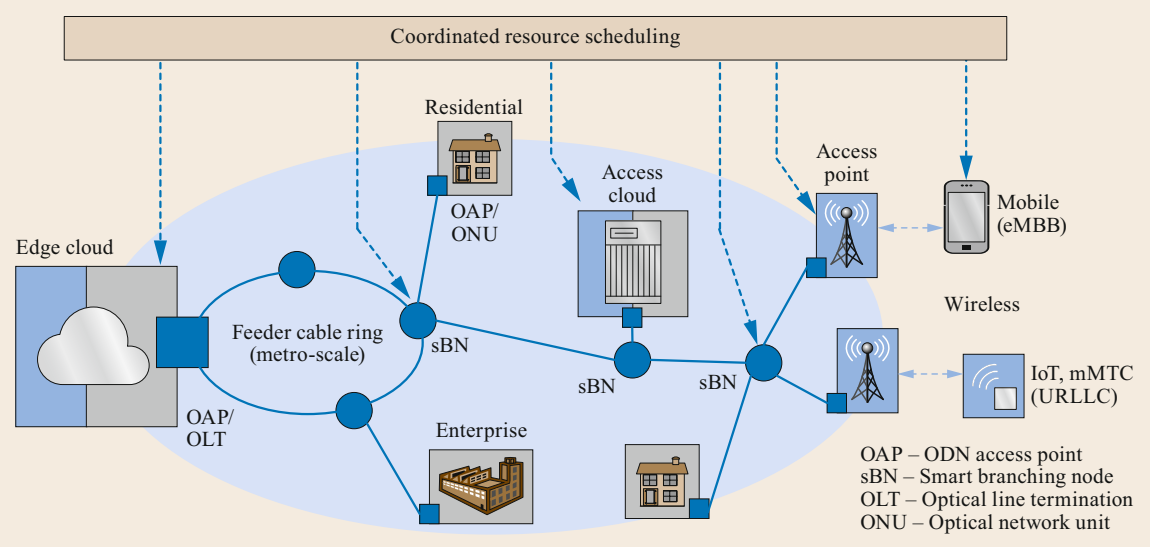


Fig. 29.8 End-to-end metro-access network employing PON system technologies for transport and reconfigurable computing hardware and cable infrastructures

a twisted-pair wire or via DOCSIS over a coaxial cable (data over cable service interface specification), as well as other wireless services using 3G and 4G, or WiFi technologies. In the following discussion, however, these technologies are not explicitly addressed. They can be considered – just as for the 5G services – to be implemented in a hybrid architecture using PON for x-haul from the edge cloud or from the access cloud (see later) to remote nodes from where the last meters are bridged via those systems [29.34].

Besides the central edge cloud, there are multiple access clouds in the network serving smaller geographical areas, while meeting two essential objectives: low-latency connections to the end points and offloading computing and transport resources from the metro scale network for services that are predominantly needed within a constrained area. Finally, there are many (radio) access points connected to the edge cloud or to the nearest access cloud. For wireless services, the functions of CU, DU, and RU can be distributed over two or three of those entities, following the 5G architecture described in the previous section.

The optical connections from the edge cloud or from the access cloud to the end customers (or to the radio access points) are established using PON systems. (Note: the feeder cable ring in Fig. 29.8 is not shown to scale. In fact, it extends over metro distances, and its circumference can be as long as 100 km.) The PON system technologies (short-reach, long-reach, ptmp TWDM, or ptp WDM, small or large wavelength channel count) are chosen according to the respective distance, capacity, traffic characteristics, as well as operational and business driven requirements, such as segregation of customer groups, of services, and more. The end points of the fiber links (indicated in Fig. 29.8 as OAP: ODN access point), as well as the branching nodes (smart branching node: sBN) are supposed to be passive on the data plane, i.e., they do not process user data, nor do they interact with the control plane of the optical transport systems. However, they are active on the physical layer by providing means for monitoring optical parameters, switching fibers or wavelength channels, reconfiguring optical connection patterns, and more (Sects. 29.4 and 29.5).

The network resources are generally orchestrated in a coordinated way across the transmission and computing entities involved (Fig. 29.9) to accommodate differentiated needs of services and customers. However, not every service type needs strict coordination of resources to meet the agreed-upon service level agreement (SLA). A wider range of transport bandwidths, computing resources, and end-to-end latency may be tolerated for some services. Others will need a very tight coordination of bandwidth across segments at a certain time, such as discussed in Sect. 29.2.3. Although in this latter example, the bandwidth assignment is implemented close to the physical transport hardware (not implemented as VNF), it still needs support from an end-to-end resource scheduling.

By implementing this centrally coordinated end-to-end resource management across multiple segments and technologies, the network can also be flexibly partitioned to provision separate slices for multiple tenants and for differentiated services sharing a common network infrastructure.

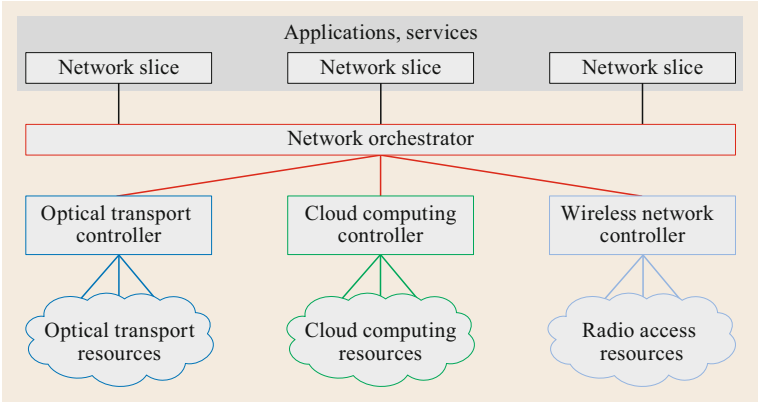


Fig. 29.9 Central orchestration of resources across network technologies and slices

29.3.3 Segments of the Optical Metro-Access Network

On the optical transport layer, the metro-access network in Fig. 29.8 comprises three major segments on different distance scales:

- Metro scale segment
  Long-reach fiber links on the metro-scale can extend over multiple tens of kilometers, up to about 100 km. They connect the access clouds to the edge cloud using high-capacity ptp channels or provide for direct ptp or ptmp links from the edge cloud to end points that can tolerate latencies in the millisecond range (e.g., residential services or wireless x-haul at the F1 interface for eMBB services). As another example, they can provide for direct interconnects between enterprise locations (e.g., branch offices of a bank or manufacturing sites). They are typically deployed in fiber cable rings from which the fibers branch off at sBNs into the access areas where they are further split for serving nodes in a ptmp topology.
- Access scale segment
  These are links of up to 20 km, starting from the access cloud and providing connectivity for latency sensitive applications and services, such as for fronthaul at the Fx interface or for URLLC services. This segment, including the access cloud, can also offload traffic from the large metro-scale network and data processing from the edge cloud for locally constrained services.
- Local scale segment
  With distances as short as 1 km or less, these links provide for short paths between neighboring end nodes to enable, e.g., cooperative data processing in wireless CoMP (cooperative multipoint) architectures, offering shortest latencies and very high capacities without burdening the links on the access or metro-scale segments. No additional fibers will be deployed for these short links. Instead, the communication will be accomplished over the drop section of the access or metro-scale ODNs deployed, i.e., by reusing the fibers connecting the end nodes to the splitters (Sect. 29.4).

The following two examples will display architecture variants in the access segment that are dedicated to accommodate special service requirements.

In Fig. 29.10, it is assumed that the access cloud provides residential and wireless services over a common PON (ONUs are not explicitly shown in the figure, for the sake of simplicity). In the distribution area, there may be one macro cell plus a few small cells, while the other end nodes are being used for FTTx services. It is further assumed that the RUs at the radio sites are connected to the CU/DU in the access cloud via Fx fronthaul links employing separate wavelength channels: an unshared ptp WDM channel for the macro cell and one or more shared ptmp TWDM channels for the small cells. The FTTx users are served via additional ptmp TWDM channels. Due to the latency constraints of the Fx fronthaul links, the distance between access cloud and distribution area will be limited to about 10 km (Fig. 29.10a). If, however, the DU functionality is located at the macro cell site, then it can be connected to the CU in a distant edge cloud via the F1 interface (assuming that no time-sensitive user plane services are involved). In the architecture shown in Fig. 29.10b, the macro cell site receives the F1 data for both its own cell as well as for the neighboring small cells. The DU at the macro cell site then connects to its own RU and to the RUs at the small cell sites via the Fx interface. The small cells are connected to the macro cell site over short links, employing a separate ptmp PON transmission system from the OLT at the macro cell to ONUs at the small cells. It operates at a different wavelength compared to the system connecting the macro site to

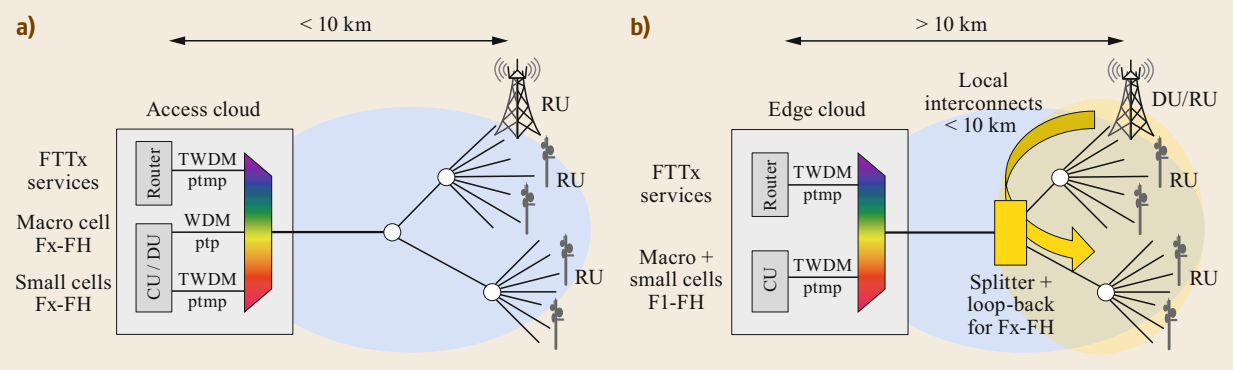


Fig. 29.10a,b The Fx fronthaul architecture limits the distance of a wireless distribution area from the CU/DU location (in the access cloud) to about 10 km **(a)**. For larger than 10 km distances the macro site is connected via the F1 interface to the CU location (now in the edge cloud) and provides for the DU functionality also for neighboring small cells (local C-RAN) **(b)**

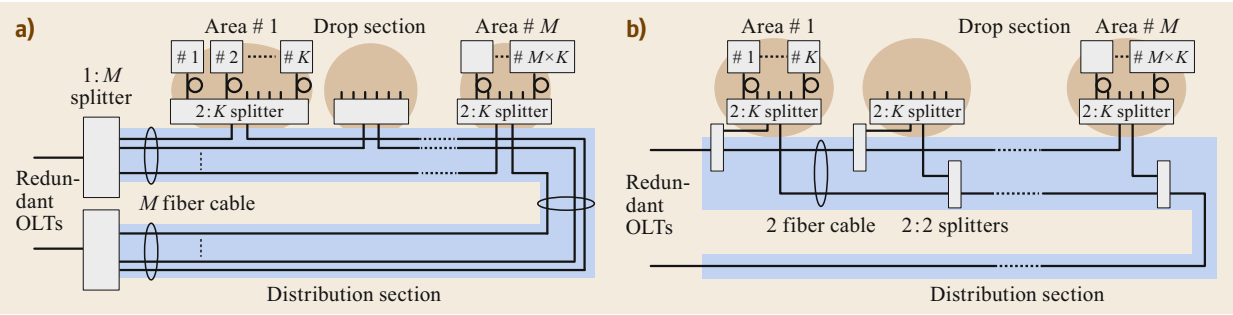


Fig. 29.11a,b Redundant fiber architecture in the distribution section of a PON. **(a)** Using a multifiber cable after the first splitter to realize a redundant tree architecture. **(b)** Using a two-fiber cable to realize a redundant bus architecture

the edge cloud and reuses the drop fibers of the PON for local loop back via the (now modified) power splitter. Further details of the implementation for this local C-RAN scenario are described in Sect. 29.4.

Network resilience is usually not a major concern in residential access. However, with critical services for enterprises, or for x-haul links to wireless nodes, a redundant network design may be needed. There is a trade-off between increased cost for deploying redundant network resources and the implications of customer dissatisfaction in the case of network outages.

Possible protection scenarios for PON systems are described in [29.8] and [29.35] (see also the discussion in Chap. 27). However, since the deployment of redundant fiber links in the drop section is frequently deemed too expensive, only Type B protection has practically been accepted by operators so far. This protection type employs two geographically disjoint feeder fiber routes to the first splitter, from where the unprotected distribution and drop sections start. On the metro scale this redundant feeder is deployed as part of the cable ring, as shown in Fig. 29.8. However, there are also ways to provide fiber protection in the distribution and drop section of access scale PONs at an effort that can be acceptable for certain network scenarios, such as x-haul for small cells.

In the architecture shown in Fig. 29.11a, the two $1:M$ splitters on the left are connected to two OLTs located in different access clouds (not shown). The cable in the distribution section contains M fiber strands, each of which connects a port of the first splitter to one input port of a $2:K$ power splitter located at the drop areas #1 to #M. The second input port of the $2:K$ splitter is connected to the second $1:M$ splitter using the remainder of the same fiber strand, as shown in Fig. 29.11. All ONUs in this architecture are on the same *split level*. The number of drop areas is limited, via the total split factor $N = K \times M$, by the power budget of the PON system used. Current systems support up to 29–35 dB ODN loss, depending on the equipment class chosen (cf. PON specifications in Sect. 29.2.2).

In the architecture shown in Fig. 29.11b, the drop areas are subtended from a single fiber strand (two strands near the drop areas) utilizing cascaded $2:2$ power splitters. The splitting losses in this architecture increase when going down the cascade in either direction, result-

ing in lower losses for the ONUs in the, respectively, first area as compared to those in the, respectively, last area. The number of drop areas in this architecture is limited by the allowed differential path loss of PONs, which in most cases amounts to 15 dB (cf. PON specifications in Sect. 29.2.2). This allows for six cascaded drop areas under realistic conditions. With splitters having asymmetric split ratios that change along the cascade in suitable ways, the number of drop areas can be increased.

29.4 Local Interconnects in Point-to-Multipoint Networks

29.4.1 Use Cases

PONs are used to connect multiple end nodes to a central point, which in turn usually provides for connection to higher network hierarchy levels, such as connecting FTTH customers to the metro network. There are, however, use cases in which the ONUs in a ptmp network need to communicate with each other, in addition to their regular communication with the OLT. In some cases, these inter-ONU links pose challenging requirements on low-latency or high-transmission capacity that are hard to meet, if the ONUs communicate with each other via a loop-back through the OLT. These use cases call for separate optical channels over the shortest possible links between the ONUs, however, without deploying additional fibers. Such use cases have been identified in wireless fronthaul architectures (Sect. 29.3), others can be encountered, e.g., in intradata center networks.

Generic Use Cases

Before showing how such direct optical links between ONUs on the same PON can be implemented on an existing ODN, some specific use cases will be introduced to show how different the requirements on the interconnection architecture can be. Suitable optical solutions will accordingly be different; they will then be shown in the remainder of this section.

First consider the generic target topology in Fig. 29.12 and how it can be mapped onto a ptmp-ODN by reusing the regular drop fibers also for inter-ONU links. The solid red lines in Fig. 29.12a indicate the logical links from a core node to a group of end nodes in a star topology. The dashed blue lines indicate logical links between certain neighboring end nodes. In this example, one end node (#6) is only connected to the core node, some others (#3, #5) are only connected to their neighbors, while the rest of the nodes are connected to both the core node and to some neighbors. This topology can be realized on a PON-like tree architecture, if the remote node (the brown box in Fig. 29.12b) also provides for the required interconnections among drop fibers connecting to the respective ONUs [29.36]. The topology in Fig. 29.12a is replicated in the background in Fig. 29.12b to indicate the mapping between both. In fact, all end nodes are a member of the drop section of the PON, i.e., they are all physically connected to the remote node, be they logically connected only to the core node (as for #6), or only to their neighbors (as for #3, #5), or both. The interconnection structure can be extended to neighboring PON-ODNs, i.e., to neighboring clusters of end nodes that are connected to different OLTs (core

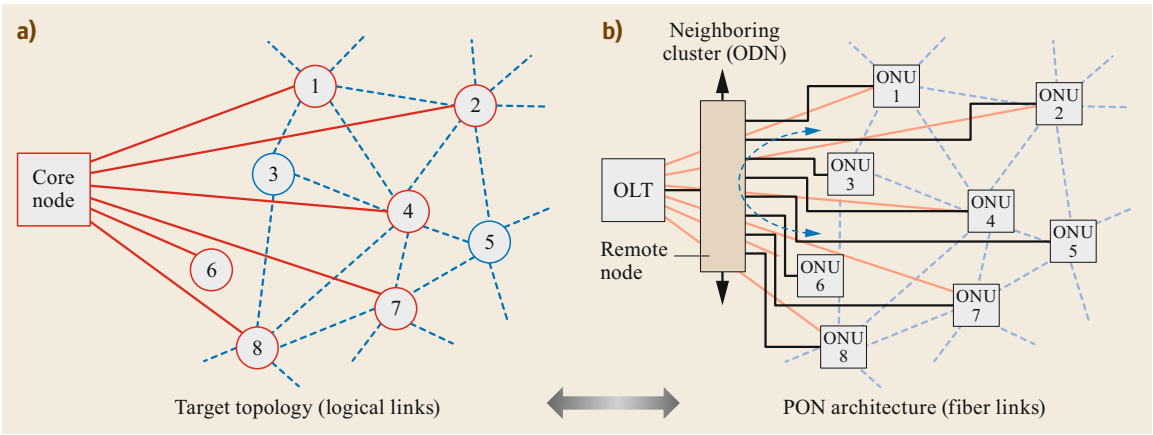


Fig. 29.12a,b Logical links between core node and end nodes, and between end nodes **(a)**; mapping the topology to a ptmp ODN where the drop fibers are reused for direct local interconnects **(b)**

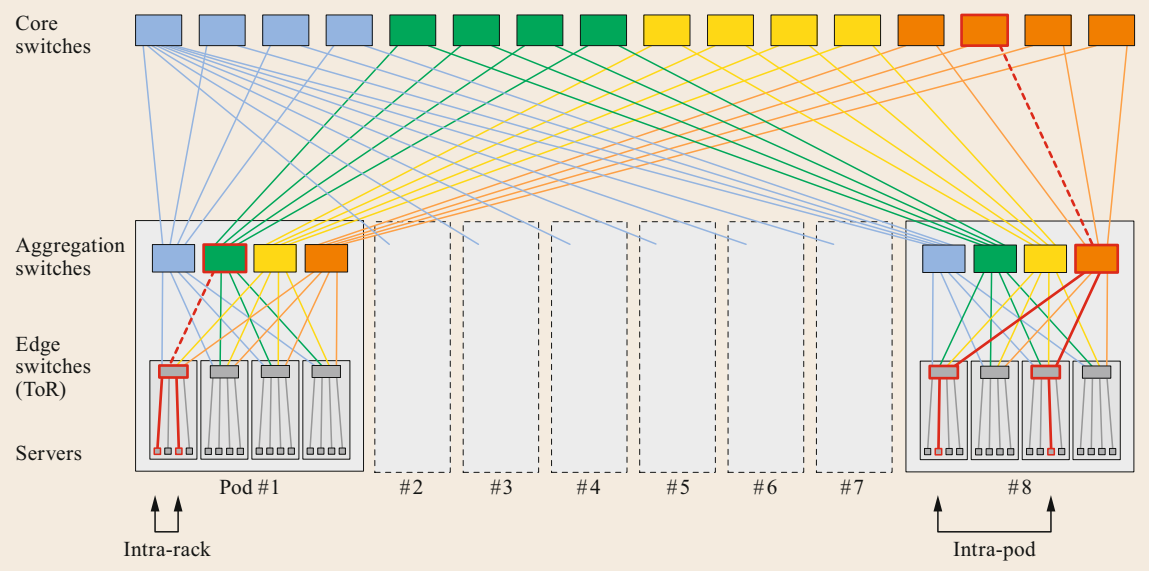


Fig. 29.13 Use case *any-to-any* in an intra-DC network; intrarack (*solid red lines* in Pod #1) and intra-pod communication (*solid red lines* in Pod #8) can be realized by intra-PON communication links via passive loop-back at the connecting node (see the text for details). Note: the line colors other than *red* are only to guide the eye; they have no specific meaning

nodes). This is indicated in the logical topology by unterminated dashed blue lines leaving from, e.g., nodes #1 and #2 towards the top, and in the PON architecture by a correspondingly connecting link exiting from the remote node towards a neighboring remote node.

The networks in Figs. 29.13–29.15 show three different examples in which the network provides for specific ONU interconnection patterns. They are motivated by requirements encountered in certain practical scenarios, as will be described next. None of the solutions has been implemented in any of today's networks yet. However, as will be shown, they can be realized on the basis of existing PON systems and architectures with moderate modifications.

Intradata Center Network (*any-to-any*)

The network in Fig. 29.13 represents a common intra-data center (DC) architecture, the fat tree or folded Clos network [29.37] (actual DC networks may differ in details, but the hierarchical structure and organization of servers and switches, as they are relevant for this discussion, are well represented by the architecture shown). In this example, the switches on all levels (core, aggregation, and edge (top-of-rack: ToR)) are identical in terms of port numbers (here eight ports), link capacity per port, and switching capacity. The traffic patterns and traffic statistics inside DCs are not always alike; they largely depend on the business that the DC has been designed for [29.38]. However, there is a common characteristic in that most of the traffic stays inside the DC, most of it being East–West traffic, much of it staying even within a rack or within a pod. Due to the unpredictability of specific connections between servers, the DC architecture must provide for fast and efficient, i.e., low-latency and high-bandwidth, any-to-any interconnection between servers in a rack (red lines in Pod #1 in Fig. 29.13) or across racks in a pod (red lines in Pod #8 in Fig. 29.13).

Assume that the ports on the aggregation switches that connect towards the servers inside a pod (pod #1 in Fig. 29.13) are equipped with TDM-PON OLT modules, and that the server ports are equipped with ONU modules. The edge switches (ToR) can then be replaced with modified passive power splitters that provide not only for the North–South connection (dashed red lines), but also for the East–West connection (solid red lines) between servers, as is generically indicated in Fig. 29.12 (for implementation details, see Sect. 29.4.3). A similar consideration applies to the intra-pod communication between racks, as shown for pod #8 in Fig. 29.13. Here, the TDM-PON OLT modules reside in the ports of the core switches, while the aggregation switches are replaced with modified power splitters.

Local C-RAN (*one-to-few*)

The second use case (Fig. 29.14) was introduced in the context of low-layer Fx fronthaul over PON-based metro-access architectures (Fig. 29.10). The low la-

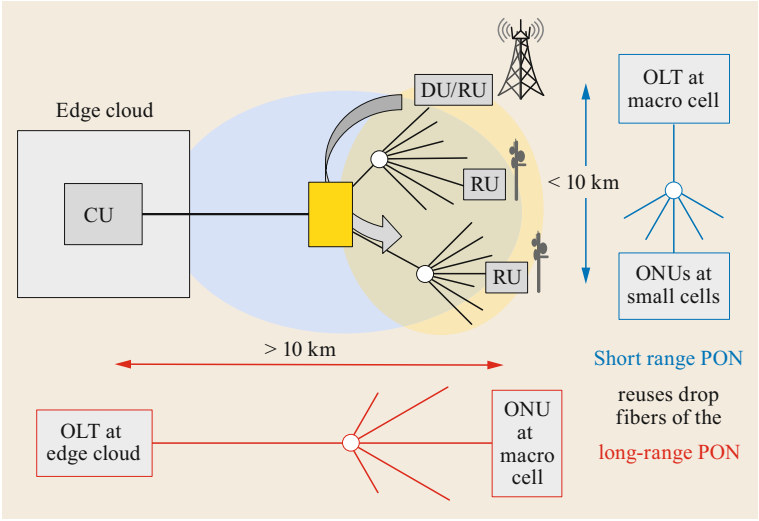


Fig. 29.14 Use case *one-to-few* for low-latency fronthaul in a local C-RAN architecture. The *blue* and *red tree diagrams* indicate how the short and long-range topologies are implemented on the common ODN (cf. Fig. 29.10, only the wireless part is shown here)

tency requirements of such links can be met by implementing a second, short-range PON (blue topology shown to the right of the metro-access network in Fig. 29.14) across the drop section of the first, long-range PON (red topology). The CU in the edge cloud is connected over the long-range PON to the DU at the distant macro cell site (latency-tolerant link) which in turn serves the RU at the macro cell as well as the RUs at the nearby small cells (low-latency links). The short range PON reuses the drop section of the long-range PON to connect the macro cell with the small cells via loop-back at the passive splitter. Depending on the number of small cells on the PON, the loop-back provides for a one-to-few connection pattern (for implementation details, see Sect. 29.4.3).

CoMP Network (*any-to-few*)

The third example is again found in wireless networks. For certain modes of radio operation (cooperative multipoint (CoMP) [29.31]), there is a need to exchange massive amounts of data between neighboring cell sites within a short time. For instance, in the mode of uplink joint reception (UL-JR), each antenna site processes its own received IQ (in-phase and quadrature component) samples together with the IQ samples received from neighboring antennas to improve the radio-link performance. Since low-layer radio data are processed, this scheme requires exchanging data at submillisecond time scales, hence short high-capacity links.

Figure 29.15a schematically shows the topology on the radio layer (top), including the links to the radio core (red) and the direct links between neighboring radio sites (blue). The mapping of this topology onto the PON layer is shown at the bottom. The square grid shown in Fig. 29.15b displays the local interconnection pattern in the case that each radio site exchanges data with its four nearest neighbors. Each site on the square grid is the center of a local neighborhood of four, and at the same time each such site is a member of the respective neighborhoods to those four sites. The remote passive node (brown) of the PON provides for the appropriate connections between the drop fibers connected to it, thus enabling the locally constrained, but (infinitely) extensible any-to-few interconnection pattern.

In the three examples described above, the local interconnects could in principle be established by placing an optical mirror at the CO-sided port of the power splitter, reflecting all ONU upstream signals (optionally on a wavelength channel separate from the regular upstream towards the OLT) back to all ONUs, where they would then be available for further processing. This brute force approach, however, in most cases is not efficient: the looped-back signals pass the splitter twice, thus experiencing excessively high losses, and each attached node must process the data from all other nodes, regardless of whether it participates in the communication or not. Therefore, other solutions have been proposed in the past that can help improve on this by utilizing time domain and/or spectral domain resources dynamically and more efficiently [29.39, 40].

In the following, we will focus on such solutions for each of the above cases that will establish dedicated connections between end nodes over a static optical interconnection pattern. Occasional reconfiguration of the optical interconnects on millisecond time scales or longer is in scope and will be briefly discussed for one use case. However, dynamic path switching on the packet or burst level will not be considered.

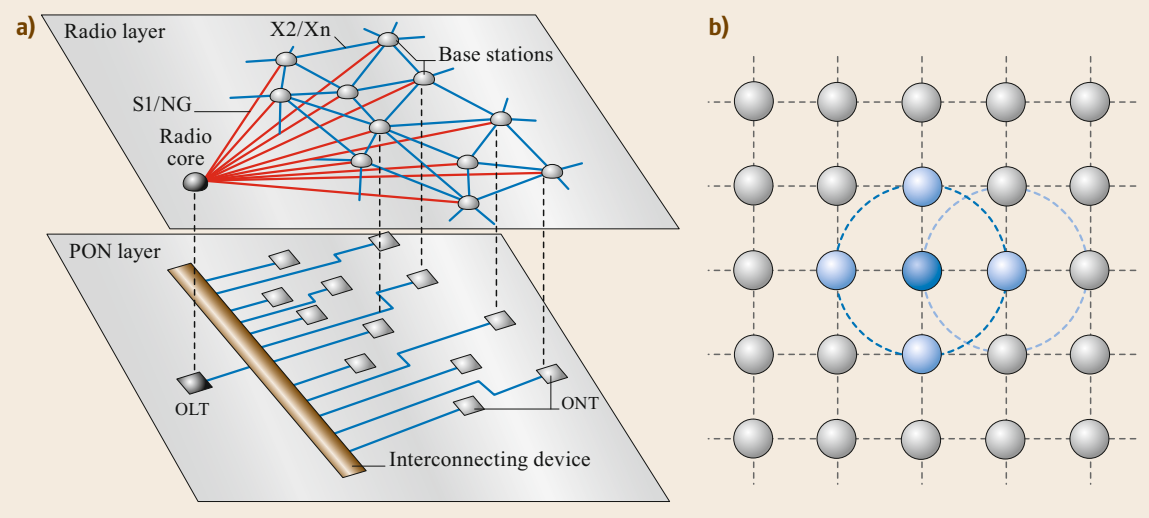


Fig. 29.15 **(a)** Use case *any-to-few* for latency-sensitive interconnects in wireless network operating in UL-JR mode. **(b)** Interconnection pattern on a square grid where each site (cf. *dark blue*) is a center of a neighborhood of four other sites (*light blue*), which, in turn, are centers of their own neighborhoods (one being indicated by the *light-blue dashed circle*)

29.4.2 Wavelength Ranges for Local Interconnects

The optical links for the local interconnects between ONUs can be established on WDM channels separate from those used for the connection between OLT and ONUs. In some cases, sharing wavelength channels among the OLT-ONU connection and the local ONU interconnects (e.g., by TDM) can also be a suitable option, as will be shown in Sect. 29.4.3.

It is assumed that the connection between the OLT and ONUs is accomplished by one of the current system technologies of the ITU-PON or IEEE-PON families (cf. Table 29.1). Multiple wavelength ranges have been specified for those systems such that a large part of the available fiber spectrum is already occupied (Fig. 29.16, see also Table 29.1). In real networks, however, there are typically only one, sometimes two, system variants implemented at the same time on a given ODN. The fiber spectrum is hence not exhausted, so that alternative transmission systems for ONU interconnects or for other special purposes (Sect. 29.5) can, in principle, be operated also in the remaining specified, but unused bands. However, to avoid potential spectrum conflicts in future network migrations, a safer approach is to work with existing PON systems in their standardized bands, or to use the auxiliary band in the range 1380–1460 nm.

So, there are essentially three options of operating ONU interconnects along with the regular OLT-ONU connection on the same ODN:

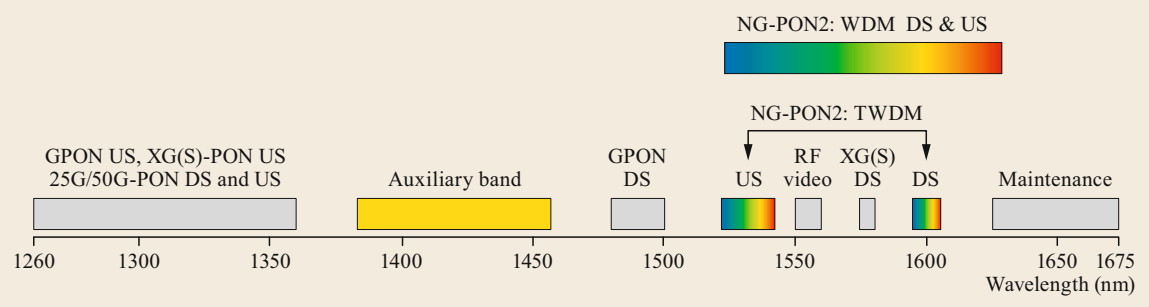


Fig. 29.16 Wavelength ranges specified for different PON generations and the additional auxiliary band (*yellow*: 1380–1460 nm) in the E-band. Only the ITU-PON variants are shown; the wavelength ranges of the IEEE-PON variants coincide with their respective ITU-PON counterparts (cf. Table 29.1)

- Interconnects in the auxiliary band, any PON system for the OLT-ONU connection:
  The E-band is currently not considered in transmission system standards due to the increased attenuation from OH-molecule absorption in old, lower-quality fibers. New fibers do not suffer from this type of attenuation, but still this wavelength range remains unused to guarantee unrestricted usability of the PON systems. For local interconnects, however, the distances are short, so that even an increased differential attenuation coefficient will not severely affect the power budget of those links.
  The boundaries of the considered auxiliary wavelength band (1380–1460 nm) leave sufficiently wide guard bands on either side (20 nm) to allow for inexpensive optical filters.
- Interconnects and OLT-ONU connection on different NG-PON2 wavelengths:
  NG-PON2 specifies multiple wavelength channels in its TWDM-PON and in its WDM-PON subnetworks. So, the interconnects can operate on one or more of those channels, while the regular OLT-ONU connection can be accomplished over another channel in those bands.
- Interconnects and OLT-ONU connection over different PON variants:
  The ONU interconnection can be established using a PON system variant that is different from the one used for the OLT-ONU connection. This typically involves system variants of different transport capacities and operating in different bands, e.g., XGS-PON and 25G-PON.

29.4.3 Remote Node Architectures

There are multiple ways to implement the interconnection links and the remote splitter nodes for the use cases introduced in the previous section. We will focus here on one implementation for each case that seems to be most favorable in view of the overall network solution.

Any-to-Any

In data centers, each server must generally be able to communicate with any other server in the data center, with lowest possible latency and high bandwidth. High bandwidth is, however, not needed continuously, but will be requested dynamically, typically on a per packet-flow basis. Depending on the aggregation level and on traffic dynamics, the interconnecting network between servers can be designed to benefit from statistical multiplexing, a TDM-PON thus being one favorable solution. This approach is particularly useful for the communication between servers in the same rack, or in the same pod, as will be shown below. The multitude of possible connections requires operation on multiple wavelength channels, so that a TWDM-PON approach will be chosen. The architecture will be discussed with reference to Fig. 29.13, here focusing on the intrarack communication.

The connections inside the pod considered are shown at the left of Fig. 29.17. The ToR switches, however, are now replaced with passive remote nodes, each connecting N aggregation switches in the pod with N servers in the rack (here $N = 4$). The nodes are made from $2N \times 2N$ star couplers along with wavelength diplexers at the North-bound interfaces for separating the downstream and upstream wavelengths and with circulators at the South-bound interfaces for wavelength agnostic separation of downstream and upstream signals, as is shown on the right-hand side of Fig. 29.17. The $2N \times 2N$ star coupler connects each of its input ports to each of its output ports by splitting the signals by $1 : 2N$ [29.41].

The ports on the aggregation switches are TWDM-PON OLTs. All N ports on a given switch (i.e., all N OLTs) operate at the same wavelength pair (downstream and upstream). The channel pairs used by the N aggregation switches in the same pod are mutually different. The servers can connect to any aggregation switch in the pod by tuning their TWDM-transceivers to the channel pair of that particular switch. The internal architecture of the remote node provides for the connection between all servers (ONUs) and each aggregation switch (OLT), and at the same time also loops back all upstream bursts from the servers to all other servers in the same rack that are connected to the same switch (OLT) via the selected channel pair.

Figure 29.17b displays the optical connections inside the remote node for one particular OLT and its connected ONUs. The connections through the star coupler are indicated by solid red lines for downstream and solid blue lines for regular upstream. Additional dashed blue lines indicate the looped-back signals among the group of connected ONUs. Since the looped-back signals are identical to the stream of bursts that is transmitted upstream to the OLT, the bursts on the loop-back paths are time aligned without collision and can be detected by the ONUs. To detect the bursts, the ONU transceivers must contain a wavelength tunable burst-mode receiver for the upstream wavelength in addition to the wavelength tunable continuous-mode receiver for the downstream signals.

The OLTs assign the bursts to either the regular OLT-ONU communication or to one or more ONU-ONU communication. The OLT and the ONU burst-mode receivers that are tuned to the same wavelength detect the entire stream. As usual in TDM-PON, the bursts received will then be further processed only at

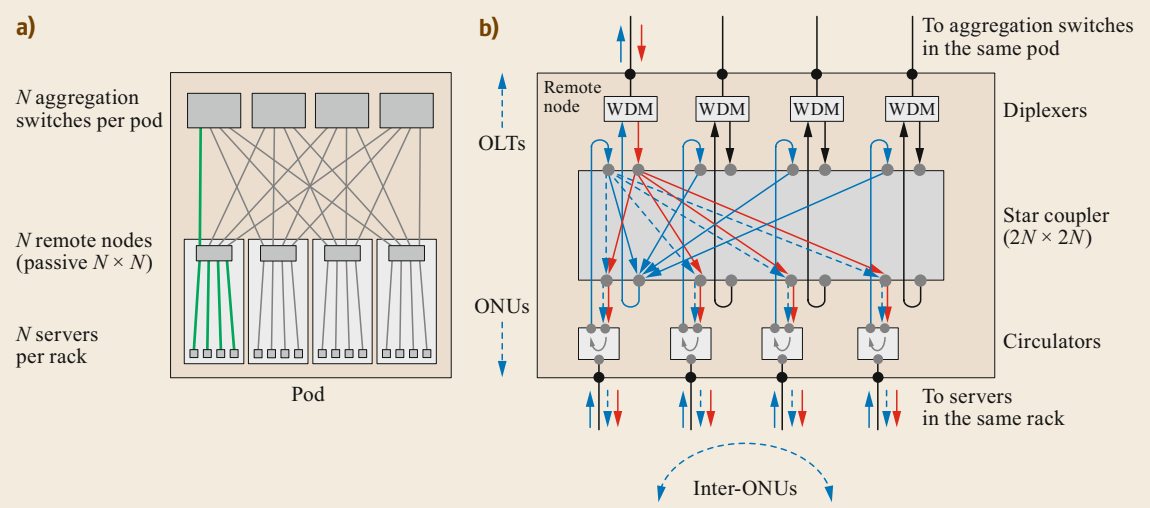


Fig. 29.17a,b Intra-pod connections (Fig. 29.13) with ToR switches replaced with passive remote nodes **(a)**; remote node architecture facilitating connections between switches and servers, as well as *any-to-any* interconnects between servers in a rack **(b)**. The connections of a particular 1 : 4 TWDM-PON channel are highlighted in green **(a)** and in *red/blue* for down/upstream **(b)**, assuming all four ONUs are wavelength tuned to the same OLT. *Dashed blue lines* show light paths for inter-ONU communication

those ONUs (and at the OLT) that they have been addressed to.

For a remote-node connected to N aggregation switches and to N servers inside the rack, the star coupler needs $2N$ ports on either side. The splitting loss for all signals (OLT-ONU and inter-ONU) then amounts to $1/2N$, i.e., 18 dB for $N = 32$. In real implementations, the total loss also includes component losses (excess loss of the star coupler, WDM diplexers, circulators) and the losses of the usually short fiber links inside the DC. Still, it will remain well below 30 dB and can, hence, be bridged by NG-PON2 optics made for the low optical power budget class N1 (up to 29 dB). In this example, the number of wavelength channels and the tuning range are well beyond current NG-PON2 specifications: 32 TWDM channels on a 100 GHz grid would cover a range of 3.2 THz, requiring widely tunable transceivers (across 24 nm in C/L-band). However, with reduced channel numbers, narrower channel spacing (25 or 50 GHz), or with slightly modified network architectures, these challenging requirements can be relaxed (for large N there will unlikely be as many aggregation switches as there are servers in a rack, therefore less wavelength channels would be sufficient). For an alternative network approach employing wavelength tuneable ptp channels, see [29.42].

One-to-Few

In the network shown in Fig. 29.14, the edge cloud serves remote nodes of a radio network over a long-distance PON. While mobile fronthaul over long distances is not an issue at the F1 interface with eMBB services, the implementation of an Fx fronthaul link over more than about 10 km is in most cases not possible due to the latency constraints at the low-layer splits [29.27]. So, a fully centralized architecture with CU and DU in the edge cloud will not be possible. However, in UDN architectures as shown in Fig. 29.14, a local C-RAN can be established for the small cells surrounding a macro cell.

In this scenario, the macro cell site is connected to the edge cloud over a long-range PON carrying the high-layer split (F1) fronthaul data for the macro cell site as well as for the neighboring small cell sites. The CU for all sites is located in the edge cloud, and the DU for all sites is located at the macro cell site. From here, the RU of the macro site is served directly (the macro DU and the macro RU are colocated), whereas the RUs of the small cells are served from the respective DU functions at the macro site over local interconnects between the end points of the long-range PON.

The optical connection for the latency-sensitive Fx interface at the low-layer split between DU and RUs is established as a local short-range PON from the macro cell site to the small cells, reusing the fibers in the drop section of the long-range PON. The passive remote node provides for the necessary loop-back function, as shown in Fig. 29.18.

The F1 fronthaul link to the macro cell site carries the data for both the macro cell and for the nearby small

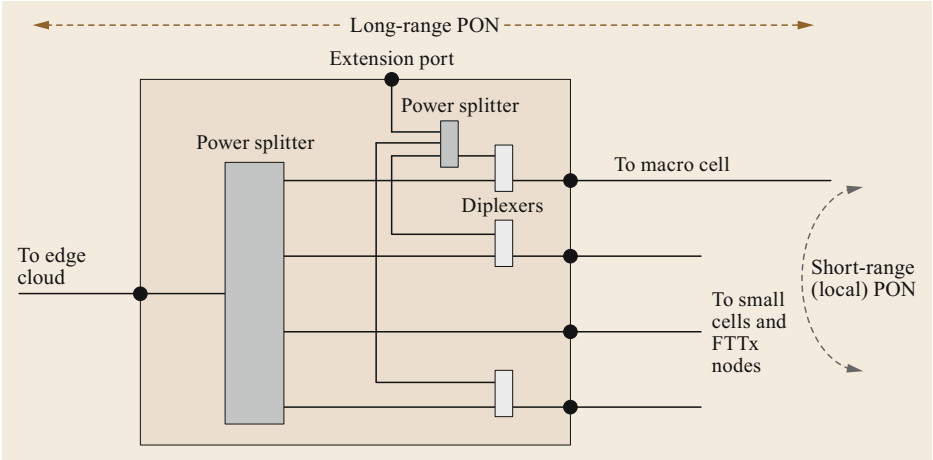


Fig. 29.18 Remote node in a long-range PON, providing for *one-to-few* interconnects between some end nodes, thus establishing a local C-RAN architecture across the drop section of the long-range PON

cells. Depending on the required capacity, aggregation, and statistical multiplexing efficiency, this link may be realized by a high-rate ptmp-TDM-PON channel multiplexed with other services on the PON, or it may be provided by a dedicated ptp-wavelength channel. The local fronthaul links from the macro cell to the small cells are established by a separate PON system with its OLT located at the macro cell site, and the ONUs located at the small cell sites. Here again, depending on capacity, aggregation, and statistical multiplexing, this may be realized by a single ptmp-TDM-PON channel or by a ptp WDM-PON.

The network as a whole, i.e., the long-range PON and the local short-range PON for the interconnects, may be composed of the wavelength channels in an NG-PON2-type system (TWDM-PON or WDM-PON) or by different system variants operating on different wavelength channels, such as XGS-PON for the long-range connection and a 25G-PON for the local fronthaul links.

The local PON experiences a smaller splitting loss than the long-range PON due to the small number of small cells, compared to the total number of attached ONUs. Together with the small loss of the short fiber links, this substantially relaxes the power budget for the short-range PON. In turn, this allows connecting additional cells also outside the ODN considered by using a spare port (extension port) on the small power splitter of the short-range PON (Fig. 29.18).

The remote-node architecture shown in Fig. 29.18 contains optical diplexers on those drop fibers that connect to the small cells. Due to the different wavelengths used for the long-range PON and for the short-range PON, those drop fibers can be used for both PONs at the same time. Not shown in the figure are the respective second diplexers in the drop section, where the data for the small cells are separated from, e.g., FTTx services provided from the edge cloud to residential customers over the same drop fibers. Alternatively, those drop fibers can be used exclusively for the small-cell fronthaul, thus enabling a clear separation of services on the physical layer. The choice obviously depends on the availability of fibers, as well as on operation preferences of the network operator.

Any-to-Few

The structure of the remote node shown in Fig. 29.19 is a generalization of the structure in Fig. 29.18. It allows for a local communication between any end node of a PON and a small group of neighboring nodes over the drop fibers. This can be used for realizing, e.g., the wireless architecture shown in Fig. 29.15. The type of splitters (here 1 : 4 splitters) in the coupling unit and the way of interconnecting their ports realizes a square grid for the attached end nodes, as shown on the right-hand side of Fig. 29.15. Topologies of higher dimensions for the attached end nodes can be realized by increasing the split ratio of the power splitters: a 3-D-cube topology is realized with 1 : 6 splitters, a 4-D-hypercube topology with 1 : 8 splitters, and so forth [29.43].

The optical communication on this complex topology is realized in a different way compared to the previous two examples. Figure 29.20a shows schematically how one particular end node (ONU #10, connected to splitter #10 in Fig. 29.19) of the network communicates with its neighboring ONUs #6, #9, #11, and #14 and with the central OLT. ONU #10 transmits the data for all of its four neighbors in a single time-domain multiplexed stream (e.g., Ethernet), broadcast to them on a single wavelength channel (λ_2 in Fig. 29.20a). At the same time, it receives multiplexed streams that are broadcast from nodes #6, #9, #11, and #14 to their respective neighbors (including ONU #10) on wavelength

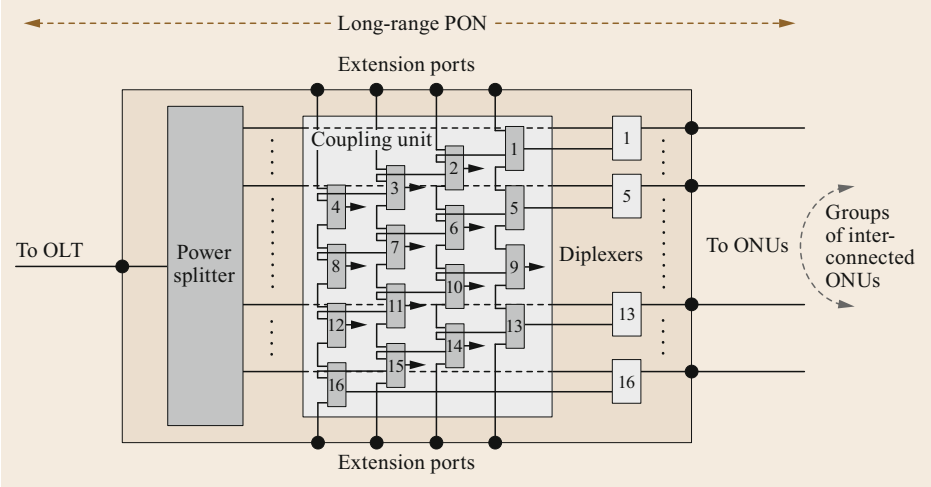


Fig. 29.19 Remote node for *any-to-few* interconnects on a square grid topology providing direct optical interconnects between any end node and its four neighbors (the ports of the main splitter are directly connected to the diplexers, cf. *shaded lines* across the coupling unit). (Note: for the sake of simplicity, not all links inside the coupling unit are shown, e.g., the splitters 1, 5, 9, and 13 each have an unused port, which can be used for more extension ports; the same applies to splitters 4, 8, 12, and 16. The *little arrows* from, e.g., splitters 2, 6, 10, and 14 to the right indicate the connection to the respective output port of the remote node via additional diplexers.)

channels λ_1, λ_3, λ_4, and λ_5. From those streams, ONU #10 extracts only those data that are relevant to it. The wavelength channels for the inter-ONU communication can be chosen on, e.g., a 100 GHz-channel grid in the auxiliary wavelength band shown in Fig. 29.16.

In addition to this local communication within each group of nodes, there is another link from ONU #10 to the central OLT via the large power splitter in the remote node in Fig. 29.19 (not shown in Fig. 29.20). This link belongs to the long-range PON, operating on a separate downstream/upstream wavelength pair ($\lambda_{DS}/\lambda_{US}$), as specified for the chosen PON system variant. The related ONU transceiver structure is schematically shown in Fig. 29.20b. Instead of using fixed wavelength optics as shown, a more flexible implementation using wavelength tunable optics will be preferable, especially when considering occasional network reconfigurations (Sect. 29.4.4).

The small size of a communication group (five ONUs in the above example) requires only a small number of wavelength channels across the entire network, regardless of the total number of nodes, i.e., regardless of its topological extension. For a grid in which each node has N neighbors to communicate with, a set of $N+1$ wavelength channels is sufficient. On larger and more complex structures it may become challenging to consistently assign just those few wavelength channels without encountering channel collisions (i.e., a node

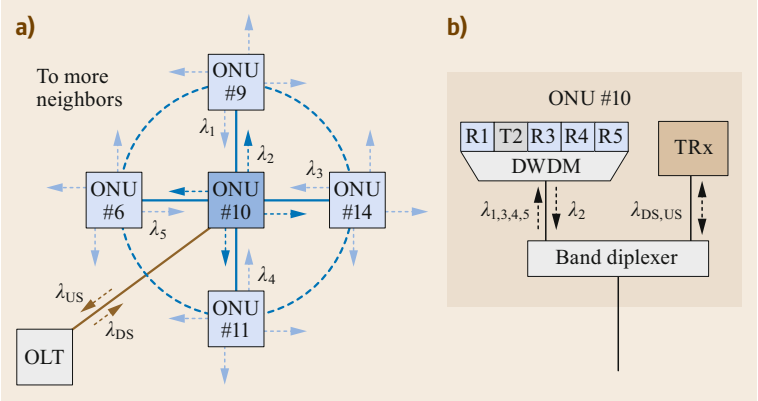


Fig. 29.20a,b Optical channels used for inter-ONU communication on a square grid in an *any-to-few* architecture, and for the link to the central OLT. **(a)** Wavelength assignment for the broadcast channels leaving each node; each wavelength is sent to all respective next neighbors. **(b)** Transceiver architecture for ONU #10 with a transceiver for the long-range PON (*TRx*), and the group of four receivers R1, R3, R4, and R5, and one transmitter T2 for the local interconnects

receiving identical wavelength channels from different neighbors); even more so, when the network is being reconfigured or new nodes are added later on. So, a few more than $N + 1$ channels would ease the implementation in such cases.

29.4.4 Cross-ODN and Reconfigurable Local Interconnects

In Fig. 29.19 it was shown that at the edges of the grid multiple splitter ports remain unconnected. These ports can be used for extending the grid into the ODN of neighboring PONs (Fig. 29.21a) or into the neighboring ODN segment of the same PON (Fig. 29.21b). This is an essential feature that is needed in cases where, e.g., an UE moves through multiple cells, and the antenna group performing UL-JR processing changes accordingly. At some point, the UE will leave the group of antennas attached to one particular ODN (segment) and will enter the group of antennas served by the neighboring ODN (segment).

If the remote nodes of both ODNs are located at the same location (e.g., a street cabinet), then the extension ports of the remote nodes can be directly connected using patch fibers (Fig. 29.21a). If the remote nodes are located in different locations, then the signals on (multiple) extension ports can be wavelength multiplexed onto a common port and be passed over, together with the long-range PON wavelength channels, to the neighboring ODN at the next higher split level (Fig. 29.21b). It must be noted that in both situations the wavelength channel assignment for the ONU interconnects must now take into account the nodes of the extended grid, i.e., also the ones on the other side in the neighboring ODN to prevent channel collisions.

In real networks, the interconnect configurations must occasionally be modified over time, be it for repairs, system migration, or network capacity enhancements. This may require a change of the interconnection topology for the network considered, hence not only the ONUs, but also the remote nodes should be reconfigurable. This will also allow using a common hardware platform for different network scenarios.

There are essentially two options to make the remote node reconfigurable (Fig. 29.22):

- The output ports of the remote node can be interchanged using an $M \times M$ optical switch after the diplexers, where M denotes the number of end nodes attached to the remote node.
- The interconnection of the splitter ports inside the coupling unit can be modified using a single-sided

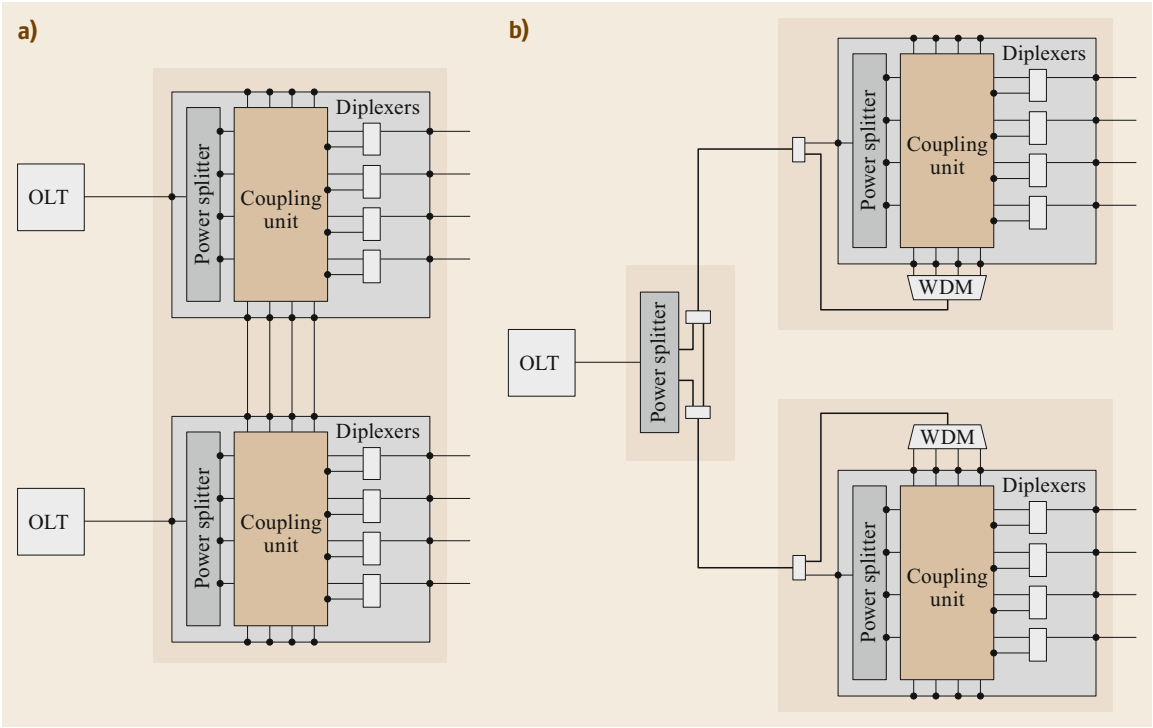


Fig. 29.21a,b Grid extensions with remote nodes located **(a)** in common locations, **(b)** in different locations

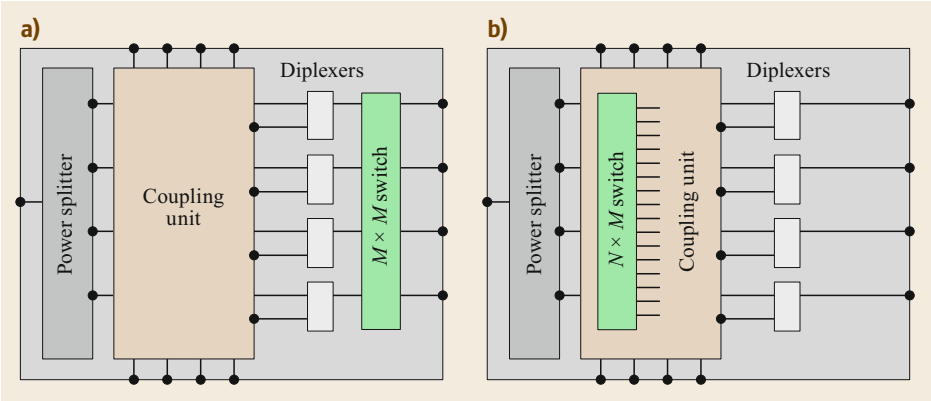


Fig. 29.22a,b Remote node for reconfigurable local interconnects. **(a)** Reconfiguring the output ports after a fixed coupling unit, and **(b)** reconfiguring the interconnects among the splitter ports inside the coupling unit

$N \times M$ optical switch (Fig. 30.12 in Chap. 30), where N indicates the number of internal splitter ports ($N = 4$ in Fig. 29.19).

In the first option, only the output ports towards the ONUs are exchanged, but the topology of the port interconnection pattern is retained.

In the second option the entire topology can be modified, meaning that the communication relation between any two end nodes may be completely different after switching. It also modifies the topological role of the extension ports. In general, this will need a reassignment of the wavelength channels for the interconnecting links.

It should be noted that the node architectures in Figs. 29.18 and 29.19 can also be realized as a remote node for a wavelength routing WDM-PON. In that case, the main power splitter on the left-hand side of those figures would be replaced with a wavelength routing element. The power splitters that establish the local interconnects (single splitter at the top of Fig. 29.18 or the splitters inside the coupling unit in Fig. 29.19) can likewise be replaced with wavelength routing elements. Both variants can also be used, if the long-reach PON is replaced with ptp-fiber links from the edge or access cloud to the end nodes. In that case, the main splitter is not used and the long-range ptp-fibers are connected directly to the diplexers at the output ports in the remote node.

29.5 Smart ODN

The ODN in optical access networks is usually considered entirely passive, as the term PON for the transmission system suggests. Hence, it is static and does not contain means for monitoring and management. Any possible intelligence that can be used for supervising the state and the operation of the network is usually available only at the end points of the link, i.e., at the OLT and at the ONUs. In practical network operations, however, a more detailed knowledge about the status of the optical links, the availability and performance of fiber connections, and more, is essential to identify existing faults or anticipate future faults in the ODN. The continuous supervision and management of the ODN becomes more important as fiber access networks are becoming more complex, and as their uninterrupted availability is becoming more essential for customers. A regular update on the network status helps avoid unexpected network interrupts and will ease repair actions during operation.

29.5.1 Motivation

These operational aspects can be addressed to some extent in the case that the network operator owns and runs both the passive fiber infrastructure and the active transmission equipment. Transmission systems provide for various measurement means that can be used to monitor the status not only of the active system equipment (e.g., transmitted and received optical powers, laser current, electrical parameters of the transceivers and their temperature), but also of the status of the ODN by, e.g., correlating these data across different groups of ONUs. Information from higher system layers, such as Ethernet frame losses and other alerts on Layer 2, can also support isolating optical link issues. All this, however, does not allow for clear and unambiguous identification of the root cause of an issue on the fiber layer, nor does it provide precise information about where the fault is located, information that is needed for quickly initiating

repair. So, additional measurement tools are typically used for monitoring and managing the fiber layer, such as an optical time-domain reflectometer (OTDR) for localizing fiber cuts, identifying increased connector losses, observing splice degradations, and more.

If the fiber infrastructure is owned and maintained by a company that does not operate the transmission system or provide the services, i.e., it rents out the fiber infrastructure to other network or service operators, then there is no way for the fiber infrastructure owner to access information that is generated by system-integrated monitoring means at the OLT or ONUs (e.g., higher-layer alerts). Depending on the particular network situation, on the business case (residential services, business/enterprise services, wireless back and fronthaul, campus networks, multiple network operators) and on the related SLAs between the infrastructure owner and the network or service operator(s), a reliable, but independent ODN management approach may become indispensable. The fiber infrastructure should become an independently managed layer of the network [29.44].

Aside from bare monitoring, there is also a need to identify and reconfigure the connectivity map of the fiber infrastructure, e.g., for protection switching in the case of fiber link failures. In FTTx networks, the links to residential customers are usually not individually protected. However, for business-critical services or for back/fronthauling aggregation nodes, such as wireless, coaxial, and DSL access points, where large numbers of end users would be affected by a failure, protection switching is essential. In addition, flexible and remote reconfiguration of certain fiber links or even of entire network segments will increasingly be asked for. Possible reconfigurations comprise, e.g., modification of the split factor or of the split ratio of power splitters, modification of the wavelength routing pattern of a WDM node, turning a node from power splitting to wavelength routing, and more, which would allow operators to react to the changing needs of service delivery or to changing customer distributions.

The above reconfigurations can be accomplished by using common optical technologies, such as wavelength-selective switches (WSS) and reconfigurable add/drop multiplexers (ROADM), as they are used in, e.g., high-capacity metro or longhaul networks. For use in access networks, however, this type of equipment is usually over-dimensioned, too bulky, too expensive, and too power hungry. Therefore, solutions are sought that are better adapted to the needs of optical access networks in terms of functionality, size, and power consumption. In access networks, the remote nodes are frequently positioned in street cabinets where electrical power is available. However, this is not always the case (e.g.,those in manholes and buried in the ground or suspended from poles in aerial fiber deployments), so that remote powering, supported by a node design for low-energy consumption, would be needed.

In this section a number of means will be introduced that allow for monitoring and reconfiguring the optical layer in access networks in as simple and inexpensive ways as possible. Starting from all passive monitoring, the described node architectures will become increasingly complex, incorporating active elements and being powered remotely by optical energy. The objective of the discussion is to elaborate what can be accomplished under access typical operations and cost constraints, while targeting at a fully manageable optical fiber layer. It is noted that only the first, most simple features are currently considered for testing in real networks, while the more sophisticated node architectures are still to be elaborated in more detail before they can be considered for introduction into real networks.

29.5.2 Optical Layer Monitoring (OLM)

The end-to-end availability of a fiber link can be tested all-optically, without resorting to higher-layer information from the transmission system, if an optical mirror with specified reflection factor is inserted at the client side of the network, separate from the ONU (Fig. 29.23a). This reflection is accomplished either by a partially reflecting mirror for the downstream data signal (e.g., 1% reflection) or by a 100% reflecting mirror for an OTDR operating in a separate wavelength range [29.45, 46].

Passive Demarcation Point Monitor

Such a demarcation point monitor (DPM) allows the operator to keep track of possible degradations of the end-to-end link quality by monitoring the remote reflection. (The arrangement shown in Fig. 29.23b achieves the desired reflection factor by appropriately choosing the wavelength responses of the coupler and of the mirror as described in the figure caption. An alternative solution is a suitably designed mirror inserted into the fiber link. That mirror would, however, have to be designed to leave all other—also future—signal wavelengths unaffected.)

When the reflection factor is measured to be seemingly reduced, this would indicate an increased link loss between the CO (central office) and the demarcation point. It can also help isolate the root cause of the degradation, because correlating this observation with the response from other DPMs on the same ODN can narrow down the options of where in the ODN the degradation may have occurred (before or after the

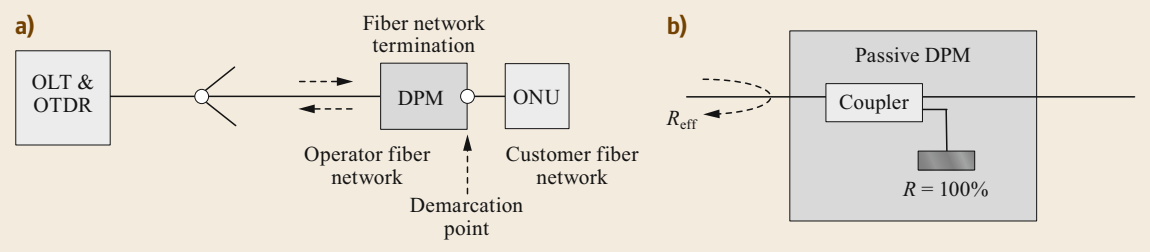


Fig. 29.23 **(a)** Optical demarcation point with integrated monitor (DPM) for terminating the operator's fiber network at the customer's premise. **(b)** Example of a passive DPM with effective reflection factor $R_{eff} = -20$ dB for measuring back-reflected downstream signals by using a −10 dB tap coupler or with reflection factor $R_{eff} = 0$ dB for OTDR measurements in the maintenance wavelength range by using a wavelength-selective coupler (cf. Table 29.1) [29.46]

splitter, or between splitters in cascaded split ODNs, etc.). In addition, this DPM provides for terminating the operator's fiber network at the customer's premise. In the case of service degradations it allows identifying whether the issue is located in the operator's network or in the customer's (home) network.

There have been proposals to use the residual reflection of the optical transceiver module in the ONU for this type of measurement [29.47]. This approach, however, yields unreliable results, since the optical transceiver modules are designed for low reflection (typically less than −30 dB). Their reflection factor is not sufficiently precise, instead it will vary from sample to sample. Moreover, the measured reflection will easily be altered by reflections from nearby optical connectors or splices, which may even vary with changing environmental conditions. Aside from this uncertainty, the approach does not meet the target of separating the operator's ODN from the customer's in-house network, to which the ONU usually belongs.

The practical usefulness of passive DPMs using simple mirrors is limited by the fact that in ptmp ODNs (PON) their reflections are in many cases indistinguishable from each other. Without prior knowledge about their location in the ODN and their distance from the CO (e.g., from fiber installation records), their reflective response in an OTDR trace cannot unambiguously be assigned to a particular customer's premise. In, e.g., FTTH deployments in multitenant buildings with multiple apartments on the same floor, it is even possible that the reflections from multiple passive DPMs overlap. Unintentional reflections from other elements in the ODN, such as bad connectors, add even more to the ambiguity of the measurement results. To identify such details more precisely, an OTDR with high spatial resolution, combined with a high dynamic range, is needed [29.45]. In commercially available OTDR equipment, these two features are usually not selectable at the same time (Chap. 36). Instead, a good spatial resolution is often achieved with short test pulses, which in turn does not allow achieving a sufficiently high sensitivity for use with high-loss ODNs.

The above-mentioned drawbacks can be alleviated by employing an *optical signature* at the passive DPM, so that it can be unambiguously identified by its unique reflection pattern. Such a signature can be the reflection at a single, individual wavelength or an individual pattern of reflections at multiple wavelengths (spectral code). Single wavelength reflections, as well as spectrally coded reflections can be interrogated using a broadband optical source (several nm wide) or a wavelength tunable optical source [29.48, 49]. The link availability can then unambiguously be tested end-to-end, provided a sufficiently high number of spectral reflection patterns (single wavelengths or complex codes) can be generated in the spectral band, and that they can be clearly identified at the receiver. (These requirements resemble those in spectrally encoded code division multiplexed systems, see Chap. 28). Similar considerations apply to using time-domain codes for DPM signatures (i.e., identifying unique pulse patterns upon reflection of a single test pulse). In either case, the large number of DPM devices with different reflection signatures would be challenging for the operator to handle in a high-split ODN. Keeping many different hardware signatures in stock and making sure that they do not get confused during installation and repairs is an operational challenge in large networks. A practical solution should be based on using devices that can be configured once during installation (set-and-forget). Suitable technical realizations for this are, however, not yet available.

Active Demarcation Point Monitor

An active DPM overcomes the discussed deficiencies of passive DPMs by its ability to identify itself with a logical ID# (identification number) that can be flexibly assigned during or after installation. In addition to a simple *alive* message, more information can be conveyed, and even more optical functionalities can be included [29.50].

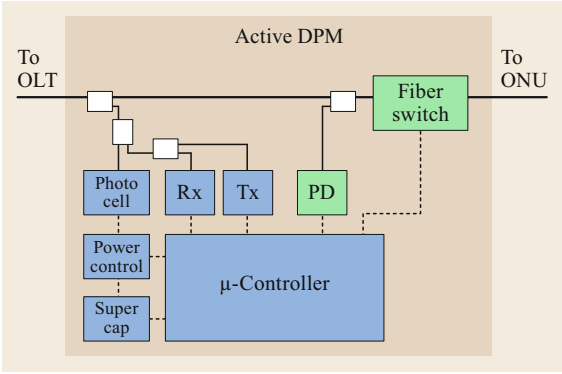


Fig. 29.24 Active demarcation point monitor (DPM) setup (*blue*: basic components for powering, communication, and control; *green*: auxiliary functions; *white*: wavelength diplexers or tap couplers)

Just like with the passive DPM, the active DPM concept does not only provide for information about the link availability, but it also helps terminate the operator's fiber network and separate the outside ODN from the customer's network. So it should remain separate from the ONU and stay operationally independent from the customer's electrical supply. The active DPM shown in Fig. 29.24 contains a photocell for converting optical into electrical energy which is stored in a rechargeable battery or in a supercapacitor (supercap). An ultra-low-energy microprocessor controls the DPM operation and communication. An optical receiver (Rx) and transmitter (Tx) are available for communication with an ODN management module (OMM) in the CO that connects to the network operations support system (OSS) for coordination of monitoring actions and data analysis (Fig. 29.27 and Sect. 29.5.4). The optical powering and communication are accomplished over optical channels in a wavelength region outside the PON transmission windows, e.g., in the auxiliary wavelength band shown in Fig. 29.16. The photocell, as well as Rx and Tx are connected to the transmission fiber by wavelength-selective couplers. These functional blocks are the basic elements for operating active remote nodes in access networks, as they are proposed in this section [29.51].

The optical power injected from the CO into the ODN for feeding the remote nodes can be as high as +20 to +25 dBm. Depending on the ODN loss and architecture (power splitting and/or wavelength routing), the DPM will then receive about 0 dBm optical power or slightly less. Even with only −10 dBm optical power received at the DPM, useful functionalities can be implemented in addition to the communication channel between DPM and OMM. Such functionalities can, e.g., comprise connection control and connection identification:

- The DPM in Fig. 29.24 can be instructed by the OMM to activate a latching (for low-energy operation) fiber switch, to cut the connection between the ODN and the customer's fiber to protect the operator's network in case of a malfunctional ONU.
- With the additional power detector (green PD in Fig. 29.24), connected to the transmission fiber via a tap coupler, it can detect the arrival of an optical power burst sent by the attached ONU. This allows associating the DPM ID# with the ID# of the connected ONU. For this to happen, the OLT instructs the ONU to send an optical burst, the arrival of which is detected at the DPM by measuring the optical power using the green (low bandwidth) PD. The DPM then reports this event via the OMM to the operations support system (OSS), which in turn correlates the burst transmission from the known ONU and the burst detection at the known DPM. Knowing the location of the DPM (from installation records), the connected ONU can hence be localized too, thus allowing for remotely creating and updating the connectivity map of the network at any time.

Intelligent Splitter Monitor

The connectivity identification scheme described in the second bullet point above can be applied also to identify which ONU is connected to which splitter port in a remote node. This information is especially important for repair actions in order not to disconnect the wrong ONU if there is an issue to be fixed with another ONU. There should be records from installation that clearly provide this information. In practice, however, it turns out that after some time, and after multiple repairs or (manual) reconfigurations, those records are not always properly maintained, hence they do not reliably represent the actual status of the network connectivity map. The intelligent splitter monitor (ISM) enables a process that allows the operator to remotely update the required data base at any time, also during network operation.

The ONU-to-splitter port mapping is based on the same sequence of steps as described above for mapping an ONU to an active DPM [29.51]. All ONUs of interest will be instructed sequentially, one by one, to send a burst upstream, while the power detectors (brown PD array) connected to the splitter ports sense on which port the power burst arrives (Fig. 29.25). The port ID#s are reported back to the OSS via the OMM where the correlation of ONU ID# and splitter port ID# is analyzed, and the port mapping is stored in the data base. In cascaded splitter architectures, this detection process can be performed sequentially on all splitters stages, thus also yielding information about which splitter is connected to which port on the preceding splitter stage.

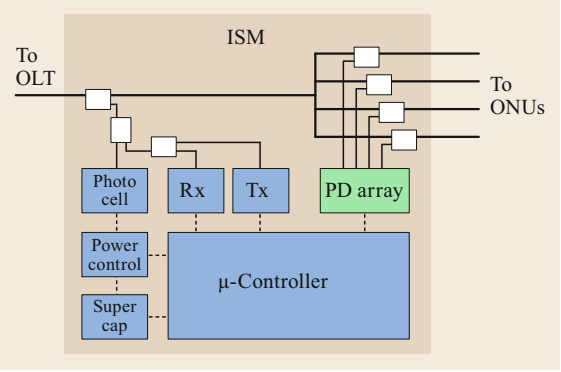


Fig. 29.25 Intelligent splitter monitor (ISM) setup (*blue*: basic components for powering, communication, and control; *green*: auxiliary functions; *white*: wavelength diplexers or tap couplers)

In PONs with many ONUs, this process can take a long time and can exhaust the available electrical energy stored in the case of remotely powered nodes. Moreover, it can lead to long outage times for the ONUs, because whenever the detection process is active for one ONU, all other ONUs (directly or indirectly) connected to the same splitter must stop transmitting to enable unimpaired identification of the respective port. Both these drawbacks are relieved by keeping all ONUs active as normal and then instructing them one by one to switch off for a short time. This *negative logic* yields the same information as the *positive logic* described above. It can also be applied with transmission systems that do not operate in burst mode, such as WDM-PON on a power-split ODN.

For an even better efficiency in terms of low-power exhaustion and short ONU outage time, the ONUs can be assigned to two groups of equal size, the ONUs in one group being active, the others being inactive (active/inactive ONU means it does/does not participate in the respective test). The power detection is then performed on all splitter ports at the same time. In the next cycle, the ONUs are reassigned to the groups in a different way, and the power detection is again repeated on all ports at the same time, and so forth, as indicated by the green and red bullets in Fig. 29.26. For a PON with eight ONUs, this process will take only $3 = \log_2(8)$ test cycles instead of eight to eventually identify the splitter ports for all eight ONUs.

After this first step, all ONUs can be mapped to one specific splitter port in the last stage. For port identification in the preceding splitter stage, only one ONU per splitter of the last stage is selected to perform the testing, and so on. In the end, all splitter ports in the network can be unambiguously assigned to an ONU or to another splitter on the succeeding stage.

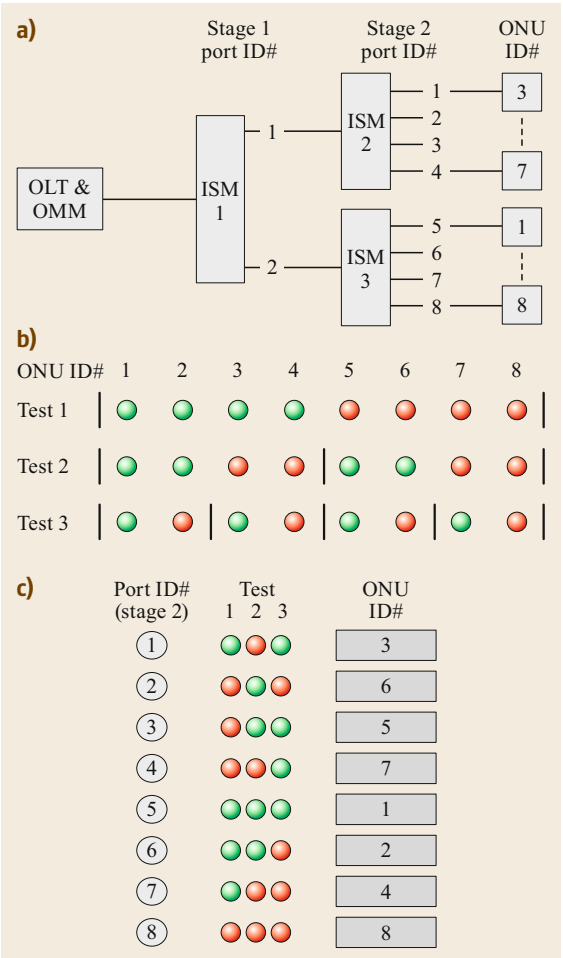


Fig. 29.26a–c Detection process for mapping eight ONUs to the splitter ports (ISM ports) in the last stage of a two-stage cascaded split ODN **(a)**. *Green/red dots* indicate active (= participates)/inactive (= does not participate) in the respective test **(b)**. The detection pattern per port (*green/red* = on/off) allows for identifying the ONU attached to it **(c)** [29.51]

Finally, for a TWDM-PON, the above tests must be performed on one wavelength channel at a time. Here again, there is the option of applying negative instead of positive logic, such that the unaffected wavelength channels can keep transmitting data while one channel is being tested. In this case, the power detectors in the ISM must be able to clearly differentiate between multiple power levels that are generated by the superposition of bursts in different wavelength channels.

Monitored ODN as a Separate Network Layer

When introducing the active DPM and the ISM together into a network, the ODN topology can be detected

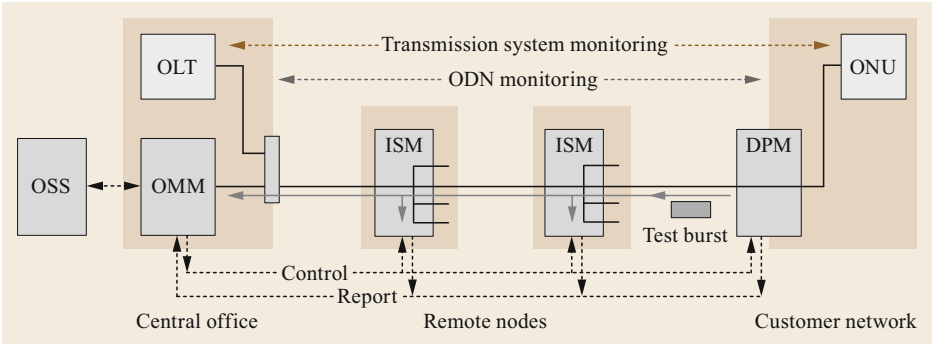


Fig. 29.27 Autonomous end-to-end ODN monitoring (the transmission system monitoring comprises internal monitoring of the optical system equipment and optionally of higher-layer signals)

autonomously, without activating the ONUs to generate test signals. Instead, the Tx of the DPMs can be triggered to emit bursts on their communication channel (in the auxiliary wavelength band), which are then used to perform port mapping in the ISMs as described above. This allows managing the ODN entirely on its own, as an autonomously managed layer of the network (Fig. 29.27).

Once this remotely managed ODN architecture is in place, other monitoring functions can also be added to the remote nodes, such as an inline OTDR. A simplified, possibly remotely powered OTDR unit that can be switched to any splitter port, would allow testing each attached fiber separately [29.52]. Unlike full end-to-end OTDR measurements on ptmp-ODNs, this testing of individual ptp fibers up to the next node can work with a reduced dynamic range of $< 10\,\mathrm{dB}$ and yields unambiguous information about potential faults on those links. The reflected signals from beyond the next node are discarded in the evaluation.

Aside from the segregation of infrastructure segments (operator's network versus customer's network) and aside from topology detection by port mapping or other monitoring functions, this architecture has all the basic elements needed to implement additional smart functions in the remote nodes, as will be discussed in the next section.

29.5.3 Smart Branching Nodes

The concept of low-power monitoring in remote nodes can be extended to functionalities that support fiber switching for network reconfigurations. The optical switches should be of latching type, which consume electrical energy only when switching from one state to another state [29.53], so that they can be implemented even in remotely powered architectures.

Simple 2×2 cross/bar switches are used, e.g., for fiber protection switching. Here, we consider more complex reconfigurations of the ODN connectivity map to accommodate occasional changes in service provisioning or customer connections. Conventional WSS or ROADMs, as they are found in large metro networks, would be over-dimensioned for optical access networks. Instead, flexible variants of simple passive branching nodes that allow choosing among a reduced set of possible configurations matching practical use cases in fiber access networks will be discussed here. In addition to the switching functionalities, the nodes should also contain monitoring functions such as those described in the previous section, so that they are called smart branching nodes (sBNs) [29.44]. It must be noted that the approaches introduced in the following are still too complex for implementation in real networks. However, they show possible paths forward towards rendering the fiber plant in access networks a more versatile asset for network operators.

Multiple approaches have been proposed in the literature to realize such reconfigurable sBNs. In the conceptually most simple case, fiber switches are used for choosing among fixed preconfigured branching configurations. In [29.54], a flexible 1 : 16 branching node was realized that can be configured for pure power splitting or for coarse wavelength routing (CWDM) or dense wavelength routing (DWDM), respectively, followed by additional power splitting with a reduced split factor. This kind of approach allows a simple implementation using state-of-the-art components. However, it is typically constrained to only a few possible configurations. In the following, two different proposals are shown that are designed for higher flexibility and for more compact implementation, respectively.

The quasi-passive reconfigurable node (QPAR) concept is based on a modular architecture that allows flexibly interconnecting multiple input ports of a node with multiple output ports while choosing between wavelength channels, splitting ratios, and power levels [29.55]. In [29.56], this generic approach was adapted to a $1 : N$ branching architecture for passive access networks (Fig. 29.28). The optical switches are of

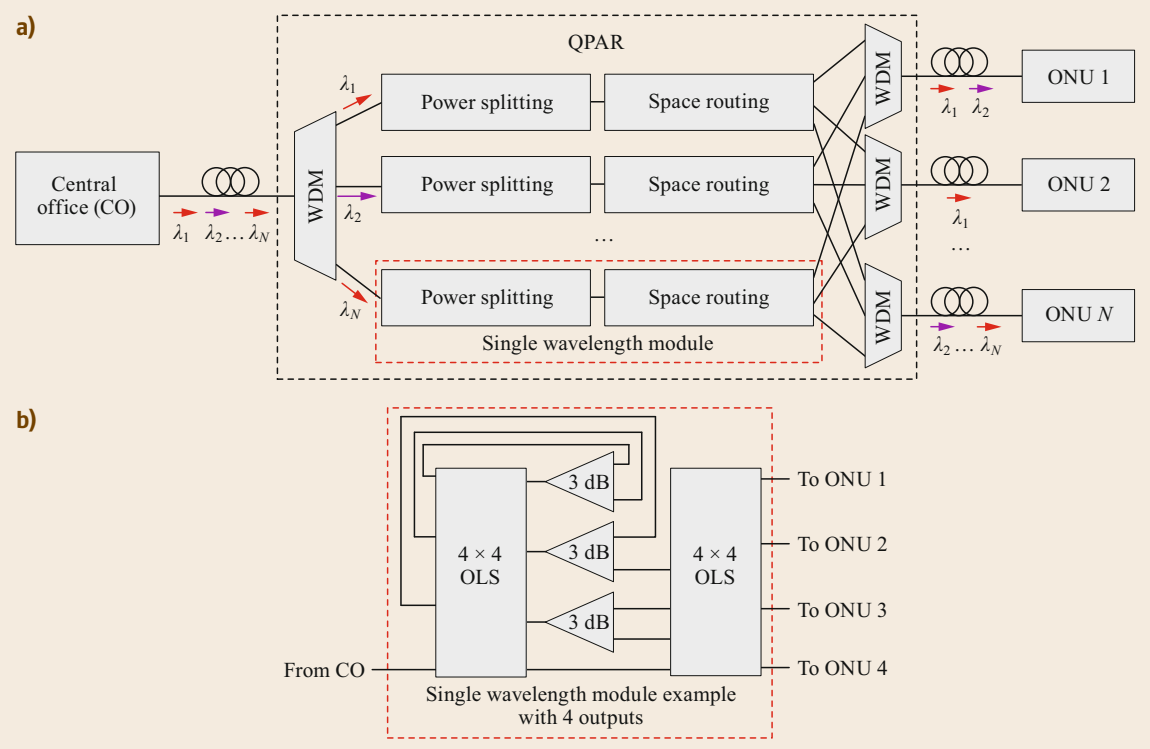


Fig. 29.28a,b Reconfigurable $1:N$ branching node employing a modular QPAR architecture: **(a)** generic structure, **(b)** example of a single-wavelength module in a 1 : 4 configuration (courtesy of Ke Wang)

latching type (OLS) to avoid energy consumption when not being switched.

The generic QPAR structure is very flexible and allows choosing between many branching configurations. However, the component count can be high, resulting in a non-negligible insertion loss when the optical signals pass through multiple elements. If instead, full flexibility is not needed, but a certain set of predefined configurations is sufficient, then a second concept holds promise in terms of low insertion loss and compact implementation in future sBNs.

Multiplane light convertors (MPLC) are passive devices that by way of successively repeated (static) optical phase modulation and Fourier transform allow lossless conversion of transverse optical modes [29.57]. This can be used for realizing an optical power splitter by transforming the mode exiting from a single-mode fiber at the input into a pattern of multiple modes side by side for an array of single-mode fibers at the output. The transform can be designed to distribute the input optical power across the output fibers in predefined ratios that change with the angle and position of the input light beam hitting the MPLC (Fig. 29.29), as well as with the wavelength. Current concepts promise close to zero insertion loss across an entire waveband with splitting ratios varying between bands, e.g., an O-band compared to a C-band [29.58]. With more sophisticated designs, this characteristic may in future become available even on the scale of individual wavelength channels. For application in an sBN, the free-space beam steering element shown in Fig. 29.29 can be replaced with a $1 \times N$ latching switch in combination with fiber segments attached to the MPLC, for realizing the required variations of angles and positions at the input.

29.5.4 Remote Optical Powering and Energy-Efficient Node Management

In many ODN deployments the remote nodes are integrated in, e.g., street cabinets with access to local electrical power supplies. From a power supply perspective, operating functionalities such as those described above are then easily accomplished. The communication between the remote node and the CO can be facilitated using an ONU inside the sBN, operated as just another end node of the PON. In the case of complete separation of the optical layer management from the PON

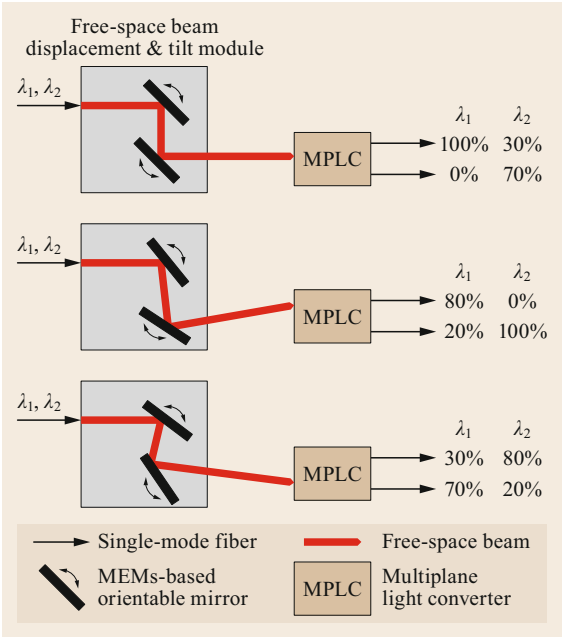


Fig. 29.29 Reconfigurable $1 \times N$ splitting/wavelength routing element by choosing different input angles and positions of a light beam hitting an MPLC (courtesy of Cailabs, Rennes). *MEMS* micro-electromechanical system

transmission system (Fig. 29.27), the OMM can host a dedicated OLT for this communication. A more cost-efficient approach for this low-bandwidth management channel will be described below, along with optical powering of the remote nodes.

In some network deployments, the remote nodes may be implemented in places without access to local power supplies, e.g., in manholes or in aerial cable infrastructures. Also for the active DPM, an independent source of energy is preferable to remain independent from the customer's power supply. The node architectures introduced in the previous sections already contained means for accommodating such scenarios by way of remote optical powering and energy-efficient communication. This approach will be described in more detail below.

For simplifying the description, DPM and sBN will collectively be termed sRN (smart remote node).

The energy-efficient smart ODN architecture is enabled by two functional blocks in the OMM and in the sRN, respectively (Fig. 29.30):

- The OMM is the central element for powering and communicating with the sRNs. It contains the OMM control and communication unit and the optical powering source. The OMM control and communication unit is the gateway towards the OSS on one side and provides an optical communication channel to the sRNs on the other side.
- The sRN contains an sRN optical powering unit that converts the optical energy provided by the powering source in the OMM into electrical energy for the sRN. It also manages the storage of the energy and its consumption by the node-integrated functions. The sRN control and communication unit receives instructions from the OSS via the OMM and returns monitoring data or status reports in the reverse direction.

In addition to these basic functionalities, there are smart optical functions included for monitoring or reconfiguration, as described in the previous sections (here, generically indicated by the green block).

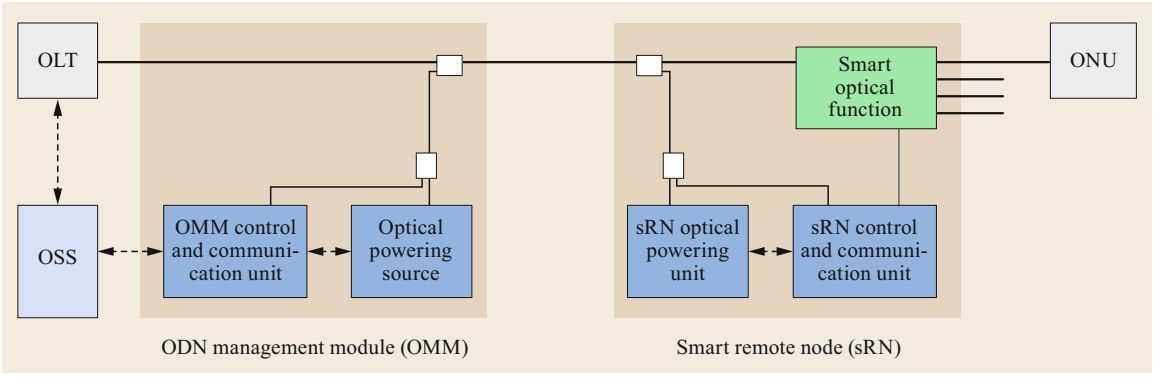


Fig. 29.30 Generic building blocks for optical powering, communication and control in smart ODNs (*dark blue*: basic functions for powering, communication, and control; *green*: auxiliary smart optical functions; *white*: wavelength diplexers or tap couplers). There may be a single or multiple fibers exiting from the sRN, depending on whether it is a DPM or an sBN, respectively

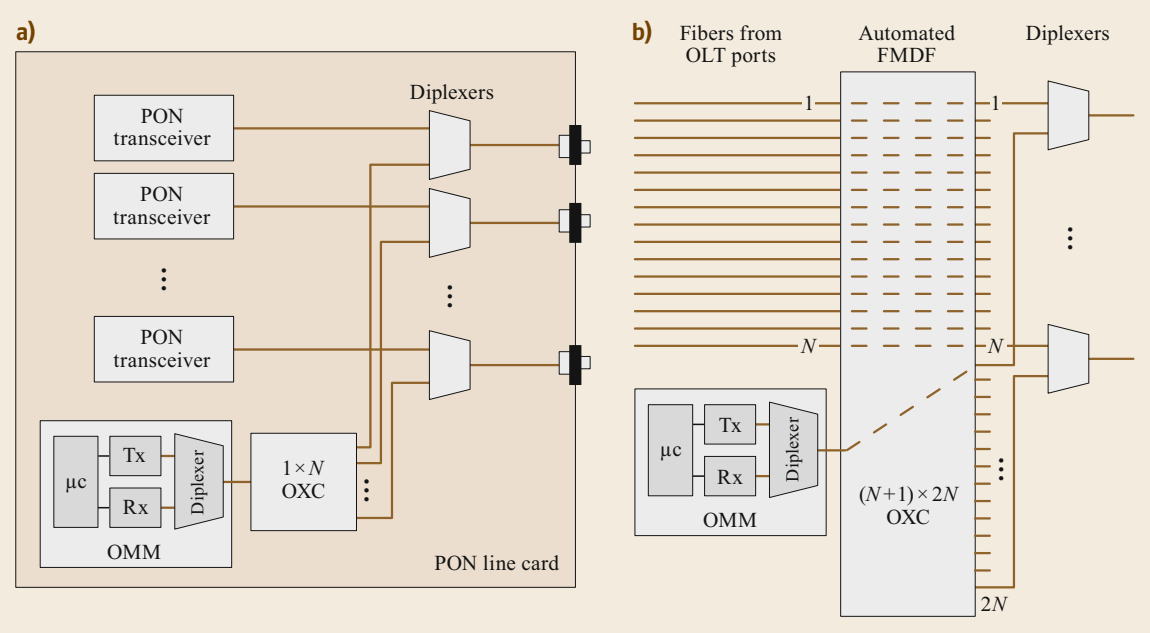


Fig. 29.31a,b OMM sharing in the central office: one OMM per line card **(a)** or per FMDF **(b)**. (The OXC on the *right* does not support independent switching of the optical communication and recharging channel)

The sRN electronics and optics are made from ultra-low power electronics, such as available for sensor applications. Similarly, low-power optics are used for the communication channel and for the smart optical functions. To meet those low-power requirements, the communication is typically accomplished at low bit rates (in the range of kb/s to few Mb/s), sending only short messages (< 100 bytes). The smart optical functions may include power detectors, optical switches, or other means to reconfigure optical connections, to name a few. The energy consumption of these functions should be reduced to a minimum, and they will be activated only for short time intervals after long intervals of inactivity, during which the local energy storage (battery or supercapacitor) can be recharged.

The PON transmission system equipment in the CO is usually hosted in racks, which in turn hosting line cards with multiple OLT ports. Providing a dedicated OMM for each OLT port, as seemingly suggested by Figs. 29.27 and 29.30, would be over-dimensioned and too costly. The OLM functionalities in the remote nodes do not have to be invoked continuously or simultaneously on each ODN at the same time. They must be available only for occasional ODN status updates, for occasional reconfigurations, or for providing access to redundant resources in the case of ODN malfunctions and repairs. Hence, a single OMM per line card or per rack can be shared across multiple ODNs by way of flexibly connecting it to any ODN as required, using a fiber optic crossconnect (OXC) that is positioned on the line card (Fig. 29.31a) or as part of an automated fiber main distribution frame (FMDF) (Fig. 29.31b).

The optical channel for remote powering, as well as the channel for communication between the OMM and the remote nodes, can be allocated in the auxiliary wavelength range 1380–1460 nm (Fig. 29.16). The optical powering channel can, in principle, also be used for downstream communication by applying on/off keying modulation, for instance. However, this coupling of powering and communication would not allow an efficient and flexible operation of the smart ODN concept. Since both processes, powering and communication, take place on different time scales (several minutes for remote powering versus milliseconds or seconds for communication), the concept would benefit from serving both processes on separate wavelength channels. This allows the OMM to communicate with those ODNs that are in the measurement phase, while others are in the recharging phase.

In this scenario, a set of three wavelengths is needed: one wavelength (λ_{OLM_p}) for powering the sRN, one for downstream communication (λ_{OLM_DS}), and one for upstream communication (λ_{OLM_US}) with the OMM. In cascaded split ODNs (e.g., ISMs in Fig. 29.26), there are multiple ways possible for assigning and using these wavelengths:

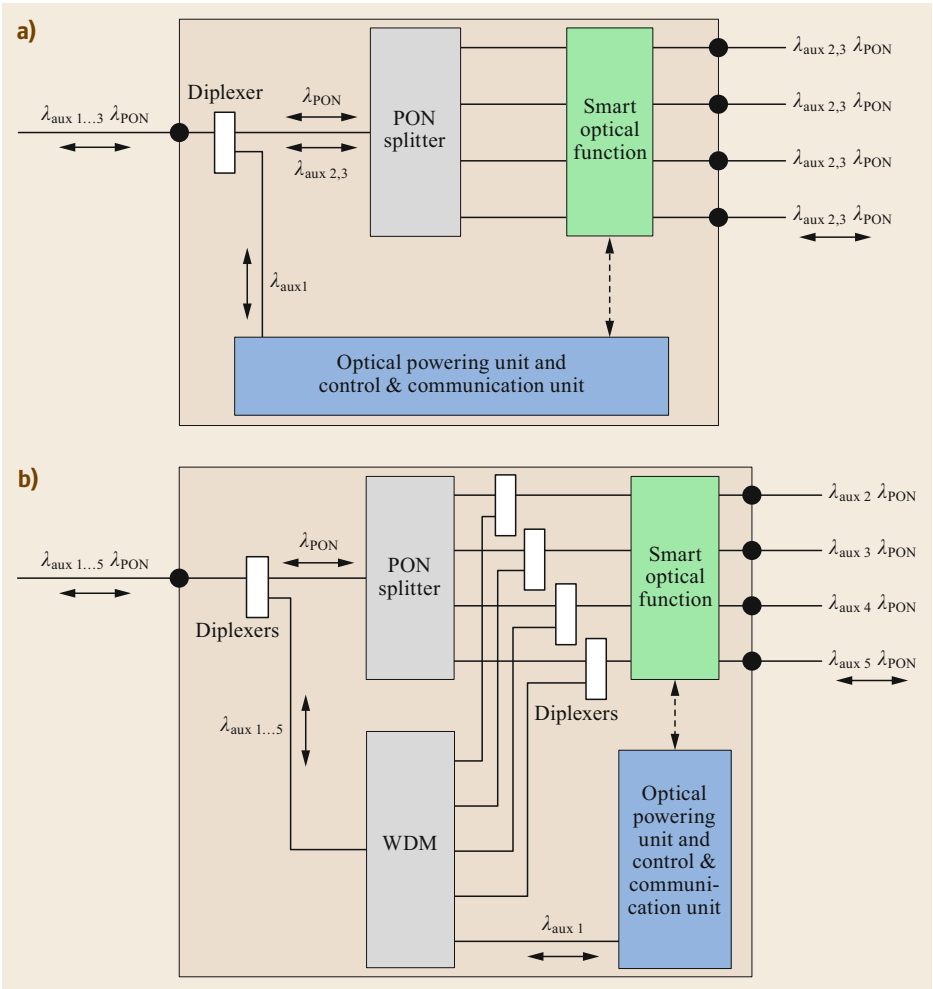


Fig. 29.32a,b Node architectures for terminating OLM wavelength channels per stage **(a)** or per node **(b)**; λ_{aux} stands for a set of OLM wavelengths (see text), λ_{PON} for the regular communication channel pairs of the PON transmission system; the functional units are generically displayed as one common block (*blue* for electrical, and *green* for optical functionalities)

- One wavelength set for the entire ODN:
  The powering and control and communication units of each intermediate sRN are connected to the transmission fiber via a tap coupler (e.g., 10%), while only the last node (e.g., DPM) terminates the OLM wavelengths via a diplexer. In this arrangement, the OLM wavelengths pass though the regular power splitter along with the PON channel pairs λ_{PON}, so that the nodes at later stages will receive less power than the ones at preceding stages. This can be mitigated to some extent by using different tap ratios for each stage.
- One wavelength set for each stage of sRNs:
  The powering and control and communication units of each sRN at a stage are connected to the transmission fiber via a diplexer terminating the OLM wavelength channels for this particular stage. In the preceding stages, those channels are either power split or wavelength routed to pass through the preceding nodes (Fig. 29.32a, power splitting).
- One wavelength set for each sRN:
  The powering and control and communication units of each sRN are connected to the transmission fiber via a diplexer and terminate the OLM channels assigned to this particular sRN. In the preceding stages, the wavelength channels have either been power split or wavelength routed to bypass the preceding nodes (Fig. 29.32b, wavelength routing).

Each of the above node architectures may be useful in particular ODN architectures. It might even be useful to choose a different architecture for the powering channel as compared to the communication channels. In any case, it is apparent that an individual wavelength set per node would be a very challenging solution in view of the total cost. Even bypassing the preceding nodes via

WDM and diplexers, as shown on the right-hand side in Fig. 29.32, increases the hardware effort beyond what may be acceptable for a *simple* management solution. So, some sort of wavelength sharing among multiple nodes is advisable.

The communication between OMM and sRN needs to be very power efficient in the case of remotely powered nodes. Multiple design factors must contribute to that: low optical transmit powers from the sRN, low bit rates, short messages, sleep mode of the sRN, and an energy-efficient communication protocol. Unless each sRN is addressed over a unique pair of communication wavelengths (which is not advised in view of the associated technical effort and cost, see above), multiple sRNs will share a common wavelength pair employing a simple multiple access protocol based on time division multiple access. It is, however, not efficient to use an advanced protocol such as used in GPON or EPON. The energy needed for running such complex protocols would exceed the energy available from the reservoir in the sRN. Instead, a very simple energy-efficient protocol can be used for this purpose (low-energy MAC protocol, [29.59]). With this protocol, the sRNs occasionally wake up from sleep mode (after minutes) and then listen to the downstream OLM channel for beacon signals emitted from the OMM at short intervals. Once an sRN detects such a beacon, it falls into snooze mode, during which a counter is activated that ensures that the sRN wakes up for communication at a scheduled time previously announced with the beacon. During the scheduled window, those sRNs that had detected the beacon will wake up fully, ready for communication. The OMM asks one sRN at a time for short communication during this time window, while all others remain quiet.

References

29.1 ITU-T: Recommendation G.983.x series on broadband optical access systems based on Passive Optical Networks (PON) (2001–2007)

29.2 ITU-T: Recommendation G.984.x series on Gigabit-capable Passive Optical Networks (GPON) (2003–2014)

29.3 IEEE: Standard for Ethernet, IEEE Std 802.3-2018 (2018)

29.4 ITU-T: Recommendation G.987.x series on 10-Gigabit-capable Passive Optical Networks (XG-PON) (2012–2017)

29.5 ITU-T: Recommendation G.9807 10-Gigabit-capable Symmetric Passive Optical Network (XGS-PON) (2016–2017)

29.6 ITU-T: Recommendation G.989.x series on 40-Gigabit-capable passive optical networks (NG-PON2) (2013–2017)

29.7 ITU-T: Recommendation G.9802.x series on Multiple-wavelength passive optical networks (MW-PONs) (2015)

29.8 ITU-T: Supplement G.Sup51 Passive optical network protection considerations (06/2017)

29.9 ITU-T: Supplement G.Sup49 Rogue optical network unit (ONU) considerations (02/2011)

29.10 B. Le Guyader, W. Poehlmann, F. Saliou, L. Jentsch, L. Guillo, P. Chanclou, T. Pfeiffer: Electrical splitting OEO G-PON reach extender demonstration, Paper P.6.8. In: *Proc. Eur. Conf. Opt. Commun.* (2013)

29.11 S. Koenig, R. Bonk, R. Schmogrow, A. Josten, D. Karnick, H. Schmuck, T. Pfeiffer, C. Koos, W. Freude, J. Leuthold: Cascade of 4 SOAs with 448 Gbit/s (224 Gbit/s) dual channel dual polarization 16QAM (QPSK) for high-capacity business paths in converged metro-access networks, Paper OTh4A.3. In: *Proc. Opt. Fiber Conf.* (2013)

29.12 Nokia: 5G new radio network, White Paper (2019)

29.13 T. Pfeiffer: Next generation mobile fronthaul and midhaul architectures, J. Opt. Commun. Netw. **7**, B38–B45 (2015)

29.14 3GPP: Technical Report TR 38.801 V14.0.0 (2017–03), Study on new radio access technology: radio access architecture and interfaces (release 14) (2017)

29.15 IEEE: Standard for packet-based fronthaul transport networks, IEEE 1914.1 (will be published in 2020)

29.16 ITU-T: Supplement G.Sup66, 5G wireless fronthaul requirements in a passive optical network context (07/2019)

29.17 Small Cell Forum: SCF DOC 159.07.02 Small cell virtualization functional splits and use cases, release 7.0 (2016)

29.18 CPRI Cooperation: CPRI specification Common Public Radio Interface (CPRI); interface specification, CPRI specification V7.0, 2015-10-09 (2015)

29.19 CPRI Cooperation: CPRI specification Common Public Radio Interface: Requirements for the eCPRI Transport Network, eCPRI Transport Network V1.2, 2018-06-25 (2018)

29.20 U. Doetsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, P. Sehier: Quantitative analysis of split base station processing and determination of advantageous architectures for LTE, Bell Labs Tech. J. **18**, 105–128 (2013)

29.21 ITU-T: Recommendation G.698.4 Multichannel bi-directional DWDM applications with port agnostic single-channel optical interfaces (2018)

29.22 ITU-T: Recommendation L.66, Optical fibre cable maintenance criteria for in-service fibre testing in access networks (05/2007)

29.23 R. Bonk, W. Poehlmann, D. van Veen, J. Galaro, R. Farah, H. Schmuck, T. Pfeiffer: The underestimated challenges of burst-mode WDM transmission in TWDM-PON, Opt. Fiber Technol. **26**, 59–70 (2015)

29.24 H. Debrégeas, R. Brenot, J.-G. Provost, S. Barbet, W. Pöhlmann, R. Borkowski, R. Bonk, T. Pfeiffer: Quasi frequency drift suppression for burst mode operation in low-cost thermally-tuned TWDM-PON, Paper Th5A.5. In: *Opt. Fiber Commun. Conf.* (2017)

29.25 CPRI Cooperation: CPRI Specification Common Public Radio Interface: eCPRI Interface Specification eCPRI specification V2.0, 2019-05-10 (2019)

29.26 IEEE: Standard for radio over Ethernet encapsulations and mappings, IEEE Std 1914.3-2018 (2018)

29.27 O-RAN Fronthaul Working Group, O-RAN Alliance: Technical specification ORAN-WG4.CUS.0-v03.00, Control, user and synchronization plane specification (2020)

29.28 S. Bidkar, J. Galaro, T. Pfeiffer: First demonstration of an ultra-low-latency fronthaul transport over a commercial TDM-PON platform, Paper Tu2K.3. In: *Opt. Fiber Commun. Conf.* (2018)

29.29 T. Tashiro, S. Kuwano, J. Terada, T. Kawamura, N. Tanaka, S. Shigematsu, N. Yoshimoto: A novel DBA scheme for TDM-PON based mobile fronthaul, Paper Tu3F.3. In: *Opt. Fiber Commun. Conf.* (2014)

29.30 R. Bonk, R. Borkowski, M. Straub, H. Schmuck, T. Pfeiffer: Demonstration of ONU activation for in-service TDM-PON allowing uninterrupted low-latency transport links, Paper W3J.4. In: *Opt. Commun. Conf.* (2019)

29.31 Open Networking Lab, AT&T: White paper central office re-architectured as data center, June (2015)

29.32 NTT: White paper flexible access system architecture (FASA), v2.0, March (2017)

29.33 ONAP: White paper, Open network automation platform (ONAP) architecture, http://www.onap.org (2018)

29.34 M.K. Weldon: *The Future X Network: A Bell Labs Perspective* (CRC, Boca Raton 2015)

29.35 F. Effenberger: PON resilience, J. Opt. Commun. Netw. **7**(3), A547–A552 (2015)

29.36 T. Pfeiffer: Converged heterogeneous optical metro-access networks, Paper Tu.5.B.1. In: *Eur. Conf. Opt. Commun.* (2010)

29.37 M. Al-Fares, A. Loukissas, A. Vahdat: A scalable, commodity data center network architecture, ACM SIGCOMM Comput. Commun. Rev. **38**(4), 63–74 (2008)

29.38 T. Benson, A. Akella, D.A. Maltz: Network traffic characteristics of data centers in the wild. In: *Proc. Internet Meas. Conf. (IMC)* (2010) pp. 267–280

29.39 J. Li, J. Chen: Passive optical network based mobile backhaul enabling ultra-low latency for communications among base stations, J. Opt. Commun. Netw. **9**, 855–863 (2017)

29.40 C. Choi, Q. Wei, T. Biermann, L. Scalia: Mobile WDM backhaul access networks with physical inter-base-station links for coordinated multipoint transmission/reception systems. In: *IEEE Global Telecommun. Conf. GLOBECOM* (2011), https://doi.org/10.1109/GLOCOM.2011.6133510

29.41 A.A.M. Saleh, H. Kogelnik: Reflective single-mode fiber-optic passive star couplers, J. Lightwave Technol. **6**, 392–398 (1988)

29.42 Y. Cheng, M. Fiorani, R. Lin, L. Wosinska, J. Chen: POTORI: a passive optical top-of-rack interconnect architecture for data centers, J. Opt. Commun. Netw. **9**, 401–411 (2017)

29.43 T. Pfeiffer: Distributing system for the communication between distributed base stations, and distributing unit therefor, Patent pending, EP2348787A1 (2010)

29.44 T. Pfeiffer: A physical layer perspective on access network sharing, Opt. Fiber Technol. **26**, 12–20 (2015)

29.45 Y. Enomoto, H. Izumita, M. Nakamura: Over 31.5 dB dynamic range optical fiber line testing system with optical fiber fault isolation function for 32-branched PON, Paper ThAA3. In: *Opt. Fiber Commun. Conf.* (2013)

29.46 T. Pfeiffer, H. Schmuck, M. Straub, J. Hehmann: Cost efficient non-service interrupting monitoring of optical fiber links in FTTH / FTTB networks, Paper Tu.1.F.2. In: *Eur. Conf. Opt. Commun.* (2008)

29.47 A. Ehrhardt, L. Schuerer, F. Escher, B. Nagel, H.-M. Foisel: ONT reflection for additional maintenance by OTDR measurements in FTTH networks, Paper JW2A.08. In: *Opt. Fiber Commun. Conf.* (2013)

29.48 C.-K. Chan, F. Tong, L.-K. Chew, K.-P. Ho, D. Lam: Fault surveillance of branched optical networks using an amplifier-generated wavelength-sweeping monitoring source, Paper ThE2. In: *Opt. Fiber Commun. Conf.* (1999)

29.49 Y. Li, D. Wang, J. Li: FTTH remote fiber monitoring using optical wavelength domain reflectometry (OWDR) and wavelength coded tag (WCT), Paper OThU3. In: *Opt. Fiber Commun. Conf.* (2006)

29.50 J. Hehmann, T. Pfeiffer: New monitoring concepts for optical access networks, Bell Labs Tech. J. **13**, 183–198 (2008)

29.51 J. Hehmann, M. Straub, L. Jentsch, M. Earnshaw, P. Anthapadmanabhan, T. Pfeiffer: Remotely powered intelligent splitter monitor for fiber access networks, Paper Tu.1.5.4. In: *Eur. Conf. Opt. Commun.* (2015)

29.52 M. Straub, L. Jentsch, J. Hehmann, T. Pfeiffer, R. Bonk: Remotely powered Inline OTDR unit with unique identification possibility of power splitter branches for use in access network applications, Paper we2.68. In: *Eur. Conf. Opt. Commun.* (2018)

29.53 Sercalo: MEMS optical switch SXLA2x2SMF, http://www.sercalo.com

29.54 B. Schrenk, M. Hofer, M. Hentschel, T. Zemen: Semi-passive power/wavelength splitting node with integrated spectrum monitoring for reconfigurable PON, Paper Th2A.30. In: *Opt. Fiber Commun. Conf.* (2017)

29.55 K. Wang, A. Gowda, Y. Bi, L.G. Kazovsky: Multi-dimensional quasi-passive reconfigurable (MDQPAR) node for future 5G optical networks, Paper Tu2K.2. In: *Opt. Fiber Commun. Conf.* (2017)

29.56 K. Wang, Y. Bi, A. Gowda, L.G. Kazovsky: Bidirectional quasi-passive reconfigurable (Bi-QPAR) remote node for future optical access networks, J. Lightwave Technol. **35**, 2109–2117 (2017)

29.57 G. Labroille, B. Denolle, P. Jian, P. Genevaux, N. Treps, J.-F. Morizur: Efficient and mode selective spatial mode multiplexer based on multi-plane light conversion, Opt. Express **22**(13), 15599–15607 (2014)
29.58 Cailabs: Variable-ratio optical splitter for controlling the proportion of energy of a light beam, Patent pending, W02018/134534 A1 (2018)
29.59 M. Roeger, B. Hiba, J. Hehmann, M. Straub, H. Schmuck, M. Hedrich, T. Pfeiffer, C. Koos, J. Leuthold, W. Freude: In-service monitoring of PON access networks with powerline independent devices, J. Opt. Commun. Netw. **6**, 1018–1027 (2014)

Thomas Pfeiffer

Bell Labs
Nokia
Stuttgart, Germany
thomas.pfeiffer@nokia-bell-labs.com

Thomas Pfeiffer received his PhD in 1986 from the Max-Planck-Institute for Solid State Physics and the University of Stuttgart. He then joined Alcatel Research Center, working on technologies for optical transmission systems and networks. He led the Global Optical Access team at Bell Labs for 15 years before becoming Principal Architect for Optical Access Networks at Nokia Bell Labs in 2017. He received the 2019 Bell Labs Fellows Award, and became an OSA Fellow Member in 2020.

30. Long-Reach Passive Optical Networks and Access/Metro Integration

David Payne, Giuseppe Talli, Marco Ruffini, Alan Hill, Paul Townsend

The physical layout of cables and nodes in many of today's passive optical networks (PONs) still dates back to the early days of copper loop installations, with customer to exchange node distances limited to a few kilometers largely by the transmission distance of analogue telephony over the copper pairs. However, modern optical communication technologies can enable much longer transmission distances, and through the use of optical amplification can effectively integrate the access and metro portions of the network into a single all-optical communication system, which is commonly referred to as "Long-reach PON" (LR-PON) or "Amplified PON". LR-PONs offer several advantages in terms of infrastructure sharing, network node consolidation and core network delayering.

This chapter describes the history of the development of LR-PONs, the technical design and the enhancements that can be added, such as flexible or dynamic wavelength assignment and the benefits for the end to end network architecture when the LR-PON is used to its full capability. We use results from recent research projects to illustrate the advantages of changing the overall network architecture to enable much higher sustained user bandwidths while reducing power consumption per user and improving economic viability. We also review recent experimental demonstrations of the end-to-end operation of such systems which validate the viability of these concepts using currently available components and technologies.

The physical layout of cables and nodes in today's fixed network dates back to the early 1900s when the copper line infrastructure was beginning to be installed. Customer to exchange node distances were determined largely by the transmission distance of analogue telephony over the copper pairs that could be used for economically viable telephone cables. These constraints limited typical distances to ≈ 2 km in denser areas where larger pair count cables were required to serve the higher density of customer locations. In more rural areas these distances could be extended to 5–7 km with careful design and more expensive technology, for example using higher gauge copper wire and *loading coils*.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_30

30.1 The History of LR-PON

These distance constraints continue to the present day and will continue until the copper technology is replaced with a longer reach technology, with optical fibers being the obvious choice for that replacement technology. The advent of broadband using digital subscriber line (DSL) technology over the copper pair infrastructure has made the distances even more constrained. To further increase bandwidths, via DSL technologies, active nodes need to be placed even closer to the customers in locations such as the street cabinets (primary cross-connect points—PCPs) for VDSL (very high speed digital subscriber line) and even the distribution points (DPs) for the latest G.fast technologies.

The copper distance constraints produced the network architecture we have today, with thousands of active traffic processing nodes being required for typical large European-sized countries with more than 40 million population. The capital and operational cost of equipping, upgrading, and maintaining these nodes is a huge economic burden for fixed network operators. Since the late 1970s, optical fiber has been considered the ideal choice for replacing this copper network, with the much greater reach of fiber seen as the enabler for significant node consolidation.

Node consolidation is where small nodes are combined onto larger nodes further within the network to significantly reduce the total count (Fig. 30.1). Node consolidation has been an on-going consideration by operators as optical fiber technology approached economic viability for access networks. The reason it has not yet happened to any significant extent is because of the legacy services running over copper cables. All these services would need to be transferred to fiber technology so that the copper cables could be removed, or at least decommissioned, enabling closure of the local node, but this would mean a full FTTH/P (fiber to the home/fiber to the premises) roll-out in the nodes being targeted which may not suit all users of the node.

Local loop unbundling (LLU), was introduced in the late 1990s and extensively used in the early 2000s to foster greater competition, particularly for the provision of broadband services. Unfortunately LLU has also inadvertently acted as a barrier to fiber deployment in the access network and the extensive node consolidation it can enable. This is because there can now be multiple competing players with interests and possibly equipment in the existing local exchanges (LEs) or central offices (COs) and closing these down would require all licensed operators using the building to agree on the closure and on moving to optical technology for their access connections [30.1]. The degree of consolidation would also be determined by the capabilities and the design of the optical technology chosen and different operators may want different performance and capabil-

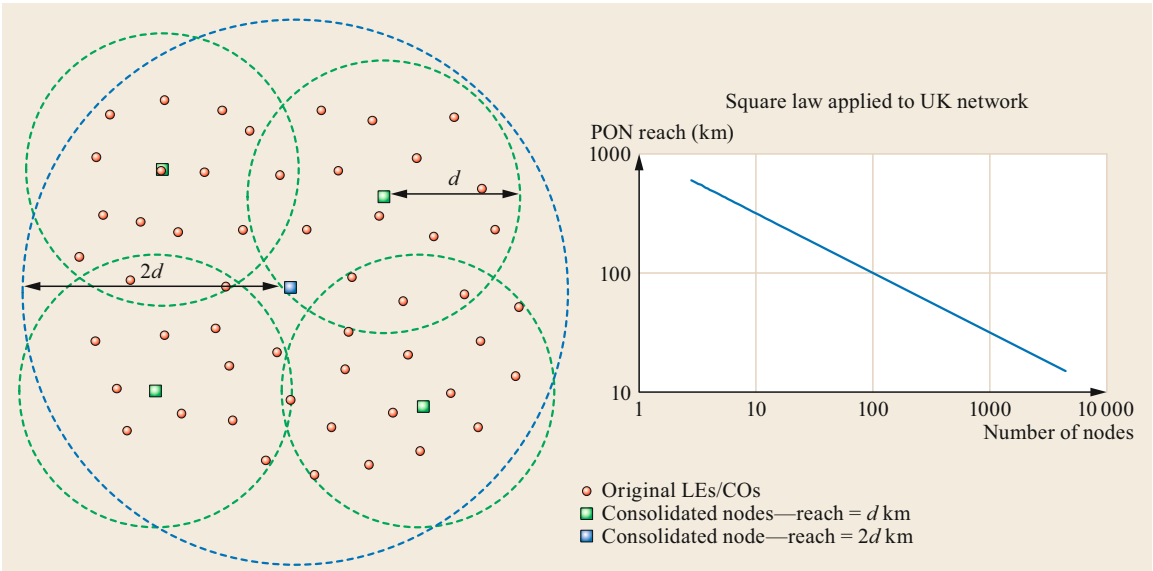


Fig. 30.1 By using longer reach access technology such as fiber to the premises (FTTP) the LEs/COs can be consolidated onto larger and more remote nodes. The degree of node consolidation is approximately a square law, so doubling the access network reach quadruples the number of nodes that can be consolidated onto a single node. The *inset chart* shows this law fitted to the UK network

ities. So the scale of future node consolidation will be determined as much by business models and regulation as technology development.

However, there is a growing move towards greater deployment of optical access networks and the passive optical network (PON) has a particular advantage for node consolidation due to the reduction in fiber cable size via the splitter points as the optical cable network approaches the consolidated node. The savings in cable cost becomes more significant as reach increases and backhaul or feeder network cable routes become longer. The PON solution therefore gives a significant advantage both in terms of capital and operational costs as the reach increases, enabling greater node consolidation with greater savings in cabling costs and greater potential for reducing the number of switching/routing nodes in the network, reducing equipment and energy costs.

The short reach of copper technologies, and the need to provide full coverage of the populated land area, means that LEs/COs are relatively uniformly distributed in populous countries, such as those in Europe. It follows therefore that the number of exchanges that can be consolidated onto a larger central node increases approximately as the square of the geographical reach of the access network. This relationship is illustrated in Fig. 30.1 which also shows the approximate degree of consolidation for the UK network as a function of the PON reach.

The square-law curve for the UK in Fig. 30.1 was normalized to results from optimization modeling of the number of nodes required for the UK mainland. It showed that with 100 km reach, 100 nodes would be required (56 : 1 node reduction) and for a 125 km reach 73 nodes would be required (76:1 node reduction). These results are very close to the simple square-law relationship shown in Fig. 30.1. It should be noted that the square-law relationship is an approximation, particularly for shorter reaches where density variations of customer sites and edge effects have a larger impact. Also for amplified PONs with long reach and greater split, protection paths must be provided such as dual parenting each original LE/CO site onto two consolidated nodes. This is required to maintain network availability but limits the effective area covered by a long-reach PON of a given reach; this reach limit is included in the curve in Fig. 30.1.

Clearly, pushing the theoretical limit of the long reach, all LEs/COs could be consolidated onto one vast node, although such a degree of consolidation would be impracticable, with serious security and reliability issues. However, it does raise the question of what degree of node consolidation and what reach of access network should be targeted. Answering this question requires an end-to-end consideration of the network architecture as the greater the scale of node consolidation the greater the opportunity for restructuring the access, metro, and core networks to make them more capable, lower cost, and have lower environmental impact by reducing use of material resources and energy requirements. It should be stressed here that long-reach PONs are an enabler for network restructuring, it is much more than an alternative access technology, and using LR-PON as an enabler for network restructuring is an important aspect of the work described in this chapter.

Using the simple square-law relation in Fig. 30.1 the degree of node consolidation for various example PON solutions, either commercially available being developed in standards, or proposed in projects, is compared in Table 30.1. This is not an exhaustive list but serves to illustrate the range of options and the potential for node consolidation and network restructuring. Although the numbers in Table 30.1 are an example and are specific to the UK, the principles apply to all countries that have telecommunications networks originally based on the reach of copper access technologies, which is typical of the developed world.

30.1.1 Historical Development of LR-PONs

When passive optical networks were invented in the early 1980s they were called *shared access networks* rather than *passive optical networks*. The objective at that time was to find a way to reduce the cost of optical fiber access networks and avoid all the problems operators had with large-pair-count copper cables which used the classic point-to-point (Pt-Pt) architecture for the access network. The huge bandwidth available on fiber compared to copper meant that this capacity could be shared over a number of customers. Power splitters were being developed (as well as WDM technologies) and sharing the fiber via the use of optical splitters was a way of significantly reducing the optical access network costs compared to a like-for-like replacement of the copper network with Pt-Pt fiber technology. This was also the time when the switch from multimode-fiber technology to single-mode fiber was occurring and, at least for telecommunications networks, this enabled greater transmission distances due to the lower loss of single-mode fiber and removal of mode dispersion, which was a major limitation of multimode-fiber transmission distance. Also single-mode fiber splitters were more uniform and had lower excess loss compared to multimode-fiber devices. The difficult problem of splicing and connectors for single-mode fiber had been solved in the late 1970s/early 1980s, so this was no longer a barrier to single-mode fiber deployment and single-mode fiber became the technology of choice for PON networks.

Table 30.1 Reach and node consolidation comparison of PON systems

PON system	Reach (km)	No. consolidated nodes	Level of network restructuring
GPON (Gigabit-capable PON)	20	2500	Limited reach—generally not used for node consolidation
GPON ex	60	278	Significant access-metro convergence possible—applicability limited by small split (high feeder fiber cabling costs)
EPON (Ethernet PON)	10	10 000	Too short for any network restructuring
10 Gb/s EPON	20	2500	Limited reach—generally not used for node consolidation
XGPON (10-Gigabit-capable PON)	20	2500	Limited reach—generally not used for node consolidation
NG-PON2 (next-generation PON2)	40	625	Some access-metro convergence
OASE LR-PON[a]	100	100	End-to-end restructuring—access-metro convergence + simplified core
DISCUS LR-PON[b]	125	73	End-to-end restructuring—access-metro convergence + simplified core + sparse rural capability
LR-WDM-PON[c]	80	156	End-to-end restructuring—access-metro convergence + some simplification of core

[a] OASE compared a number of access architectures but had 100 km reach with high split (≈ 1000 way) for the LR-PON option
[b] DISCUS had two LR-PON architectures, one a standard LR-PON with a single amplifier node (AN) and a second with even longer reach for sparse rural areas where the ANs were in a chain or open-ring topology (Sect. 30.1.1)
[c] LR-WDM-PON is a long-reach WDM-POM that uses wavelength-division multiplexing (WDM) devices rather than power splitters and can serve as a metro-access convergence network

The term *passive optical network* was adopted later to emphasize the difference, and advantages, when compared with a competing fiber access technology that used an active star topology with an opto-electronic node placed at the cabinet location. From this active cabinet a Pt-Pt fiber distribution network went to the customer premises while a feeder fiber pair went back to the exchange. This active cabinet (which ironically is now where the VDSL node is placed for *superfast* broadband over copper pairs) was replaced by a passive optical splitter in the PON solution and hence the term passive optical network came about and has stuck. It should be stressed that the passive part of a PON was only the fiber network from the local exchange to customer, that is, the traditional access network and what we now call the optical distribution network or ODN. The LE/CO contains significant quantities of optical and electronic processing equipment and conventional PONs terminate on the LE/CO, leaving this equipment in place. The LR-PON architecture however bypasses the LE/CO and replaces the equipment with small, low power consumption, optical amplifiers which could be placed in the local cable chamber or a street cabinet, facilitating closure of the buildings housing the equipment in these LE/CO nodes, while the ODN remains passive with only splitter components and no active cabinets.

The reach of a passive optical network is limited by the available optical power budget which has to be shared across the $10\log(n)$ dB splitting loss plus excess splitter losses of order $0.5\log_2(n)$ dB and the fiber loss. In early commercial PONs a typical split ratio of $n = 32$ was targeted which gives a splitter loss of approximately 17.5 dB. Early PONs predominantly used the 1300 nm window and a spliced cable loss of around 0.45 dB was often assumed. With a ≈ 28-dB power budget available (at that time) and a 3 dB margin then there was only about 9 dB for fiber loss and terminating connectors which gave an effective reach of approximately 15 km. Although 15 km was a significant increase over the reach of copper technology and could have enabled some moderate level of node consolidation, early PON systems were deployed on a piecemeal basis and exchange closure was generally not an option, so very little node consolidation occurred. The relatively short reach of these early PONs meant that the degree of node consolidation was always going to be fairly limited; what was needed was a more aggressive and targeted FTTP roll out and a much longer geographical reach PON. This meant PONs with much greater power budgets being required.

By the end of the 1980s erbium-doped fiber amplifiers (EDFAs) were becoming practical tools for optical communications systems and work with amplified PONs was being carried out at major laboratories around the world [30.2, 3].

These early experiments demonstrated the potential of amplified PONs but a more pragmatic bidirectional solution was needed by operators. A research program was set up in BT looking at a practical LR-PON solution that could provide a significant level of node consolidation while delivering the current telephony and foreseen future broadband services. This work focused on an LR-PON with up to 100 km reach, up to 3500-way split, operating in the downstream direction at 1.2 Gb/s and upstream at 300 Mb/s. This work is mentioned because

it directly led to the first EU project addressing long-reach PONs, the EU ACT *PLANET* project [30.4]. This project targeted an amplified PON with 2.5 Gb/s downstream, 311 Mb/s upstream, a split of 2048 ways and reach of 100 km. It should be noted that the term Long Reach or LR-PON was not used at this time and amplified PONs were instead called *SuperPONs*. This term was used to emphasize that the power budget enhancement obtained from the use of optical amplifiers was used for increasing the optical split and the system bitrates as well as the physical reach. The PLANET project successfully delivered a working *SuperPON* demonstration [30.5] meeting its target objectives.

Another EU project called SONATA explored the ultimate limits of SuperPON architectures with massive two-way PONs having up to 50 000 terminals per PON [30.6]. To accommodate the huge split in the upstream direction gated optical amplifiers were used to minimize the noise contributions from amplifiers and optical network units (ONUs) in the time slots when they were not transmitting.

Research continued up to the millennium bubble burst, when research in many areas in telecommunications waned. However interest in LR-PON was resumed in the early part of the millennium, but with emphasis on the ability of LR-PONs to change network architectures. One of the triggering papers for this was [30.7]. This paper described the economic challenges that were facing operators and in particular the disparity between the costs of supporting bandwidth growth and the amount of revenue that could be obtained from the customer base. This is where the long-reach capability of SuperPONs came to the fore and the term long reach or LR-PON took over from the term SuperPON. The long reach of LR-PONs could be used to close down the vast majority of LEs/COs and carry traffic deep into the network to larger centralized core nodes (CNs) for traffic processing, routing, and switching. In addition, if the number of CNs could be reduced sufficiently then the core network could also be simplified with significant cost savings which could be used to help finance the optical distribution network to the end customers. Similarly, the long reach also enabled elimination of the metro network and its separate transmission systems, a further cost saving which helps transfer capital investment from the metro and core networks towards the access fiber network.

The economic analysis referred to above stimulated a fresh wave of research into LR-PON with greater emphasis on the reach and number of customers captured per PON. It also included the use of dense wavelength-division multiplexing DWDM over the LR-PON infrastructure. WDM techniques had been considered in PONs since the initial inception of splitter-based shared-access fiber networks. Indeed, the first splitter- and star-coupler-based networks were predicated on WDM rather than TDM (time-division multiplexed) [30.8] for the configuration of customers onto a fiber platform with programmable topologies, which could be considered the forerunner of the software-defined access network (SDAN).

In 2006, a study [30.9] proposed a hybrid DWDM-TDMA (time-division multiple access) LR-PON architecture that used the wavelength domain to increase the total number of customers connected to a head-end DWDM/TDMA terminal or optical line termination (OLT) device and a common long backhaul feeder fiber pair. In this approach, a fixed pair of DWDM wavelength channels is allocated to each physical sub-PON, one for downstream and one for upstream. Users share the downstream/upstream channel bandwidth by means of a time-domain TDM/TDMA protocol as in a conventional PON. This scheme shared the cost of the head-end, backhaul fiber and optical amplifiers over a greater number of customers than single wavelength LR-PON solutions. This hybrid DWDM-TDMA approach was further developed and studied in the EU PIEMAN project [30.10, 11].

When DWDM is introduced into the LR-PON solution, the transmitter at the customer ONU must have its laser transmitter wavelength accurately aligned to the allocated wavelength channel. Also to avoid deployment and sparing problems the ONU should be colorless, that is it should either tune to the required wavelength or use a predefined downstream wavelength as a seed reference wavelength for the upstream wavelength channel. PIEMAN investigated both approaches building two hybrid LR-PON demonstrators. The wavelength referencing experiment used an integrated reflective electroabsorption modulator (EAM) [30.12] plus semiconductor optical amplifier (SOA) for the upstream transmitter at the customer ONU, and achieved 135.1 km total reach, 10 Gb/s symmetrical bitrate, 64 wavelengths (32 up and 32 down) each serving a 256-way split, a total of 8192 customers [30.10]. The second experiment used a low-cost tuneable laser for the upstream transmitter. The reflective EAM-SOA transmitter approach suffers from Rayleigh back scatter from the downstream seed wavelength required for the upstream transmitters and so limits the optical split in the upstream direction. The tuneable transmitter approach does not suffer from this problem and greater ODN splits can be achieved. In this latter experiment the split was increased to 512 ways with 16 384 total customers [30.11].

The use of DWDM to increase the number of customers connected via the feeder fiber and OLT reduces the cost per customer of these elements, however when these costs have reduced sufficiently there is a dimin-

ishing return on further increases in the number of customers connected.

Another version of PON that uses DWDM is WDM-PON and this approach uses a wavelength multiplexer/de-multiplexer at a splitter node rather than power splitters (usually a single WDM splitting node for each PON installation although a dual-stage WDM split has also been proposed) [30.13]. This WDM-PON architecture has the advantage that it does not suffer from the $10\log(n)$ splitting loss of the conventional PON power splitter. WDM-PON can also be long reach by use of reach extenders in a similar way to TDMA-PONs. Split in these PONs is limited to the number of wavelength channels supported by the WDM devices used and also the bandwidth in each wavelength channel is dedicated to a single end user terminal or ONT (optical network terminal), see [30.13] as an example of WDM-PON.

Coherent optical technology has also been demonstrated for PON systems [30.14, 15]. Coherent technology has the advantage of increased power budget because of the greater sensitivity of coherent receivers and the extra signal processing that can be applied to coherent systems to mitigate some impairments. However those gains come at increased cost of the equipment, which is particularly important at the user end, and hence economics will dictate the adoption and take up of coherent PON solutions. Coherent technology also has the advantage of greater spectral efficiency including greater packing of optical carriers in a given optical window. This can be exploited in both WDM-PON- or hybrid DWDM-TDMA-type architectures.

DWDM-TDMA with flexible wavelength assignment is another approach for the use of WDM technologies. In this type of system some wavelengths are used for increasing the capacity to customers connected to LR-PON systems and others to provide dedicated or bespoke Pt-Pt services to targeted customers such as large businesses. Using wavelengths for capacity and service enhancement enables the LR-PON architecture to serve all customer types from residential through to large business and also mobile base-stations, but requires a more flexible and dynamic wavelength assignment scheme than those systems that fix the wavelength with WDM devices in the fiber network.

30.2 Architectural Design for the End-to-End Network Using LR-PON

Following the financial crisis in the IT and telecommunications industries at the beginning of the new millennium and then the financial crisis of 2008 which affected economies all over the world, the future economic viability and the ever increasing energy demands became (and still are) major challenges facing telecommunications networks. Continuing to build and grow networks with the three tier access, metro, and hierarchical core networks was not going to be viable if projected growths in bandwidth demands were to be realized. What was going to be required was a fresh look at the end-to-end architecture of the future network and LR-PON is an enabling technology that can facilitate both access-metro convergence and simplification of the core network. It does this by reaching deep into the network bypassing the majority of LE/CO buildings and consolidating traffic onto a much reduced number of CNs. If the number of remaining CNs can be reduced sufficiently then the core network can be configured as a single-layer flat optical core network with a full mesh of optical light paths interconnecting these remaining nodes that contain traffic processing equipment. The flat core architecture was shown to be the lowest cost once user bandwidths passed a threshold and would then remain so as bandwidth continued to grow [30.16]. The advantage is dependent on the number of nodes in the core network with a general tendency for the benefits to increase as the number of CNs is reduced. Reducing the number of CNs requires longer reach LR-PONs. An example of LR-PON used for end-to-end network restructuring can be found in [30.17, 18]. Such an architecture is outlined in Fig. 30.2. It uses 100 to 125 km LR-PONs to enable massive node consolidation of the order of 50 : 1 reduction in LEs/COs for countries with European population densities, so that for example the UK could be served with approximately 70–100 traffic processing nodes in the core of the network. The metro network is subsumed into the LR-PON infrastructure and no longer requires separate transmission equipment. The LR-PONs terminate onto consolidated metro/core nodes (MC-nodes), which are the only traffic processing nodes in the network. One important and distinguishing feature of the design of these nodes was the introduction of a fully transparent optical switching layer that interconnected the access and core sides of the node and also all the equipment required within the node. This optical switching layer introduced a flexibility for interconnection and reconfigurability that could enable full network function virtualization (NFV), even at the physical fiber layer, and when coupled with software-defined networking (SDN) and NFV in the processing layers (routers, switches etc.) enabled the realization of a full SDN/NFV architecture.

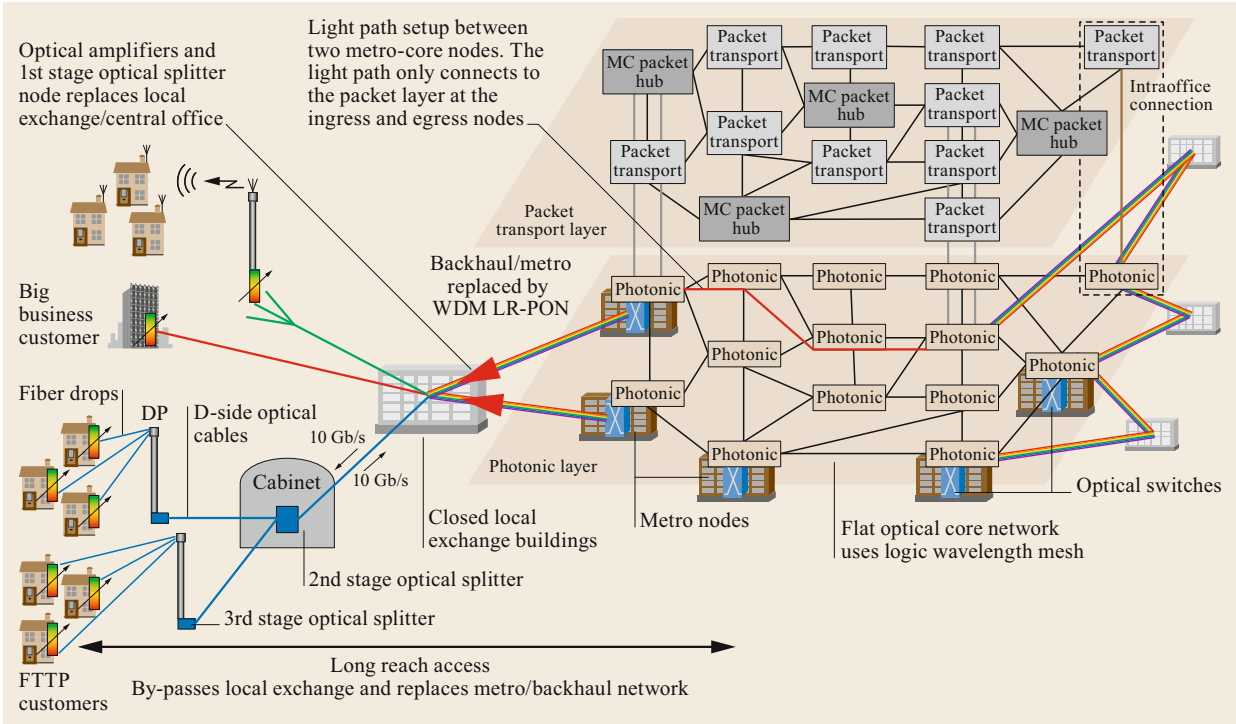


Fig. 30.2 An end-to-end architecture proposed in [30.16]. It consists of LR-PONs with DWDM/TDMA which bypasses LEs/COS and eliminates the metro-access network transmission systems. The LR-PONs are dual parented onto two separate MC-nodes which are interconnected via a flat optical core network (© IEEE 2016. Reprinted, with permission, from [30.16])

Reducing the number of CNs increases the reach requirement on the LR-PON network which not only stretches the technical challenges of the LR-PON design but also increases the round-trip time (RTT) of the access network from customer to the first traffic processing node (the MC-node).

As described later in Sect. 30.2.4, important advantages can be obtained by converging the mobile services on a common optical access infrastructure such as an LR-PON employing DWDM/TDMA. However, the RTT increase in LR-PON introduces problems for certain wireless systems such as wireless front hauling, which has the potential for reducing antenna site costs for mobile networks. Simply reducing the maximum reach of the LR-PON would jeopardize the economic benefits of the flat optical core network pushing up the costs for the whole network and all users. The compromise solution proposed in [30.18] was to place some processing equipment in selected ANs to terminate the low-latency mobile networks. This increases cost, but those costs would be supported by those mobile systems and services requiring the low latency and not increase the costs for all the other systems and services not requiring it and avoids any cross subsidization of mobile service from fixed network services.

30.2.1 Backhaul Cable Topology Options

The architecture outlined above focused on a flexible DWDM-TDMA LR-PON so that flexibility of bandwidth provisioning could be applied in both the wavelength and time domains. The general LR-PON design is based on the classic *lollipop* architecture which has a long backhaul section up to 90 km and a single AN which drops the ODN section with a reach of ≈10 km and 512- or 1024-way split as a standard design. The long backhaul cable routes will generally pass other LE/CO sites also containing ANs and cables can be arranged in rings, chains, or tree and branch topologies. Ring topologies are often proposed as these follow on from the backhaul/metro network SDH (synchronous digital hierarchy) rings that were used in the early years of fiber deployment in the metro and core networks. Ring systems are usually closed via paths through the core network but this is not really necessary for LR-PON systems and open rings or chain/bus architectures for the cable routes can also be considered. A cable chain architecture connecting LR-PON systems in a number of LE/CO sites is shown in Fig. 30.3 where the LR-PON systems are dual parented onto two separated MC-nodes for protection/resilience purposes.

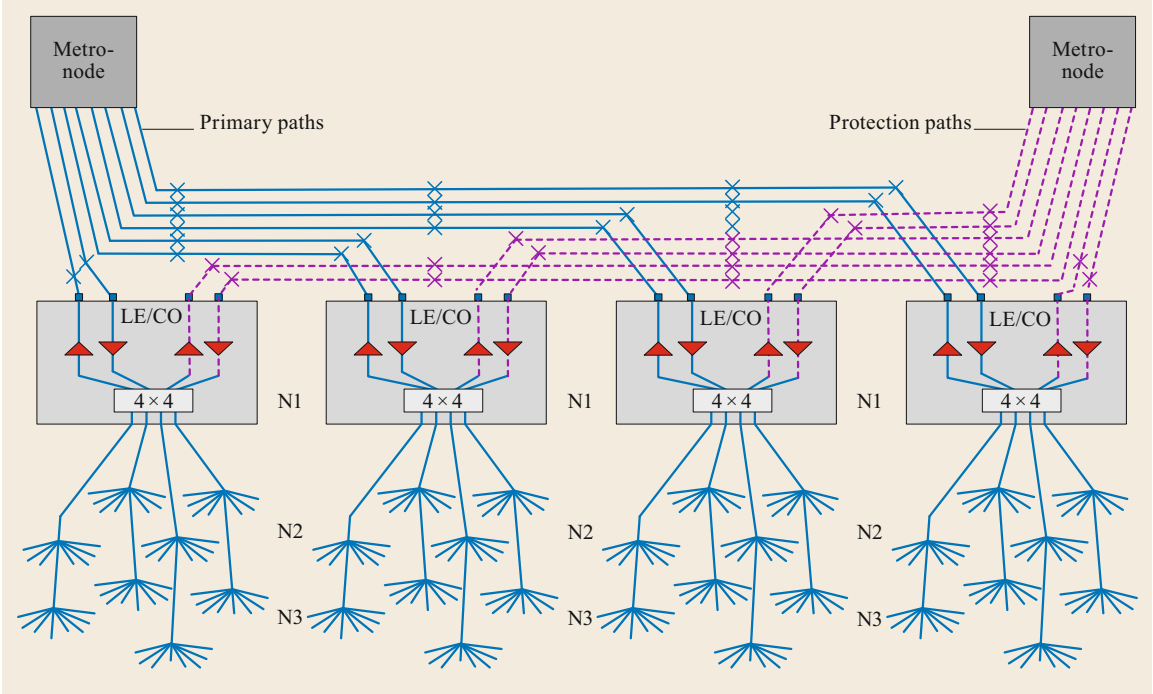


Fig. 30.3 An open ring or chain backhaul fiber cable topology. This has the advantage of reuse of cable structures used by earlier ring-based technologies and also aggregates the fibers into larger cable sizes which reduces the cost per fiber via greater sharing of common costs such as sheathing, duct, and trenching

The project described in [30.18] also proposed two designs for the LR-PON architecture: one general solution which used the *lollipop* design (the design is the same regardless of backhaul cable topology) with a nominal reach of 100 km which would be used for the majority of the network where customer sites were in regions of moderate to high density; and a second solution aimed at sparse rural areas where longer reach may be required to enable efficient node consolidation. The sparsely populated region solution comprises a chain of ANs, where part of the amplifying function in the AN is as a line amplifier, enabling greater reach to be achieved. This effectively distributes the total ODN split over a number of ANs in separated LE/CO sites. The structure of the sparse rural solution is shown in Fig. 30.4.

The operation is more complex than the conventional LR-PON configurations and will therefore be described in more detail. There is a mix of single fiber and two fiber working in the backhaul fiber sections with single fiber working in the ODN on the customer side of the ANs (note single or dual fiber working can also be applied to the *lollipop* architecture). To aid clarity, Fig. 30.4 also shows the up and downstream wavelength bands in various parts of the network (the downstream band is arbitrarily shown as green and the upstream band as red). For simplicity the option of splitting the up and down stream bands into two separate bands is shown, but the principles would also apply to interleaved up and downstream wavelengths.

In normal operation, downstream and upstream traffic, to and from the active MC-node, is transmitted in separate fiber paths. However, one of the main design features of the amplified chain is that the downstream and upstream paths can be reversed for protection in case of equipment or fiber path failure, when the protection CN becomes the active node. This implies that the downstream and upstream paths in the chain should be designed with the same criteria and also that the AN should be designed symmetrically in terms of the two paths. For this reason, as shown in Fig. 30.4, the upstream traffic can also be transmitted towards the protection MC-node in the fiber path that is transmitting the active downstream traffic. The upstream traffic transmitted towards the protection MC-node is idle traffic, which could be blocked using for example an optical switch, but the simpler design presented in Fig. 30.4 has the advantage that the protection reverse of the traffic can be triggered without any active reconfiguration of the AN, apart from turning off the primary downstream amplifier into the ODN and turning on the protection path amplifier. Wavelength collisions for LR-

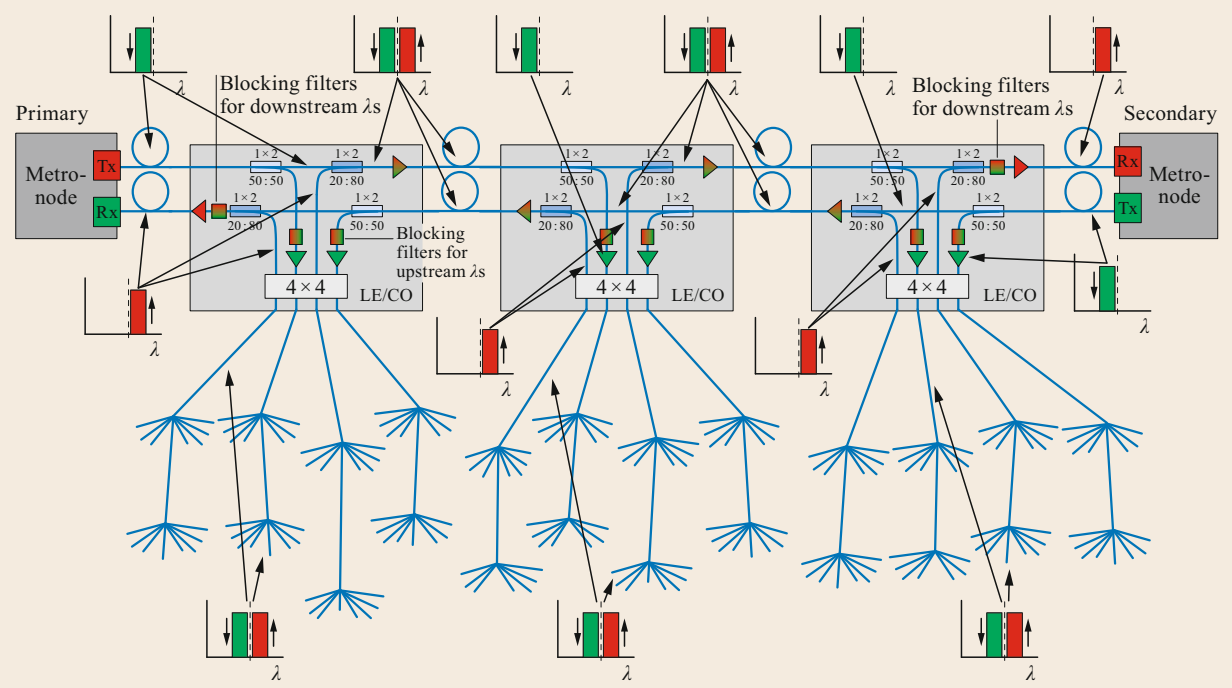


Fig. 30.4 The DISCUS project design for LR-PON that would be suitable for sparse rural areas. It places amplifiers in the backhaul fiber path at each AN which enables greater distances to be achieved

PON wavelengths are avoided by the LR-PON protocol since in this configuration, all the ANs in the chain form a single LR-PON controlled by a single instance of the LR-PON protocol, so only 1 ONU is transmitting an upstream burst at any time on any one wavelength channel.

30.2.2 Options for the Wavelength Plan

The use of DWDM over LR-PON raises the question of how best to optimize the wavelength plan for the fiber network. If we take the C-band as an example, the obvious choice would be to split the band into two halves; one for upstream and one for downstream wavelengths. The individual wavelength channels are separated or combined at the MC-node using AWG (arrayed waveguide grating) devices. An alternative approach would be to interleave the up and downstream wavelengths. This approach has a number of advantages. It minimizes crosstalk between channels in the copropagating directions, which can be important if wavelength channels are flexibly assigned and any mix of wavelength usage and modulation scheme can end up adjacent to each other. For example, 100 Gb/s coherent modulation signals using dual-polarization quadrature phase-shift keying (DP-QPSK) for Pt-Pt services could be adjacent to 10 Gb/s on-off-keying (OOK) LR-PON protocol channels and over the long-reach backhaul section could produce nonlinear crosstalk impairments. A second advantage is that the design of the tuneable laser at the ONU can be simplified by referencing to the adjacent, precisely defined, downstream wavelengths and interpolating between them to tune the upstream transmitter. This simplifies and potentially reduces the cost of any wavelength referencing components and systems within the ONU transmitter (Sect. 30.3.2).

In terms of wavelength bands for operation, the C-band has the distinct advantage of having mature optical components for DWDM operation, low fiber losses, and the availability of EDFAs as a high-performance amplification platform. Operation of DWDM LR-PON in other bands (for example the O-band) is also possible since SOAs are an alternative amplification solution. However, this can be seen as a longer term development since the full suite of required DWDM components and SOAs, particularly in fully integrated structures, are not yet mature and commercially available for these other wavelength bands.

30.2.3 Resilience in LR-PON

Generally operators do not provide protection in the access network for mass market services as this would be an expensive addition to the costs of this part of the

network. Those customers that need protection to their premises will pay for the necessary equipment and work required to offer full protection into their premises. This is also true for PON solutions which generally will not have protection within the ODN without specific requests and payment from customers, and a range of options have been developed for PON protection for those customers that require it; for an introduction of these techniques see Chap. 27.

However, the long reach of the LR-PON architecture increases the probability of a fiber cut to an extent that it becomes difficult to maintain high availability. The main increase in length is in the backhaul fiber network, which is the part of the network that historically would have protection mechanisms in place against fiber cuts and failures in the transmission equipment. This is therefore the area of the LR-PON that requires the implementation of protection mechanisms as a standard feature [30.19]. It can be seen in Figs. 30.2–30.4 that the backhaul fiber paths, the amplifiers in the ANs, and the OLT equipment are all protected by dual parenting the AN onto two MC-nodes so that in the event of a failure in the backhaul network the system would switch over to the standby node. Experiments have shown that the switch over time could meet the typical 50 ms nominal target for core protection mechanisms [30.17]. This dual parenting approach also increases protection against core network failures while minimizing equipment required for protection within the core network. Moving failed traffic paths onto an alternate MC-node provides access to a greater number of alternate paths through the core network. The wide separation of the protecting MC-nodes from the primary MC-node is also sufficient to act as a disaster recovery system in the event of the loss of a complete MC-node.

Further equipment cost savings can also be obtained by using an optical switching layer in the MC-node to provide $N:1$ sharing of LR-PON terminating equipment. By having a pool of protection equipment at the MC-node sufficient to recover from a worst-case failure (complete loss of an MC-node) the optical switching layer can be used to switch this equipment to the appropriate ports for the standby LR-PON protection paths. This saving in protection equipment can more than compensate for the cost of the optical switching layer.

30.2.4 Mobile Networks and LR-PONs

Mobile access networks have developed independently from the fixed networks used to deliver services for residential customers. Typically, radio access networks (RANs) for 2G, 3G, and 4G are connected via coarse wavelength-division multiplexed (CWDM) optics and in some cases with wireless connection to the aggregation switches which may also be used by wired network elements. However, in the access space, dedicated fiber links are often arranged so that they do not provide either the use of common technologies nor of a simple sharing of the fiber infrastructure. In the core segment the situation is a little better where typically the transport mechanisms are shared for the wired and wireless networks, but the control, operation, and management functions are still separate.

In legacy mobile networks, base stations are typically sparsely deployed (in the range from several hundred meters to several kilometers) and have worked cost-effectively without the need for tight integration with the fixed access network. However the next, 5th generation of mobile networks, will see a very significant increase in the density of access points, if target capacity increases of 100 times are to be met. At the same time there could be even stronger growth in fixed network capacity due to both fixed network traffic growth and increasing offload of traffic from mobile networks to the fixed network. This, coupled with the increase in number of small cells will cause issues of economic viability, which should be addressed by increasing the sharing of resources between the fixed and mobile network domains.

Cloud RAN or centralized RAN (C-RAN) was proposed as a solution to decrease the cost of remote radio units (RRUs), by simplifying them and moving all data processing into a centralized location where resources can be reused across multiple cells while also benefiting from statistical multiplexing gains of aggregated traffic across several cells. However, such C-RAN connectivity requires much higher capacity [30.20] and very low latency and jitter tolerance in the fiber transmission system connecting the remote radio unit (RRU) to the baseband unit (BBU).

An additional class of challenging applications in the realm of 5G networks are those requiring ultrareliable and low-latency communications (uRLLC), which might require end-to-end latency of the order of one millisecond or below. Such low latency implies that these *end-to-end* services need to be locally terminated [30.21], providing only a local reach (e.g., in terms of data sharing) and will need to be replicated at every edge node where those specific services are required. While PONs are a cost-effective mechanism for sharing access network infrastructure, such low latency requirements cannot be satisfied unless the mobile and PON scheduling mechanism are integrated [30.22] and the required reach of the services using the cloud RAN transmission systems is limited to a few tens of kilometers.

The LR-PON approach exacerbates the issue due to its longer transmission distance, which of course increases the minimum latency or RTT, making it difficult

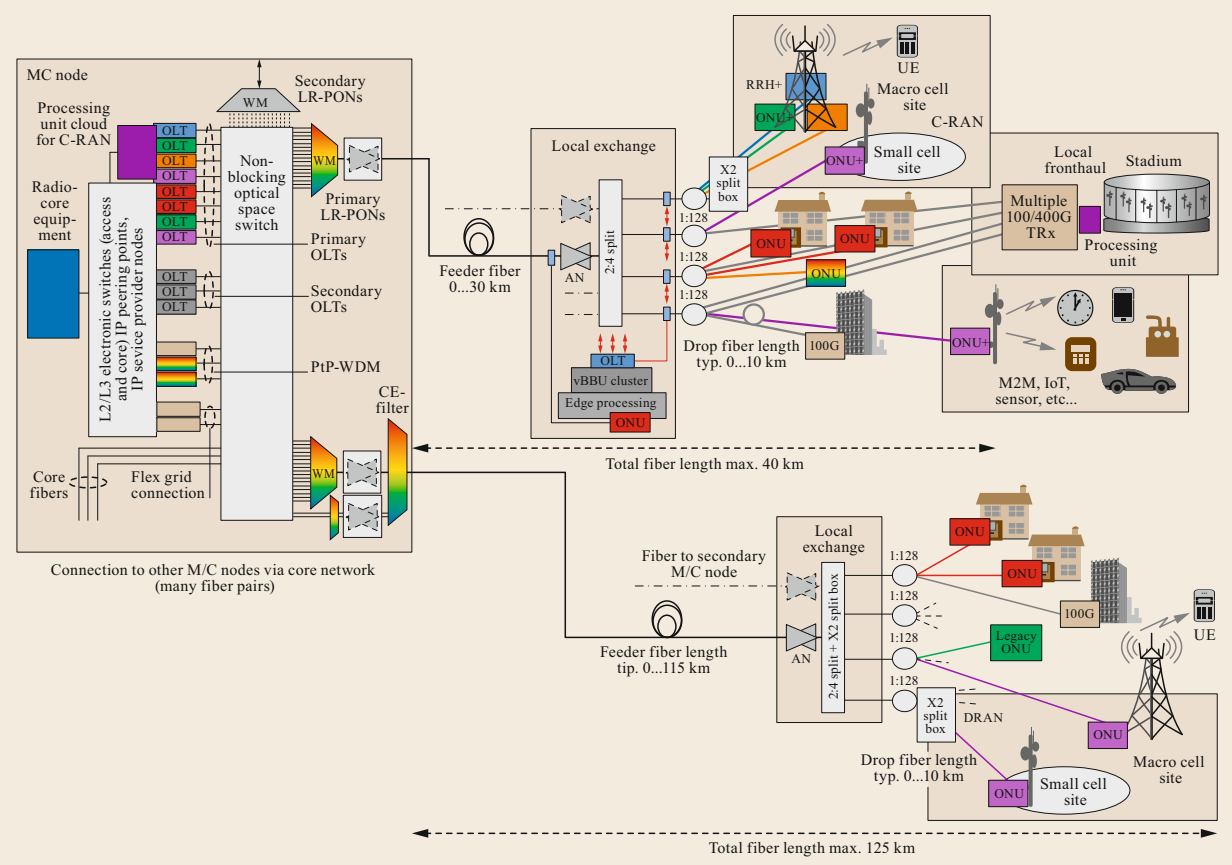


Fig. 30.5 LR-PON options for supporting wireless front haul (30 km reach limit) and wireless over longer reach LR-PON (up to 125 km)

to meet the requirements of some of the 5G latency targets. Thus, for these ultralow-latency services, the LR-PON ability to cost-effectively bypass the metro transmission traffic cannot be exploited.

Possible architectural solutions for LR-PON designed to satisfy such constraints are shown in Fig. 30.5. Note that, for simplicity, the ODN splitter below the AN is shown at a single location but the discussion also applies to a distributed split as shown in Figs. 30.3 and 30.4. The LR-PON designs in Fig. 30.5 also show single fiber working in both the ODN and backhaul network but again the designs will apply to two-fiber working solutions.

The left part of Fig. 30.5 shows the MC-node design. It comprises the primary and secondary (protection) OLTs for the flexible DWDM-TDMA PON and the Pt-Pt-WDM transceivers (which could operate using direct or coherent detection). The nonblocking optical space switch can distribute any input port to any output port independently of the number of wavelength channels on that port. Layer 2 (L2) and Layer 3 (L3) electrical switches and routers aggregate and distribute the data. The core transmission part has been omitted in this figure for sake of simplicity.

The right-hand part of Fig. 30.5 shows the ODN part of the LR-PON, together with different options for service termination. At the top we show, for example, a local exchange at a distance of 10 km from the access points, with the ability to terminate some of the PON wavelengths. This could be used for example to terminate a fronthaul connection to a local processing facility (e.g., edge cloud). The AN site (the old LE/CO site) comprises 1 EDFA for each direction (Fig. 30.5).

One well-known issue of fronthauling is that it requires very high transmission rates and each sector of a macro cell/small cell can easily require a dedicated 10 Gbit/s DWDM-TDMA PON wavelength channel and possibly a high-speed (100 Gbit/s or beyond) transceiver. The number of connectable sites is therefore limited to the number of downstream and upstream wavelength channels that can be made available for fronthauling applications. Thus, most of the wavelength

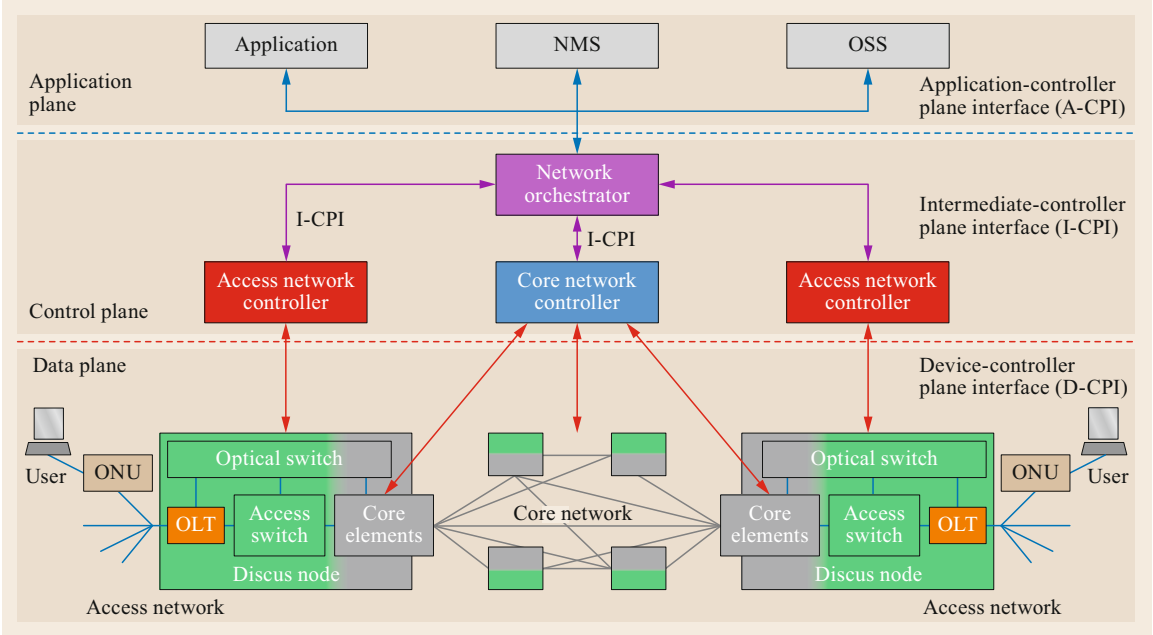


Fig. 30.6 Network controller function partitioning showing the access and core network controllers with the overarching network orchestrator

channels in centralized-RAN wireless integration scenarios will be dedicated to the wireless service whereas the fixed network wired services will only require a limited number of wavelength channels. Such a deployment scenario may be applicable for urban areas where a large number of customers are located close to the MC-nodes. It should be noted that adding termination points in the ODN can increase substantially the overall network cost, which raises the issue of whether this cost should be borne by the mobile services requiring such modifications, rather than be cross-subsidized by other network services that do not require low latency. However, these problems are regulatory rather than technical and it is important for a solution to be proposed by the relevant bodies.

The lower right-hand part of Fig. 30.5 is an option for an LR-PON where placing processing nodes within 10 km of the access points is not cost-effective, for example in a suburban or rural scenario. In these cases the backhaul distance will exceed the distance limits imposed by fronthaul and it could even exceed 100 km. In this scenario, distributed radio access network (D-RAN) mobile stations can be connected to shared PON channels and transported to the MC-node.

One final note we would like to make is that in the discussion so far we have considered traditional CPRI (Common Public Radio Interface) fronthaul, which has the disadvantage of requiring high sustained transport bit rates for relatively little traffic bandwidth. To overcome this disadvantage, modifications to the conventional CPRI-based C-RAN architecture have been proposed [30.20, 23] and have been considered by the NGMN (Next Generation Mobile Networks) Alliance [30.24] and other standardization bodies. The most attractive solution for the mentioned challenges is the use of midhauling with dual-site processing. At the last point in downlink direction, where the user data statistics still take effect, the interface capacity dynamically ranges from zero (for no traffic) up to about 20% of the CPRI rate in the case of fully loaded radio channels [30.20]. The LR-PON solution is flexible enough to support both distributed-RAN and centralized-RAN solutions coexisting together in a single WDM-TDMA LR-PON.

30.2.5 Control Plane Design for Converged Access-to-Core Node Architectures

One of the requirements that 5G services put on transport networks is the delivery of assured quality of service (QoS), which allows multiple heterogeneous services to be carried and processed, respectively, across a common transport and cloud infrastructure. The network control plane needs to assure that end-to-end resources are appropriately provisioned to meet latency and capacity requirements.

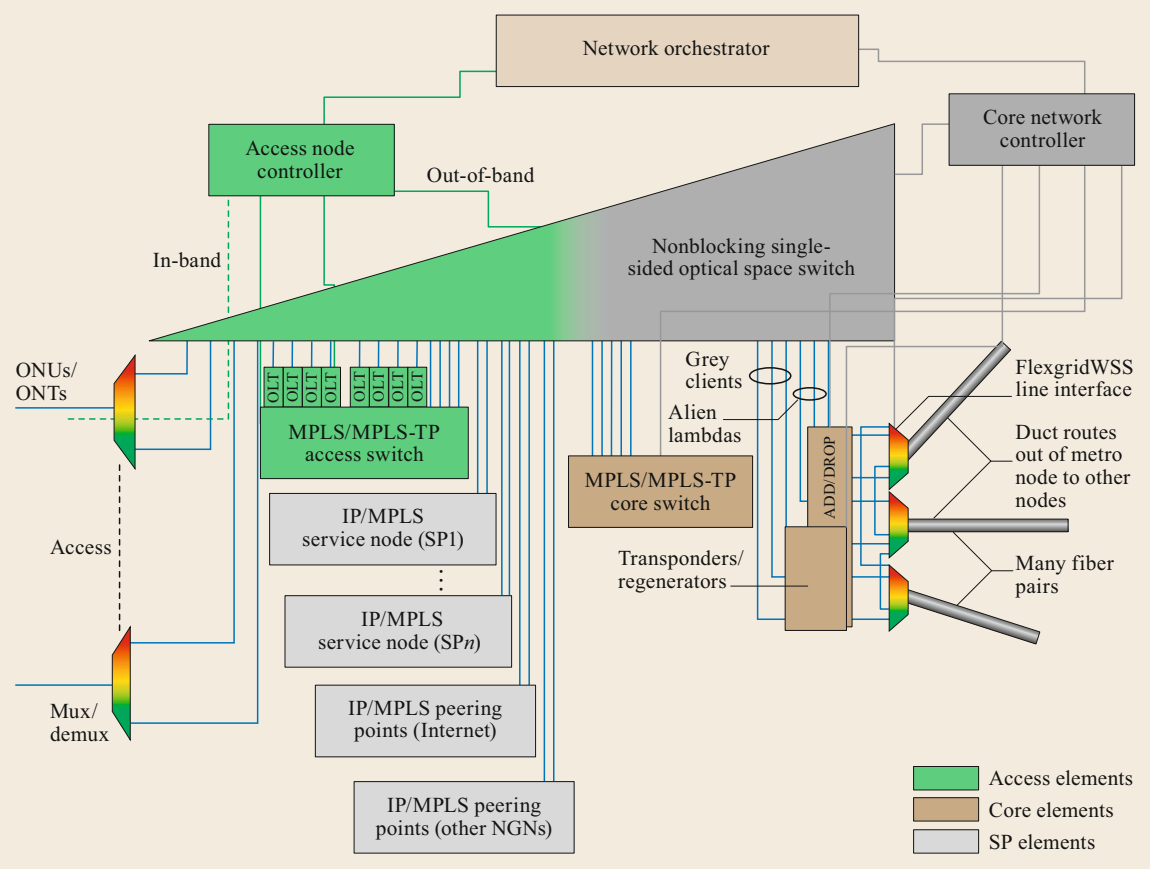


Fig. 30.7 Partitioning of the different access and core elements, under the control of the access and core network controllers, respectively, and the orchestrator controlling both controllers in order to set up an end-to-end service request

In an LR-PON-based architecture with a flat optical core network, the selected approach is that of a hierarchical control plane structure. The strong consolidation ability of an LR-PON affects the design of its control plane architecture. The highly simplified network architecture, with only two OEO (optical to electrical to optical) interface points (e.g., at the source and destination metro/core nodes), lends itself to a control plane structure with one core network controller, which can also operate as orchestrator, and a number of access network controllers. It should also be noted that if the size of the network is such that it cannot be covered by one flat optical core network, an additional cross-core network orchestration stage should be employed, to coordinate transmission across the multiple flat optical core networks.

The structure in Fig. 30.6 shows a breakdown between access and core controllers [30.25]. The division is functional, as the access controller takes care of the access portion of the MC-node, while the core controller is in charge of the core transmission part. For backwards compatibility with existing systems, the latter could further delegate network provisioning to existing systems, such as generalized multiprotocol label switching (GMPLS).

In the architecture shown in Fig. 30.6, the network orchestrator provides the north-bound API (or application-control plane interface, in open network foundation (ONF) terminology) to the application plane, interacting directly with services/applications or to a network management system and operation support system. After receiving a connection setup request, the orchestrator has the task of calculating the end-to-end path, by selecting the MC-node pairs needed for the connectivity and a suitable transmission path across the core. This can be at the level of virtual path creation (e.g., a multiprotocol label-switching MPLS path across an existing optical link) or trigger the creation of additional capacity at the physical layer.

The next step is for the access network controllers on the two MC-nodes to carry out the necessary connections to link to the end user on the access side and

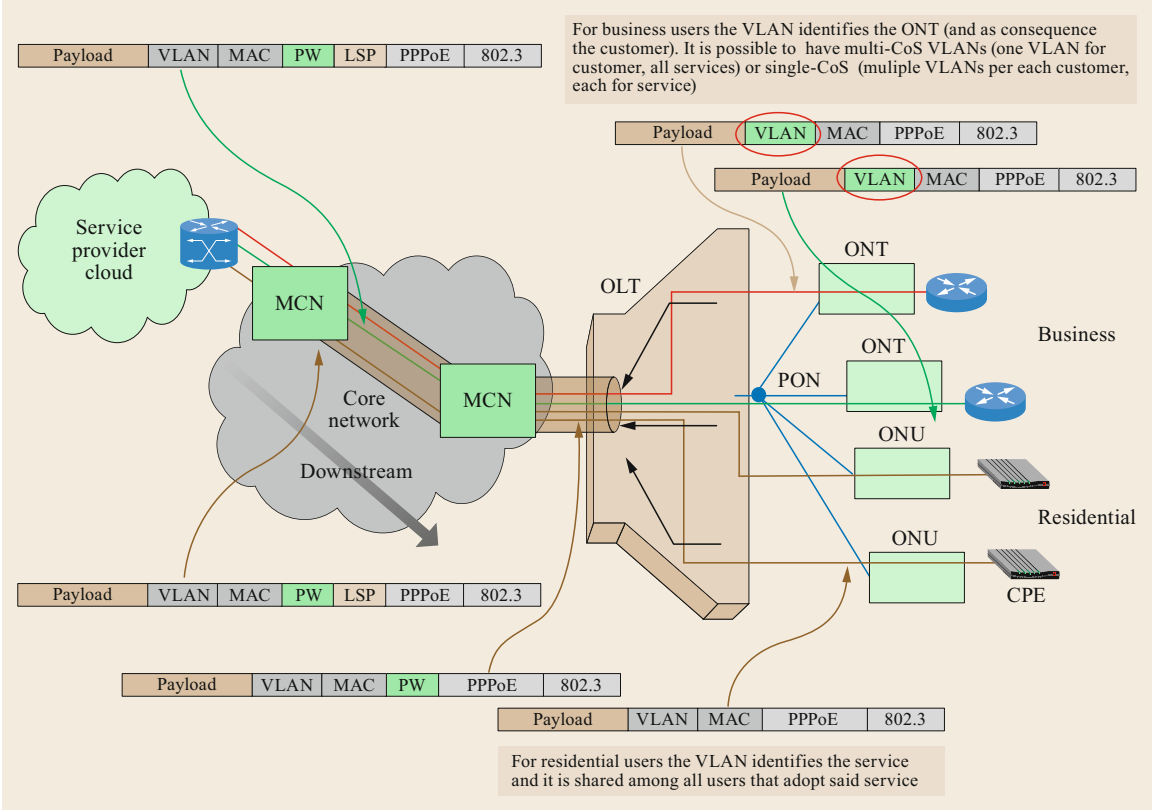


Fig. 30.8 A service implementation example using LSP and PW objects

to the appropriate core-facing connection on the core side.

It should be noted that Fig. 30.6 shows a general LR-PON control plane architecture, and different practical implementations have been carried out, such as in [30.26], which adopted the Telefonica Netphony ABNO (Application-Based Network Operations) implementation to merge orchestrator and core network controller.

Figure 30.7 abstracts the main network components of an LR-PON metro/core node and, while it does not provide full details, it classifies them into access-facing and core-facing elements. On the access side (green color) we find the OLTs and access switch. The MPLS Service Provider equipment is also part of the access side, although they need not be physically colocated with the metro/core node. The optical switch, which is the element that provides flexibility to the architecture, enabling flexible any-to-any connectivity of the components, is virtualized into an access-facing part and a core-facing part. Finally, the network core elements include an aggregator core switch, with any reconfigurable optical add-drop multiplexer (ROADM) and transponder required for backbone communication.

Any request for connectivity, whether or not it is associated with the creation of additional wavelength channels, operates on two sets of stacked MPLS labels. The inner label, the PseudoWire (PW), identifies an end-to-end path, and is assigned to a specific service type (e.g., video on demand, Internet, VoIP, bandwidth on demand). The PW also univocally identifies the service provider (SP) and the terminating OLT. Each service provider will aggregate multiple PWs into a label-switched path (LSP), identified by an outer label, which is transported through the core.

In principle, an SP can have several LSPs for each pair of MC-nodes, for example having different quality of service requirements, but at least one per pair is required.

Label-switching operations across MC-nodes are shown in Fig. 30.8. Starting from the right-hand side, the SP will push a PW label, in agreement with the access control plane, to a new service flow. All similar flows (e.g., with similar QoS requirement) directed towards the same MC-node can be associated with the same LSP, as the access aggregation switch installs the LSP label on the packet headers. Packets are then forwarded to the MC-node where the LSP label is ex-

amined to determine whether the aggregate flow is to be processed at this node or forwarded through the core towards a different MC-node. Once it arrives at the destination MC-node, the LSP label is removed by the access switch, which uses the PW label to determine QoS behavior of individual PW flows and reach the correct OLT. The OLT then strips off the PW label forwarding the packet to the ONU.

30.3 DBA and DWA Protocol Implications for LR-PON

The ability to assign capacity flexibly to end user terminals has been a recognized advantage of the point-to-multipoint topology of PON systems from the very initial concepts of splitter-based fiber networks and this included flexible wavelength assignment as well as finer granularity bandwidth assignment in the electrical TDM domain. This section outlines some of the additional issues that arise for dynamic bandwidth assignment (DBA) and dynamic wavelength assignment (DWA) for wavelength- and time-division multiplexed DWDM-TDMA LR-PONs.

30.3.1 Dynamic Bandwidth Assignment (DBA)

LR-PONs present greater challenges for DBA algorithms than Gigabit-capable PON (GPON) and Ethernet PON (EPON). Their long physical reach of up to 100 km will increase users' data traffic scheduling delays due to DBA round-trip delays and, in order to provide a high degree of node consolidation, with larger access network areas and greater customer numbers, larger total splits will be required; 1024-way split has been demonstrated [30.29]. 10-Gigabit-capable PON (XGPON) [30.30] already allows for 1023 ONUs, and up to 16 383 allocation identifiers (Alloc-IDs) for separate upstream traffic flows. The increased delays and large numbers of flows have implications for the choice of DBA algorithm. For a comprehensive survey of DBA schemes for LR-PONs the reader can refer to [30.31]. The need for DBA performance to be studied for large split has been recognized [30.31], in terms of delay, utilization/efficiency/throughput, and service level agreement. Some of the key issues are discussed in more detail below.

Delays in ONU to OLT communication and DBA computation time can produce inconsistencies between successive ONU traffic queue reports, i.e., Alloc-ID queue length status reports (SRs), as illustrated in Fig. 30.9. This is because traffic already granted by previous DBA processes, for transmission during the DBA execution time of the current (i-th) DBA process, DBA process$_i$, might still be present in its queue when the current DBA process' SR$_i$ is transmitted, causing over-

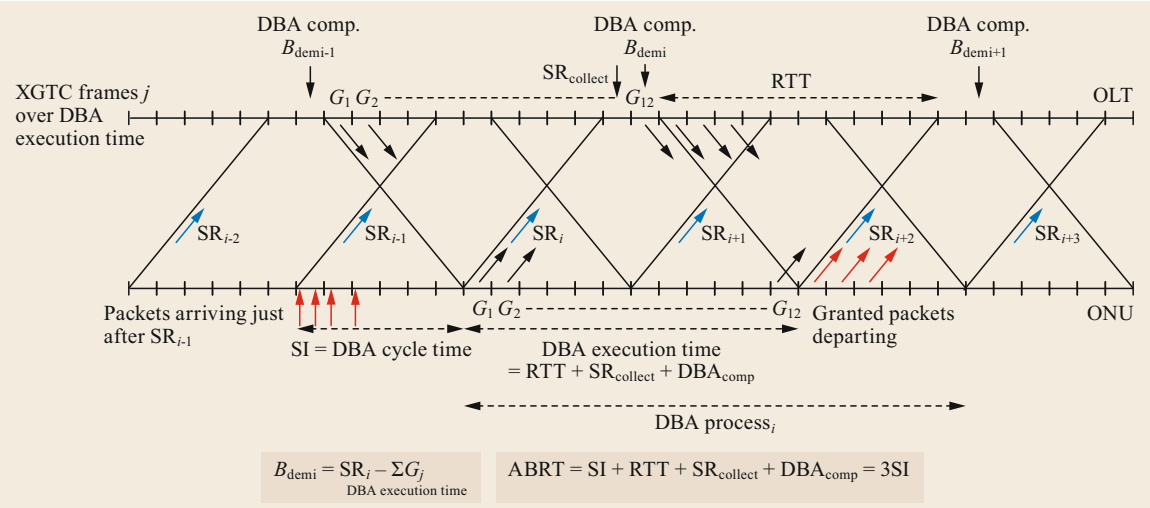


Fig. 30.9 Prevention of overgranting by subtracting previously granted traffic from a queue's status reports as in [30.27], for overlapping DBA processes having an example SI = 6 XGTC frames, RTT = 10 frames and DBA execution time = 12 frames. The assured bandwidth restoration time (ABRT) applies to the GIANT DBA algorithm [30.28], assuming it takes just one XGTC frame to compute an Alloc-ID's bandwidth demand

reporting. This can lead to overgranting of bandwidth due to incorrect bandwidth demand estimates B_{demi}. To prevent this waste of bandwidth in next-generation PON (NG-PON), the minimum guaranteed transmission container content (MGTC) scheme [30.27] subtracts, from the current SR_i, any outgoing traffic already granted by previous DBA processes. This is illustrated in Fig. 30.9 for DBA cycle time equal to half the DBA execution time, but applies whatever the ratio of the two. The DBA cycle time is the interval between the starts of successive DBA processes. Simulations reported in [30.27] have shown that if this algorithm is applied to a PON with 32 ONUs and just 20 km reach (e.g., a typical GPON system) then there are significant reductions in average scheduling delay for high load and short DBA cycle time situations, compared to schemes that only use the status report with no adjustment for traffic sent in previous grants. This simple algorithm can be reformulated to be equivalent to a proposed *newly arrived frames plus* (NA+) scheme [30.32], which is a running account that calculates the current bandwidth demand from the traffic arriving between previous and current SRs, and corrects this with a backlog term from a previous bandwidth demand and grant.

Correct bandwidth demand values are crucially important in LR-PONs that employ multiple overlapping DBA processes, as in Fig. 30.9. Overlapping the DBA processes reduces the scheduling delays, especially when using DBA cycle times shorter than the DBA execution time. However, increasing the number of DBA cycles increases protocol overheads and there is therefore a compromise in terms of reduced scheduling delays and PON protocol efficiency. Overlapping of DBA processes is implicitly achievable in long-reach GPON (LR-GPON) [30.32] and XGPON. But in long-reach EPON (LR-EPON) it required *multithreading*, a process equivalent to overlapping of DBA processes, to be introduced. Simulations [30.33] of small numbers of ONUs (16) and 100 km reach show significant reductions in average packet delay at high load, and significant increases in throughput (0.76–0.96). Simulations [30.32] for LR-PONs with 32 ONUs and reach up to 100 km show that multithreading is far more important for LR-EPON than for LR-GPON. Reduction in average delay at 0.8 load for LR-GPON is $\approx$ 1 ms, but for LR-EPON it is far more significant at $\approx$ 500 ms, because of the longer polling cycles needed to maintain DBA efficiency.

DBA performance is conventionally determined by simulation of mean packet delays and jitter. But for LR-PON, assured bandwidth restoration time (ABRT) is an important and useful delay measure [30.30]. For LR-GPON it has been shown that the maximum balanced load, beyond which queues begin to grow, and average packet delay and packet loss ratio rapidly increase, can be calculated analytically for different DBA algorithms [30.34], simply by using the required ABRT as the delay measure and the assumption that over the service interval (SI) (SI = DBA cycle time) the numbers of arriving and granted packets are equal. Note that the required ABRT is the worst-case time delay (not mean) between an Alloc-ID increasing its bandwidth demand and receiving its full provisioned assured (plus fixed) bandwidth, when previously not receiving it due to insufficient bandwidth demand [30.30]. The target value is 2 ms, but the expectation is a few ms.

The DBA algorithms used in [30.34] were based on the ones developed by the GIANT project [30.28], and a bandwidth update (BU) algorithm similar to the memoryless one for broadband PON (BPON) [30.35]. For a 100 km 10 Gb/s symmetric LR-GPON, with small numbers of ONUs (e.g., 32) and user Alloc-IDs (e.g., 512), overlapping of DBA processes allows the 2 ms ABRT target to be reached by both GIANT and BU, without sacrificing the achievable load ($\approx$ 0.94), by using SI < DBA execution time. 56 ONUs and 4096 Alloc-IDs achieve $\approx$ 0.78 load. But the performance of a 100 km-large split LR-GPON, having the maximum number of ONUs (1023) and Alloc-IDs (16 383), depends on the ONU burst overhead time (guard time, preamble, and delimiter) of XGPON [30.36]. If the short 256 bit objective value could be used, the 2 ms ABRT target could be achieved with usefully high load (0.78), by using SI < DBA execution time. Even higher load of 0.87 could be obtained with SI = DBA execution time by accepting just 2.5 ms ABRT. But with large numbers of ONUs (1023) and Alloc-IDs (16 383), the preamble requirements of burst-mode electronic dispersion compensation (BM-EDC) (Sect. 30.4.1) may preclude the 2 ms ABRT target from being achieved with useful load, as the requirements are expected to bring the ONU burst overhead time towards the worst-case value of 2048 bits. Nevertheless even with this worst-case value, the expected few ms ABRT can be achieved with useful loads by using SI $\geq$ DBA execution time. GIANT and BU algorithms both achieve > 0.6 load around ABRT = 3 ms, and BU outperforms GIANT from ABRT = 3.75 ms (with $\approx$ 0.77 load) and beyond [30.34].

30.3.2 Dynamic Wavelength Assignment

In optically amplified DWDM-TDMA LR-PONs, with up to 1024-way split and 100 km reach, and where multiple wavelengths are used for flexible assignment of capacity, DWA protocols have many challenges to overcome, during both normal operation and ONU

initialization and activation (start-up) [30.37]. These are predominantly due to the interplay between wavelength control/referencing (laser tuning precision), and physical layer impairments due to linear and interferometric crosstalk (Xtalk), the latter always dominating in the following analyses. Causes of impairment include Xtalk isolation of filters and wavelength demultiplexers, inadequate laser side-mode suppression ratio (SMSR), intensity modulation extinction ratio, and burst extinction ratio.

The most severe challenge is for serial number (SN) acquisition during ONU initialization, which is necessary for the OLT to identify the ONU that is starting up. This is particularly so when multiple ONUs attempt to start up simultaneously and their upstream ONU bursts collide, causing interferometric crosstalk with any working signal present and also between themselves. Conventionally, collisions can occur during quiet windows (no working signal) at full operational data rate (e.g., 10 Gb/s), despite the use of a random delay protocol [30.30]. At each attempt, ONUs randomly choose a delay between 0 and 48 μs in which to transmit their ONU bursts, to reduce the probability of collisions (even if all ONUs are equidistant from the OLT), and thereby reduce the number of rounds of ONU start-up attempts (quiet windows) required. Initial assessments of the potential performance of wavelength referencing solutions, Xtalk impairments and protocols were given in [30.38], but more recent theoretical work has shown that improved performance can be expected compared to that reported in [30.38]. Ideally, the laser wavelength should be close to the center of the intended wavelength channel. But if the laser tuning precision is initially poor (noncalibrated), it may be in the wrong channel before any feedback control from the OLT can correct it, disturbing operational ONU transmissions using that channel. To avoid disturbance in this case, use of a low data rate signal from a distributed Bragg reflector (DBR) laser transmitting a 1 MHz tone at low power has been proposed [30.39]. This is similar to the radio frequency (RF) pilot tone technique used to carry the auxiliary management and control channel (AMCC) in the PtP WDM PON part of NG-PON2 [30.40] Annex B, to provide information for assigning and allocating wavelengths. The low-frequency modulation of the interfering channel from a start-up ONU compared to the high bit rate signal of a working ONU enables electrical filtering to be used to detect the low-power signal in the presence of the higher power working ONU signal. Simultaneous wavelength monitoring and control of multiple lasers is also possible, as well as ranging and ONU activation, using different subcarriers or codes [30.39].

The Gaussian statistical theory in [30.41] can be used to model the impact of multiple interferers beating with a working signal. To ensure it applies only to ONUs attempting to start up simultaneously in the same wrong channel, and not to side modes from other working wavelength channels, a tuneable transmitter filter in the ONU, as shown in Fig. 30.10, improves the laser SMSR as discussed in Sect. 30.4.1. This is necessary if 16 or more wavelength channels and low power penalties are required in 100 km hybrid time and wavelength division multiplexing (TWDM) LR-PONs with 1024-way split. Using equation (7) of [30.41] for an optically preamplified receiver, total interferometric Xtalk ratio $C = kP_{max}/P_S$ in the range −16.7 to −21.5 dB (depending on the precise extinction ratio from 0 to 0.15, defined as optical power in the *0* level divided by optical power in the *1* level) would be sufficient to accommodate interferometric Xtalk between k interferers and the working signal P_S, for a 1 dB power penalty and Q value of 3.09. Obviously the more interferers there are, the lower their individual power P_{max} can be. Even when tolerating a modest number $k = 10$ of interferers into the 10 Gb/s receiver, suitable for providing an acceptable outage probability for SN acquisition [30.38], the very low power levels required in the low data rate receiver would result in impractically low speeds, which may be as slow as ≈ 1 b/s for a 1 dB power penalty. For a single successful start-up ONU, the dominant noise variance term is $(B'/B)P_{Smax}^2$, caused by the high-speed signal's electrical power spectral density detected within the low data rate receiver bandwidth B', where B is the high-speed signal bandwidth. Precise speeds would depend on extinction ratio (0.1 assumed here) and power level dynamic range P_{Smax}/P_S (5 dB is assumed here for power leveling increments and polarization variations). Even lower speeds would result from poorer extinction ratios, or if more than 10 ONUs simultaneously attempt to start up, and without power leveling. Power leveling is absolutely essential to limit the levels of both linear and interferometric Xtalk.

To avoid startup ONUs and working ONU data collisions synchronized quiet windows across all wavelength channels could be used. This would ensure that noncalibrated ONUs can be detected, whichever channel they tune to. This can also be done at full power and speed, which minimizes protocol delays. A full initialization protocol and multiple OLT synchronization details are described in [30.37] for a TWDM PON with 40 km reach. But synchronous quiet windows place restrictions on the transmission formats of the different channels, and on the service providers who operate them, and who may not wish to implement it. For example, some wavelengths may not operate PON protocols,

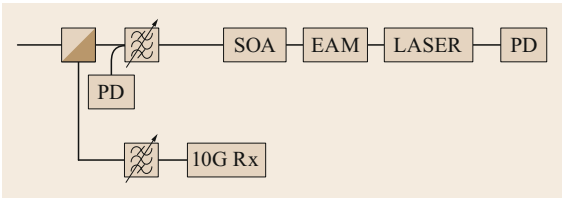


Fig. 30.10 Wavelength referencing scheme for coarse calibration in manufacture of tuneable laser located in customer premises

but support high-speed Pt-Pt services that do not have quiet windows.

To overcome the need for synchronized quiet windows, two potential alternative wavelength referencing solutions were proposed in [30.38]. Coarse calibration of the laser in manufacture at only one power level per wavelength should be sufficient to bring the laser to within the correct channel passband. Since accurate factory calibration is a lengthy process, which is today responsible for a substantial part of the tuneable laser cost, coarse calibration at only one power level could substantially reduce the factory calibration time and hence the cost. A possible schematic structure of the ONU employing coarse calibration is shown in Fig. 30.10. The tuneable filter at the transmitter improves the SMSR of nonoptimal side modes into neighboring channels, also relaxing the requirements on the laser calibration. A SOA after the laser adjusts the power level as required. This combination could provide tuning precision within $\pm 10-15$ GHz at all power levels to ensure start-up in the correct channel. Power leveling can either be performed by *blindly* incrementing the power, until successful detection of an ONU burst [30.42], or by calculation from a downstream received signal strength indication (RSSI) measurement at the ONU [30.40] Annex D.3. The latter needs start-up to be attempted at only one power level, which means that it is faster. However, the imprecision of the RSSI measurement could cause a greater power level dynamic range, worsening the interferometric Xtalk. Once the ONU is acquired by the OLT fine tuning can be provided by feedback control from the OLT.

The second potential solution is to use interleaved upstream and downstream wavelength channels, as proposed in [30.38], with a wavelength monitor (either a split contact SOA, or a two-section photodiode behind the laser), to measure the adjacent downstream channels in conjunction with the tuneable Tx filter, and interpolate the upstream wavelength between them. With this method, tighter tuning precision is expected, providing self-calibration without OLT feedback control. Both of these potential wavelength referencing solutions have yet to be demonstrated.

Both solutions allow ONUs to start up in the correct wavelength channel. This potentially allows start-up at full operational speed, within normal quiet windows. The dominant interferometric Xtalk variance term in the adjacent channels is in $(k^2-k)P_{\max}^2/X^2$, which are the $(k^2-k)/2$ beats between all pairs of the k colliding ONU laser bursts in the correct channel, X is the Xtalk isolation of a wavelength that is offset from the center of the correct channel by the worst-case wavelength error (tuning precision). The beats are detected at greatly reduced level in the adjacent channels due to the Xtalk isolation of the wavelength demultiplexer at the OLT. In the correct channel, the dominant interferometric Xtalk variance term, in $(B/\Delta f_{\text{opt}})(N-1)P_{\min}P_{\text{off}}$, is signal-ASE beat noise between the burst power $P_{\min}$ of the single successful start-up ONU and the off power P_{off} (when not enabled), as measured through the wavelength demultiplexer's optical bandwidth Δf_{opt}, of all $N-1$ off laser/SOAs that use that channel but are not bursting at that time. It remains to be verified experimentally precisely how many interferers k starting up simultaneously can beat together in the correct channel, without excessive interferometric Xtalk and error bursts in adjacent channels, while allowing a single successful ONU to start up at full operational speed, in the presence of all other Xtalk impairments [30.37].

To conclude, in dealing with the most severe interferometric Xtalk situation, when ONUs start up simultaneously and their bursts collide with a working signal or each other, the random delay protocol and power leveling become crucial protocols for reducing the Xtalk impact. Nevertheless, with poor laser tuning precision into the wrong channel, interferometric Xtalk can result in extremely low data rates for SN acquisition. Two proposed methods of ensuring start-up within the correct channel, i.e., coarse calibration in manufacture at just one power level, and self-calibration using interleaved US/DS wavelength channels, both offer the potential for starting up at full operational speed within quiet windows. A tuneable transmitter filter at the ONU to improve the laser SMSR is essential for all tuning precisions.

30.4 Physical Layer Design

Many of the components required to realize the physical layer of flexible LR-PONs with dynamic DWDM wavelength assignment are already commercially available, or can be readily adapted from existing products.

30.4.1 Key Enabling Technologies and Novel Components

However, a number of key new enabling technologies are needed, which are mainly associated with the upstream channel and the requirement to support wavelength-tuneable, extended-reach burst-mode transmission from a low cost ONU. Considering dense use of the wavelength domain with, for example, 40 channels divided equally between the upstream and downstream channels, the ONU transmitter would need to tune over 15 nm with a 50 GHz grid or 30 nm with a 100 GHz grid, which corresponds to a substantial part of the entire C-band. The stringent cost target for the ONU means that the transmitter must also have good prospects for low-cost volume production. The next subsystems in upstream path that must be specifically selected for this particular application are the optical amplifiers in the AN at the old LE/CO sites and the preamplifier in the MC-node. In contrast to other optical communication system applications, the optical amplifiers employed in the upstream direction must be resilient to input power and wavelength variations caused by the burst-mode nature of the TDMA upstream traffic together with the dynamic wavelength assignment protocol. Specifically, saturation-induced gain transients should be minimized in order to avoid the introduction of significant impairments.

The burst-mode nature of the upstream path in conjunction with the nonuniform loss distribution in the ODN, leads to a very large burst-to-burst dynamic range in the power of the signals arriving at the OLT receiver (typically 15 dB or higher without dynamic power leveling). Consequently, a burst-mode receiver is required, that can quickly adjust the gain of its electrical amplifiers on a burst-by-burst basis to avoid distortions and equalize the power of the bursts. The long reach (100 km or beyond) targeted by these converged metro-access networks in combination with a bit rate of 10 Gb/s or beyond requires the use of some form of chromatic dispersion compensation. This is especially true in the upstream direction of the network, where highly optimized transmitter technologies may be too costly. An efficient solution is to use burst-mode electronic dispersion compensation (BM-EDC), whereby the impairments originating from chromatic dispersion are compensated electronically. Critically, BM-EDC requires a linear optical receiver and therefore the upstream links of the LR-PON require a linear burst-mode receiver (BMRx) in the OLT.

Integrated Tuneable Transmitter

The laser in the ONU at the customer site, which acts as the light source for the upstream transmission, needs to be widely tuneable in order to allow the channel reconfigurability targeted by these architectures. If we consider an LR-PON system in the C-band with around 40 channels, the laser would need to tune to the 50 GHz ITU-T grid over at least half the C-band when half of the C-band is assigned to the upstream channels and the other half to the downstream, or over the whole C-band on the 100 GHz grid when the upstream and downstream channels occupy alternate channels. While fast (sub-μs scale) tuning speed is not required for the proposed network architecture, an important requirement is the need for the laser wavelength to be accurately set within the bandwidth of the assigned channel when the ONU output is active for transmission, in order not to interfere with the traffic of other PON channels.

The transmitter module should also be able to blank its output power to very low light levels, when not transmitting a burst, or when tuning to a new channel, in order to avoid interference with the traffic of other users. Additional design parameters are: sufficiently narrow linewidth to enable 10 Gbit/s transmission over 100 km or more standard single-mode fiber (less than a few 100 MHz), single-mode operation with sufficient side-mode suppression ratio (SMSR $>$ 35 dB), and an overall transmitter output power into fiber of at least $+5$ dBm. It is important to minimize the cost of the ONU equipment in a PON hence the tuneable upstream transmitter has to be manufacturable at low cost, including the tuning scheme.

While commercial tuneable transmitters exist, their cost ranges from several hundred euro to many thousands and it is clear that costs must be reduced for widespread adoption in an access application. Integration is also likely to be required in order to reduce the ONU cost, as packaging of discrete components is a time-consuming and expensive step. The requirement to operate at 10 Gb/s with the chirp performance necessary to transmit over the reach target of the LR-PON means that directly modulated lasers are not a viable solution and external modulation will need to be employed. EAMs can provide a solution as an integrated modulator that allows operation at 10 Gb/s with low chirp. In order to overcome the loss introduced by the EAM and to allow for high-power operation of the transmitter an optical amplifier will also need to be in-

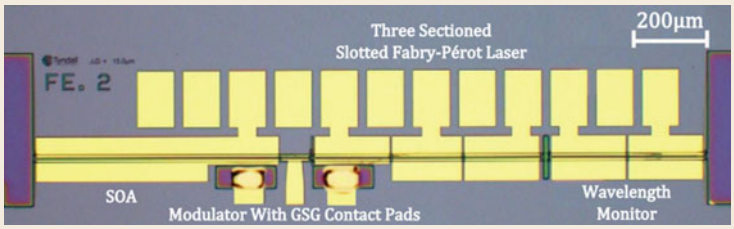


Fig. 30.11 Monolithically integrated tuneable transmitter design employing a SFP tuneable laser [30.43]

cluded in the transmitter with an SOA being a practical choice for ease of integration with the other semiconductor components. The SOA also provides two other key functions for the ONU transmitter: the optical gating function to blank the output of the transmitter when other ONUs on the same channel are transmitting or when the ONU is tuning [30.44]; and fine control of the output power for power leveling if that is implemented. In comparison with a transmitter based on a directly modulated laser, such as in currently deployed PON transmitters, the integration of the EAM and the SOA allows for a simpler operation of the tuneable laser. This structure improves the wavelength stability of the tuneable laser since the laser is always biased with constant currents and constant power even when the ONU is blanked.

The most common designs of widely tuneable lasers used presently are the modern variants of the sampled-grating distributed Bragg reflector (SG-DBR) lasers [30.45] and the super-structure grating DBR (SSG-DBR) lasers [30.46]. The main disadvantage of these types of lasers is their fabrication complexity which includes multiple epitaxial growths requiring high tolerance steps for creating the grating structures, which in turn leads to a higher cost. Over the past few years, new designs for tuneable lasers have been developed by etching slots into Fabry–Pérot lasers. The slots are etched through the upper waveguide, but not through the active quantum well region. This introduces reflections into the cavity and perturbs the longitudinal mode. By injecting currents into each section the local gain and refractive index can be changed and the overall gain and phase resonances can be adjusted thereby achieving tunability of the device [30.47]. The slotted Fabry–Pérot (SFP) lasers require only a single epitaxial growth process and standard lithography techniques, which greatly reduces the fabrication complexity and gives high yield [30.48, 49] and should achieve a production cost similar to Fabry–Pérot lasers. Deep etches in the waveguide can also be used to integrate other components with the SFP tuneable laser and since this is done lithographically it will not increase the fabrication costs substantially. Figure 30.11 shows a picture of a complete transmitter design integrating the SFP laser with a wavelength and power monitor [30.50], an electroabsorption modulator (EAM) for data modulation and a semiconductor optical amplifier for control of the optical output power.

Another time-consuming and expensive procedure for tuneable laser manufacturing is precharacterization at source. A laser with 80 channels over a 50 GHz grid might require up to 30 min to be calibrated precisely in wavelength for one power level. Because of the interdependency between output power and wavelength, the calibration would need to be performed at each power level. As mentioned in Sect. 30.3.2, the coarse wavelength calibration of the tuneable laser at only one power level could substantially reduce the overall time required for factory calibration and hence the cost.

The main technical issues that need to be addressed by the wavelength referencing and control are rogue wavelength behavior, fine tuning, and SMSR control. Rogue behavior happens when an ONU transmits on a wavelength assigned to another set of users and hence interferes with the TDMA traffic. This could happen maliciously to disrupt the traffic in the PON or because of faulty equipment, ageing, etc. The ONU laser might start up on the wrong channel or it might drift outside its assigned channel during operation. The ONU can be actively controlled to avoid this behavior, for example with feedback from the OLT [30.39, 51, 52]. A passive control could also be used by introducing filters. SMSR can also become a critical issue when a number of ONU transmitters working at different wavelengths are multiplexed on a PON, as the out-of-band side-modes from a transmitter can fall within the band of another channel. Assuming a worst-case scenario, the strongest side modes from the various active ONUs could all overlap within the band of another active channel and the noise generated would effectively add. As an example, 40 channels would add up as roughly 16 dB reduction in SMSR. High-quality tuneable lasers might have an SMSR high enough to allow operation even in the worst case described. However, the use of a cheaper tuneable laser implies that the SMSR might not be as controllable and that drifts in temperature and ageing might require fine tuning of the laser to minimize the impact of poorer SMSR performance.

Figure 30.10 shows an example of one possible full ONU structure, which includes a filter between the network fiber and the integrated tuneable transmitter. This filter serves several functions: it reduces injected power from the downstream channels into the ONU laser; it suppresses out-of-channel emission from the ONU laser in the upstream direction; it can be used to select calibrated downstream wavelengths; and it can prevent spurious wavelengths being transmitted upstream when the ONU is tuning to the correct wavelength. As the filter suppresses out-of-channel emission from the ONU laser, the SMSR requirements of the tuneable laser are reduced. For example, if the tuneable laser, as an individual component, has only 25 dB SMSR, the filter can easily increase that to 35 dB. This lower SMSR requirement of the ONU laser component also helps to simplify laser design and reduce cost.

Transient Stabilized EDFAs and Linear SOAs

Important elements of the LR-PON are the optical amplifiers in the AN at the old LE/CO sites. These must be carefully designed in order to support the long reach and the high number of users targeted by the LR-PON architecture. A key parameter is the noise figure of the amplifier which must be as low as possible in order to maintain the OSNR (optical signal-to-noise ratio) of the upstream channels within acceptable levels. The amplifiers should also provide a high gain and high output power in order to overcome the high splitting loss of the access and the long metro section of the LR-PON. Solutions that employ multiple single-channel amplifiers multiplexed in the AN have been demonstrated to be able to support long reach (beyond 100 km) and large numbers of users per channel (512 or 1024) [30.10]. However, from a system flexibility point of view the best option would be to use a single multichannel amplifier for each direction in the LE, which would maintain the transparency of the AN. Cost and form factor of the AN equipment would also be lower for the single multichannel amplifier solution.

EDFAs are very strong candidates for this application due to their good overall performance in terms of low noise figure, high gain and high output power and the ability to provide these characteristics in a system with a large number of channels. EDFAs also have the advantage of being mature components that are widely used in metro and core networks which makes them attractive for deployment in an access scenario.

SOAs are also an attractive amplification platform due to their small size, the potential for hybrid and monolithic integration, and the capability to be designed to operate in wavelength regions outside the C-band. The major disadvantage of SOAs is the distortion induced by the pattern-dependent gain saturation when they are operated in the saturation regime. Recently SOAs with large output saturation powers, and hence wide linear operation regions, have become commercially available [30.53]. Compared to earlier generations of SOAs, these devices also exhibit relatively low noise figures of the order of 7 dB. While the output power of these linear SOAs is sufficient for application in the AN of LR-PON systems, the gain of these devices is usually limited to around 12−15 dB, which requires the use of a cascade of at least two devices in order to achieve the gain required by the AN. An example of an AN configuration using these devices can be found in [30.54]. Despite the promising results obtained using linear SOAs, in the remainder of this section we will focus on designs of the AN based on EDFAs which, due to their superior noise figure and higher saturated output power, achieve the ultimate performance of the LR-PON systems.

Another important aspect of the amplifiers used in the upstream direction is that they need to be able to operate with high dynamic range burst signals. In the past EDFAs have been typically designed to operate with continuous traffic in conventional metro and core networks. The high dynamic range of the high-speed burst-mode upstream channels of a PON architecture can generate significant impairments due to saturation-induced gain transients in EDFAs unless suitable mitigation strategies are employed. However, the requirement to support channel add/drop in reconfigurable WDM networks containing ROADMs, has driven recent developments in EDFA transient control circuitry [30.55] that also provide solutions for supporting burst-mode traffic in LR-PONs.

The fast gain control in these EDFAs is usually based on a mix of feed-forward and feed-back gain control. The feed-forward part of the control acts on the pump laser by measuring the input power to the EDFA with a fast photodiode, followed by a fast control scheme which reduces the pump laser bias when the input power decreases. The feed-back part is used to improve the precision of the gain setting and operates on slower time scales using the power readings of the input and output power. The fast gain control in these EDFAs is capable of reducing the majority of the gain transient. However, there is a residual transient usually of the order of +/−0.5 dB depending on the duration of the bursts and the wavelength [30.55, 56]. In LR-PON applications where a number of these EDFAs are concatenated (particularly for the amplified chain solution for rural areas see Fig. 30.5) the residual transient could potentially accumulate. Hence, it is important to test the amplifier chain using burst patterns that are as close as possible to realistic traffic patterns.

As an example in [30.56] the impact of upstream burst traffic on EDFAs located in an LR-PON with chained ANs is examined using an upstream signal, generated by a pair of ONUs creating alternating bursts of different powers, as the test data channels and ASE (amplified spontaneous emission) to emulate a fully loaded system with 40 channels, concatenating a total of 5 EDFAs. The amplitude of the ASE used to emulate the upstream channels was modulated asynchronously relative to the ONUs bursts with a 50% duty cycle causing a large change in the overall input power of the upstream EDFAs. The accumulated residual EDFA gain transient increases with the increase in the period of the add and drop events but tends to a constant value of around 1.2 dB for periods beyond 100 μs, which is similar to the maximum burst duration for current PON standards. In [30.56], the bit error rate (BER) is also measured in real time on the alternating bursts using an field programmable gate array (FPGA) setup, showing only a marginal impact of the 1.2 dB residual EDFA transient.

Burst-Mode Electronic Dispersion Compensation (BM-EDC) in LR-PONs

The problem of the dispersion compensation in the LR-PON can be addressed in several different ways. Dispersion compensation fiber (DCF) or dispersion compensation modules using fiber gratings can be used at the amplifier locations in order to fully or partially compensate the dispersion in the downstream or upstream link. However, the DCF or fiber gratings would need to be predetermined at deployment stage and will likely require specific planning and engineering depending on network topology and the customer distribution. The use of predistortion or chirp optimized transmitters to partially compensate the dispersion of the link has also been demonstrated in LR-PON in combination with other dispersion compensation techniques.

Another solution to the dispersion compensation problem is the use of EDC at the receiver. EDC has the advantage that it adapts to the dispersion of the link and hence it does not require preplanning at the deployment stage. There are several ways to implement EDC for optical systems, from the more complex schemes using coherent receivers followed by digital signal processing (DSP), which provide the best performance with thousands of km of standard single-mode fiber compensation, to the simpler schemes using a direct detection receiver and analog or mixed-signal feed-forward equalizer (FFE) combined with a decision-feedback equalizer (DFE). With direct detection, phase and polarization information is lost, which limits the dispersion tolerance achievable by the EDC compared to the coherent receiver schemes or to DCF. However, dispersion tolerance of around 100 km as required in LR-PON is achievable at 10 Gb/s using direct detection as demonstrated experimentally in [30.57]. Analog or mixed-signal fractionally spaced FFE/DFE have found wide application in optical receivers, where the use of deep submicron CMOS (complementary metal oxide semiconductor) technologies enables circuits with a relatively low power consumption, around 1 W, to be implemented. For access applications, where cost and power consumption are major concerns, the FFE/DFE-based EDC has many advantages.

The use of EDC in the LR-PON downstream direction would require an increase in complexity at the ONU receiver compared to currently available PONs, but it could make use of relatively standard approaches that are common in electrical backplanes. In the upstream direction the cost and complexity added to the OLT receiver can be shared by all the users of the LR-PON. EDC also has the advantage of presenting no insertion loss, a negligible physical volume, reduced CapEx and inventory cost since the same chip can adapt to all the dispersion values of the LR-PON and, depending on the EDC implementation choice, a negligible additional power consumption.

Recently, there has also been interest to support larger differential reaches in PONs, as evident in recent updates to the ITU-T GPON and XGPON standards which now support a 40 km differential reach rather than the conventional 20 km [30.58]. If EDC can be made adaptive from one burst to the next (BM-EDC), high differential reach can be supported [30.57, 59, 60]. Experimental demonstrations of FFE/DFE-based BM-EDC in optical links suitable for LR-PONs using off-line processing have been presented in [30.60], where the BM-EDC can also be used to correct distortion introduced by the ONU transmitter.

Recent work has shown that the BM-EDC chip can replace the BM-CDR (clock and data recovery) chip and hence chip size and the power consumption target for the BM-EDC functionality is even further relaxed [30.60]. As explained in detail in [30.60] the FFE is effectively performing the same function as an oversampling BM-CDR which requires selection of a new optimum clock phase for each burst. This is an important result as it also shows how the BM-EDC is effectively substituting for the BM-CDR and hence it is not increasing the number of elements in the receiver chain compared to an OLT without BM-EDC.

Several challenges need to be overcome in order to realize BM-EDC. Firstly, BM-EDC requires a linear receiver front-end, in order to accurately preserve the pulse-shape of the incoming optical signal and hence be able to compensate for the effects of dispersion and bandwidth restriction. The upstream receiver of a PON

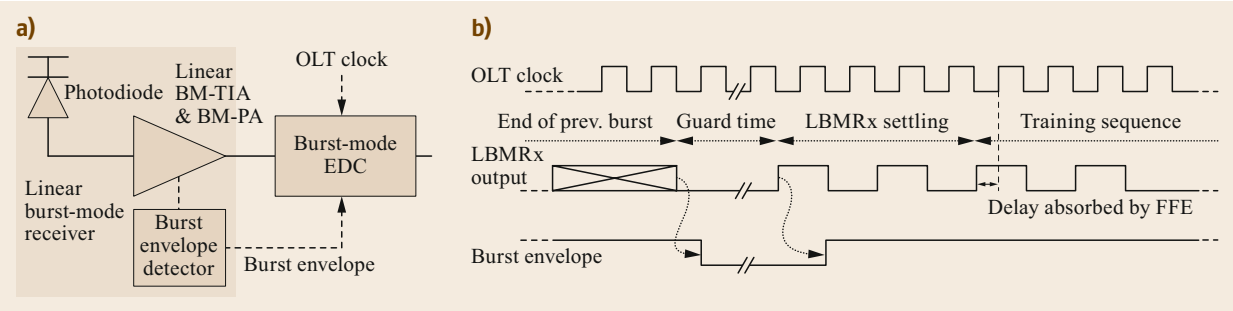


Fig. 30.12a,b OLT receiver with BM-EDC **(a)** and timing relationships **(b)**

is a specialized component that is able to work with a wide dynamic range of the optical power of the burst coming from different ONUs, which could be of the order 20 dB or more if power leveling is not implemented, and at its output equalizes the amplitudes of the bursts. Such a component is called a burst-mode receiver and in today's PON the amplitude equalization is obtained using several stages of amplification with variable gain adjusted based on the average optical power of the burst. However, these receivers usually employ limiting amplifiers, which introduce nonlinear distortion on the pulse shape, reducing the efficacy of EDC. BM-EDC therefore requires the use of a linear BMRx, rather than today's burst-mode receivers which are limiting in nature. Linear receivers working in burst mode at 10 Gb/s have been recently demonstrated in the literature [30.61]. Figure 30.12a shows a top-level diagram of an OLT receiver implementing BM-EDC. The linear BMRx linearly amplifies the incoming bursts such that these all have the same amplitude. Additionally, the linear BMRx can provide a burst envelope signal to the BM-EDC chip for timing purposes. Figure 30.12b shows the timing relationships between relevant signals: the preamble of each burst consists of a first part which is used by the linear BMRx to properly set its gain, and a second part which contains the training sequence for the BM-EDC.

BM-EDC also requires a training sequence embedded in the preamble of each burst [30.59]. The number of training symbols needs to be small in order to minimize the impact on traffic efficiency. Tap adaptation within 500 bits has been observed experimentally [30.60] for the relatively simple case of the widely used least mean square (LMS) algorithm used for tap adaptation, which is compatible with burst-mode applications provided the LMS algorithm is implemented in an IC (integrated circuit) with low additional overhead. The results of these experiments can be seen in Fig. 30.13 corresponding to a chromatic dispersion of 1400 ps/nm. When using the more efficient recursive least squares (RLS) algorithm, faster convergence

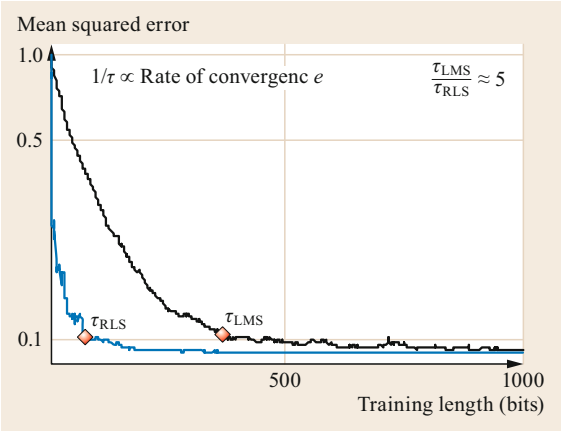


Fig. 30.13 Comparison of LMS and RLS algorithms

within around 100 bits can be observed at the cost of an increased algorithm complexity. The overhead introduced in the preamble by the training sequence of both algorithms is compatible with current PON protocols and further trade-offs between algorithm complexity and training sequence length are also possible [30.60].

30.4.2 Physical Layer System Design

The system design of an LR-PON physical layer is substantially different from a conventional standard-reach PON. The main difference is the introduction of optical amplifiers to increase both the reach and the split ratio supported by the system. The location of the optical amplifiers is bounded by deployment considerations and the network topology as introduced in Sect. 30.2. Optical amplifiers introduce optical noise, in the form of ASE, which generates a system impairment that is not present in conventional standard-reach PONs. The location of the optical amplifiers also introduces an intrinsic asymmetry in the LR-PON, where optical preamplifiers located before the OLT receiver at the MC-node can be used in the upstream direction, since they are shared by all users of the LR-PON, while in the down-

stream direction optical preamplifiers are less likely to be employed in the ONU receiver for cost reasons. As a result, the upstream receiver of LR-PONs works in a regime where the sensitivity is affected by both the thermal noise and the optical noise present on the upstream signal, while the downstream receiver sensitivity is mainly affected by thermal noise. In the following subsections the specifications of the upstream and downstream transceivers for LR-PONs will be first reviewed and then these specifications will be used as the starting point of the system level design of the LR-PON physical layer for two different topologies for dense urban areas and sparsely populated rural areas.

Upstream and Downstream Transceiver Specifications

The performance of transmitters and receivers for both the upstream and downstream link are of key importance for the design of the full system. In the PON standards the maximum output power varies with the classes of the ONUs, with powers up to +9 dBm. Usually these high powers are obtained using directly modulated lasers. However, the maximum output power that might be available from a widely tuneable transmitter are slightly lower due to the more complex structures required. The SOA in the output of the transmitter is the key limiting element in terms of power. A typical 1 dB saturation power for a monolithically integrated SOA is around +10 dBm. This corresponds to the peak output power before the SOA starts to saturate and create distortion on the signal due to gain patterning and hence corresponds to the maximum power in the *1* level of the output signal. Higher power SOAs are available as discrete components, but they are still highly specialized components that require a large bias current (around 500 mA) and their prospect for integration is not clear. Considering a 2 dB coupling loss combined with loss for diplexer or a circulator and a modulation loss of 3 dB, the output power of an SOA-based widely tuneable transmitter is in the order of +5 dBm.

In conventional nonamplified PONs, the burst-mode receiver works in a thermal-noise-limited regime, where the sensitivity is affected by the thermal noise of the photodiode and by the noise figure of the variable gain electrical amplifiers. However, in LR-PONs, because of the use of optical amplifiers in the upstream link, the burst-mode receiver of the OLT works in a different regime, where the sensitivity is affected by both the thermal noise and the optical noise present on the upstream signal. While the optical noise is inherently present in an optical amplified link, the OLT burst-mode receiver cannot work in the fully optically preamplified regime, where the thermal noise becomes negligible, because of the requirements of maintaining a large dynamic range. If power leveling is implemented then the dynamic range performance can be relaxed and the full optical preamplified regime could be exploited. However, in the following description we consider a burst-mode receiver for a PON system operating with a large dynamic range.

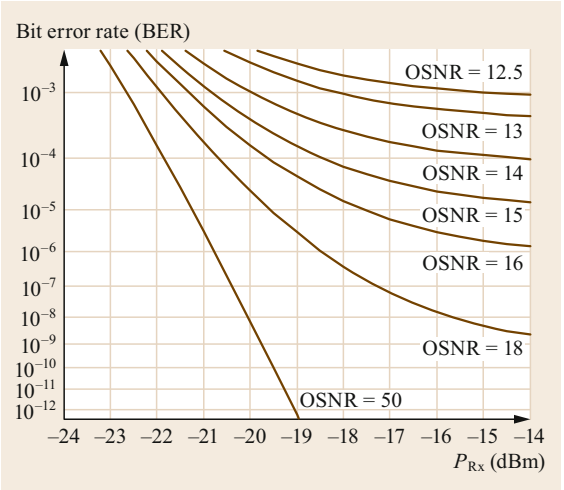


Fig. 30.14 BER as a function of the received power for varying levels of OSNR for a typical PIN-based 10 Gb/s receiver (electrical bandwidth 7.5 GHz, ER 9 dB, receiver sensitivity at 1×10^{10} of −19.5 dB, optical filter 3 dB bandwidth 25 GHz)

In order to estimate the performance of a burst-mode receiver working in the regime where thermal noise is significant, we can model a receiver working at 10 Gb/s, with an equivalent noise bandwidth of 7.5 GHz, which provides a sensitivity of −19.5 dBm at 1×10^{-10} BER for an extinction ratio of 9 dB with no optical noise present (OSNR > 50 dB) as shown in Fig. 30.14, which is a typical performance for a PIN-based 10 Gb/s receiver. We can now include the effect of the optical noise, by reducing the OSNR of the incoming signal assuming a bandwidth of 25 GHz for the optical filter, which is typical of filters used for 50 GHz-spaced WDM signals. The FEC (forward-error correction) threshold of 1×10^{-3} BER is reached for different values of the received power depending on the OSNR of the upstream signal. However, in a burst-mode system without power leveling, the minimum received power is limited by the dynamic range of the incoming signal and by the overload level of the burst-mode receiver, which is fixed by its design. If we take the example of [30.61], an overload at 1×10^{-3} BER higher than 0 dBm has been demonstrated for a linear burst-mode receiver and an experimental demonstration of an LR-PON with a dynamic range of 18 dB has been presented in [30.54] using such a receiver. Using these

results we can assume a received power of −18 dB for the lowest power of the dynamic range which results in a minimum required OSNR of 13 dB at the FEC threshold of 1×10^{-3} BER, without including system margin or nonideal receiver behaviors.

In the downstream link transmitter, the maximum output power does not directly affect the performance of the link because of the optical amplifiers located in the MC-nodes and ANs prior to the backhaul fiber loss and the high split loss ODN. The typical output power of around 0 dBm of commercially available tuneable transmitters is sufficient to maintain a high OSNR value. Because ONU downstream receivers are unlikely to employ optical amplification for cost reasons, their sensitivity will hence be limited by thermal noise alone. It should also be noted that because the downstream link operates in continuous mode the transmitter does not present any special burst-mode requirements. The NG-PON2 standard specifies ONU receivers for the downstream link at 10 Gb/s with a tuneable filter with a sensitivity of −28 dBm at 1×10^{-3} BER [30.40]. Sensitivities of −30 dBm were demonstrated in [30.54, 56] using a receiver with an APD (avalanche photodiode) and a tuneable filter, which complies with the values required by the NG-PON2 standard accounting for margin and manufacturing tolerances.

Dense Urban Networks

In a densely populated area, where there are a large number of subscribers within a 10–20-km reach of an LE/CO site, a single AN can be used between the MC-node and the ODN. This is the most common LR-PON design in the literature and it is sometimes referred to as tree-like or *lollipop* LR-PON topology (Fig. 30.16). As discussed earlier in Sect. 30.2.3 the AN can be dual parented for resiliency and protection purposes, with the protection path usually giving the more stringent constraint in terms of length. Overall reaches in the region of 100 km need to be targeted in order to support the node consolidation in the core network as discussed in Sect. 30.1.1.

The fiber link between the MC-node and the AN (usually referred to as backhaul link) can be relatively long and can introduce a large loss, for example 27 dB for 90 km considering end-of-life loss for the deployed fiber of 0.3 dB/km. Conversely, the ODN also introduces high loss because of the need to support a large number of users (up to 1024), which corresponds to 35 dB of loss considering the $10 \log(1024)$ dB splitting loss plus excess splitter losses of order $0.5 \log_2(1024)$ dB, and added to this is the ODN fiber loss. The design of the LR-PON should also consider the intrinsic asymmetry of the LR-PON, where optical preamplifiers can be used in the upstream direction before the OLT receiver at the MC-node, since they are shared by all users of the LR-PON, while in the downstream direction optical preamplifiers are less likely to be employed in the ONU receiver for cost reasons.

Because of this asymmetric design for the two signal directions, the downstream will be mostly constrained by the sensitivity of the ONU receiver and in turn by the power launched in the ODN from the AN, which requires the use of high-power optical amplifiers and a multichannel design. Conversely, the upstream link is constrained by the OSNR and the required dynamic range at the OLT burst-mode receiver. The OSNR of the upstream signal is mostly affected by the optical amplifier in the AN, which should therefore have a low noise figure. However, since the link towards the MC-node can also present a large loss, the optical preamplifier in the MC-node will further reduce the upstream OSNR and it needs to be taken into account in the system design. The dynamic range between consecutive bursts of the upstream signal also introduces a stringent requirement on the performance of the OLT burst-mode receiver, particularly if power leveling, outlined briefly in Sect. 30.3.2, is not implemented.

Figure 30.15 shows a simple and effective design for the AN where a 4×4 coupler part of the ODN split is included in the node to separate and aggregate the downstream and upstream path optical amplifiers and also to provide access to the protection path optical amplifiers [30.54]. At the metro/core node side, optical amplifiers are used to boost the downstream before it enters the backhaul fiber and to amplify the upstream before the OLT receiver. A fixed DWDM component used in conjunction with an optical switch provides the wavelength routing flexibility to and from the OLTs.

In the upstream path the power of the ONU transmitter can be fixed at +5 dBm, which is a reasonable assumption for a low cost widely tuneable transmitter. The power at the input of the EDFA in the AN is limited by ODN loss, which limits the OSNR of the upstream signal after the AN. The ODN loss depends on the fiber present in the ODN and on the split ratio, which includes the 4×4 coupler inside the AN. For the following calculation a worst-case ODN fiber length of 20 km is assumed. The gain of the EDFA in the AN can be determined by the difference between the output power launched into the backhaul fiber and the input power. The power launched in the backhaul fiber is limited for practical reasons in order to reduce nonlinear effects on the upstream signal. With this in mind, we limit the power of the loud bursts to less than +3 dBm, with slightly different powers for the three split ratios considered due to the different accumulation of the differential loss in the ODN. This corresponds to the upper limit of the dynamic range and the power

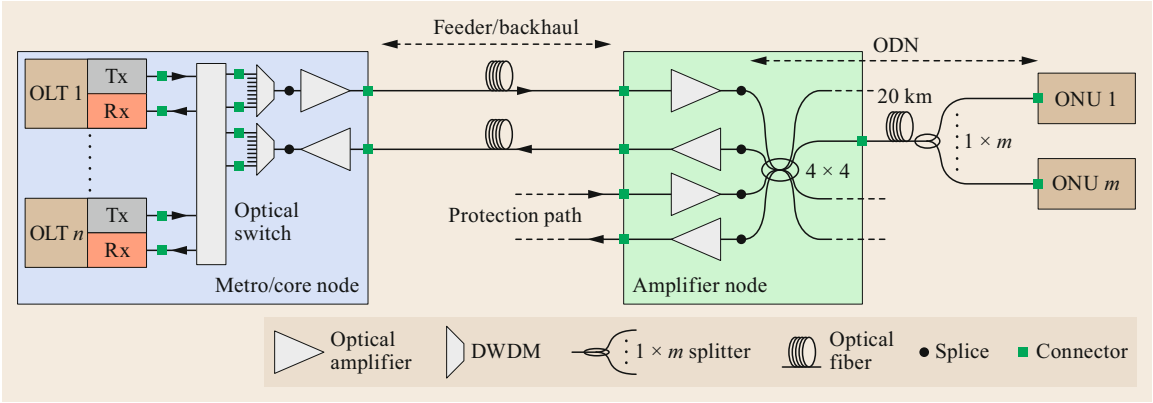


Fig. 30.15 Physical layer description of the tree-like architecture

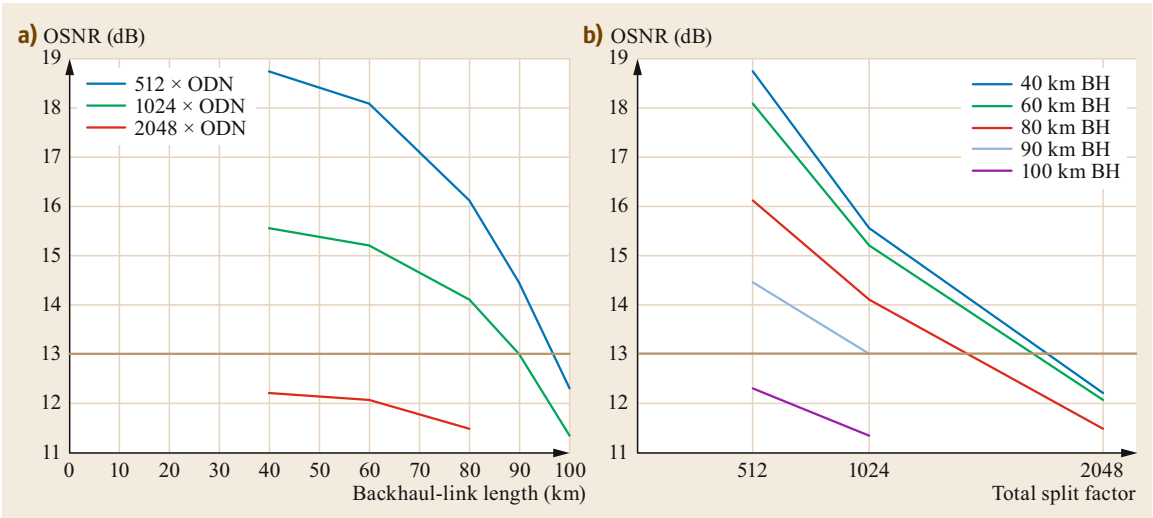


Fig. 30.16a,b *Lollipop* architecture OSNR at OLT receiver in upstream direction as a function of the backhaul fiber length **(a)** and as a function of the overall ODN split ratio **(b)**

of the soft burst can be obtained by subtracting the dynamic range. The dynamic range is considered to be introduced only by the differential loss of the splitters/couplers, connectors, and splices in the ODN path. For the three overall split ratios considered, 512, 1024, and 2048 the dynamic ranges used in the following calculation are, respectively, 9.2, 10.2, and 11.2 dB. The OSNR of the upstream signal is further degraded by the EDFA in the CN and the effect on the OSNR depends on the loss of the backhaul link. The gain of the EDFA in the CN is then determined by the input power required by the burst-mode receiver for the soft packets (highest attenuation path). In other words, the dynamic range of the upstream is adjusted to fit within the dynamic range of the burst-mode receiver using the EDFA gain.

By using the design in Fig. 30.15 with the values in Table 30.2 for the component losses and the amplifier gains in Table 30.3 as a function of the split ratio, the OSNR of the soft burst of the upstream signal can be calculated as a function of the split ratio and of the length of the backhaul link as presented in Fig. 30.16a,b. Considering a minimum required OSNR of 13 dB for the soft burst it is clear the maximum split achievable by the long-reach PON is 1024. The results also show that the length of the backhaul link plays an important role beyond 80 km both for 1024 and 512 split ratio. For a link length of 90 km and a split ratio of 1024 the OSNR is just equal to the required 13 dB with no margin, while for 100 km the required OSNR cannot be met by any split ratio analyzed. Considering that the calculation included the loss of 20 km of fiber in the ODN, an overall reach of 100 km can be achieved for this LR-PON design with a split ratio of 1024 conserving a moderate margin.

Table 30.2 Nominal and worst-case losses of components used in the LR-PON designs

	Insertion loss (nominal)	Insertion loss (worst case)	
1×2 Coupler 80%	−1.0	−1.2	dB
1×2 Coupler 20%	−7.0	−8.0	dB
4×4 Coupler	−6.0	−7.0	dB
Connector	0	−0.5	dB
Splice	0	−0.3	dB
Band filter	–	−0.5	dB
Optical switch	2	2	dB
Wavelength mux/demux	6	6	dB
Fibre loss end of life (EOL)	–	−0.3	dB/km

Conversely, the design of the downstream link is determined by the sensitivity of the receiver in the ONU and by the ODN loss. As mentioned above, the NG-PON2 standard specifies ONU receivers at 10 Gb/s with a tuneable filter with a sensitivity of −28 dBm at 1×10^{-3} BER [30.40], which is also used in these calculations. The AN EDFA output power can be calculated for the various maximum ODN losses and the receiver sensitivity, while the gain depends on the input power. For 512 and 1024 split ratio and 20 km of ODN fiber the output power of the EDFA per channel needs to be at least 9.5 dBm and 13 dBm, respectively. Because of the 4×4 coupler located in the AN, the power at the output of the AN is ≈ 7 dB lower and hence around 2.5 dBm and 6 dBm, respectively, which reduces the concerns about eye safety levels and nonlinearities in the ODN fiber.

Since the system targets multiwavelength operation, with up to 40 channels in each direction, the overall power of the EDFAs is 16 dB higher than the single channel power, which brings the overall power for the EDFA in the AN to +25.5 and +29 dBm, respectively, for 512 and 1024 split ratio. The EDFA power required by the 1024 split ratio is particularly high, but still achievable with commercially available devices. It should be noted that the EDFAs in the downstream path are always operating in continuous mode which simplifies the requirements on the pump control scheme and the overall complexity of the EDFAs. The downstream link from the MC-node to the AN does not introduce substantial challenges in terms of OSNR or power budget. With the large number of channels in the downstream direction, nonlinearities and in particular nonlinear crosstalk should be avoided by maintaining a launched power in the backhaul fiber around 0 dBm per channel, even if channel interleaving is not employed. Table 30.4 summarizes the output powers per channel and for all 40 channels for the EDFAs located in the AN for split ratios of 512 and 1024.

Sparse Rural Networks

The amplified chain design distributes the overall split that is achievable between a number of chained ANs and hence it is more suited for sparsely populated rural areas. The basic design criteria for the amplified chain, shown earlier in Fig. 30.4, is that the channel powers at the input of each AN should be the same for the whole chain. This implies that the optical amplifiers acting as line amplifiers along the chain should exactly compensate the loss of the fiber span after the AN as well as the internal loss inside the node due to the additional components, which include the asymmetric 20/80 couplers, connectors, splices, and dispersion compensation modules if present. Because in general the upstream channel is limited by a lower OSNR due to the large losses introduced by the ODN prior to the first stage of amplification, the asymmetric 20/80 couplers provide a better compromise compared to a symmetric 50/50 coupler in terms of the insertion losses of the upstream add-path. The structure of an AN for the amplified chain design is shown in Fig. 30.17 which also shows the flow of the downstream and upstream traffic. In practice, it

Table 30.3 EDFA gain as a function of the split ratio for different backhaul lengths

Split	AN upstream					CN upstream				
	40 km (dB)	60 km (dB)	80 km (dB)	90 km (dB)	100 km (dB)	40 km (dB)	60 km (dB)	80 km (dB)	90 km (dB)	100 km (dB)
512	25	25	25	25	25	13	19	25	28	31
1024	28.5	28.5	28.5	28.5	28.5	13	19	25	28	31
2048	32	32	32	32	32	13	19	25	28	31

Table 30.4 Output power required for the AN EDFAs

Split	Upstream 1 channel (dBm)	Upstream 40 channels (dBm)	Downstream 1 channel (dBm)	Downstream 40 channels (dBm)
512	−4	12	9.5	25.5
1024	−3	13	13	29

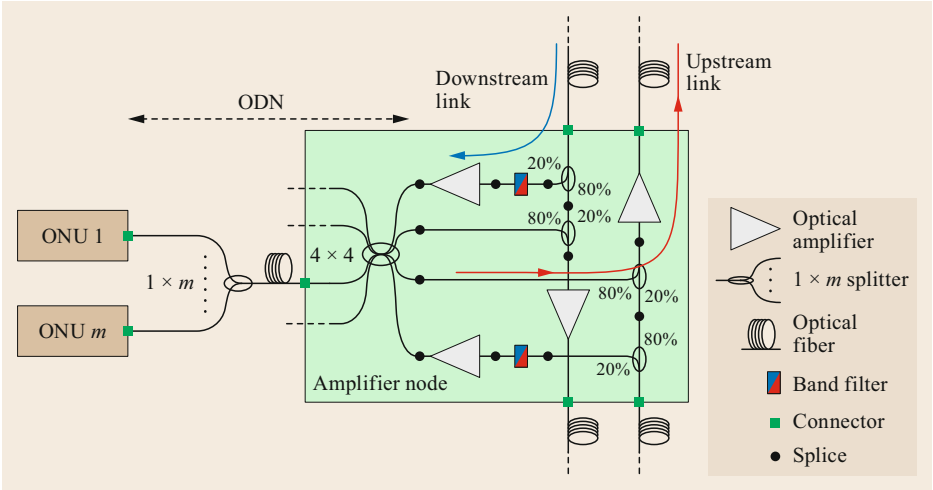


Fig. 30.17 Node structure for the chain of ANs

would be costly and inefficient to specify the optical amplifier gains for each AN. Instead schemes providing variable gain optical amplification should be employed and tailored to the span loss during commissioning of the system.

The lowest gain for the optical amplifier is given by the case where the fiber span is zero and the amplifier needs only to compensate for the losses of the two 20/80 splitters that are in the backhaul fiber path, one connected to the 80% port and the other to the 20% one. Considering realistic losses for the two 20/80 splitters and the splices between the components, their overall loss can be estimated to be 9.5 dB. The AN is assumed to be connectorized at the input and output, which adds 0.5 dB for each connector. The minimum gain of the optical amplifier is hence 10.5 dB. Conversely, the maximum gain is determined by the maximum length of the fiber span, by adding the fiber loss to the minimum gain. We consider a maximum distance between nodes of 60 km, which is a typical maximum length for unamplified spans in metro ring networks. Considering an end-of-life fiber loss of 0.3 dB/km, the maximum gain required from the amplifier is 28.5 dB. This value is easily achievable using EDFAs with a dual-stage configuration. To simplify the calculation of the system OSNR during the design phase we assume that the variable gain optical amplifiers are obtained using amplifiers with a fixed gain of 28.5 dB followed by an optical attenuator, which corresponds to a worst case in terms of OSNR.

The downstream and upstream path can be reversed for protection in case of equipment failure, when the protection CN becomes the active node. For this reason, the downstream and upstream paths in the chain should be designed with the same criteria and using amplifiers able to support the worst case for either path in terms of the maximum output power.

In the upstream path the power at the input of the first optical amplifier is constrained by the ONU transmitter power and by the losses in the ODN and inside the AN. Since the gain of the optical amplifier is effectively fixed by the span loss, the power per channel remains the same at the input of each optical amplifier in the ANs. The final optical amplifier in the chain is located in the MC-node and is used to overcome the loss introduced by the wavelength demux and optical switch and to align the power of the incoming upstream channels to the burst-mode receiver dynamic range.

Following these considerations the only variables in the design are the ODN loss (directly related to the number of users supported by each AN) and the number of ANs in the chain. These two parameters will determine the OSNR at the OLT receiver and hence the performance of the system. In the evaluation of the effect of the ODN split, it is critical to consider the worst-case loss introduced by the splitters and the fiber. Hence, we assume the maximum length of the ODN to be 20 km, which is typical of deployed PONs today, and then vary the splitting factor which causes the ODN loss to vary in discrete steps.

The OSNR at the OLT receiver can then be calculated for the three ODN losses, corresponding to 128, 256, 512 split ratios, as a function of the number of the ANs present in the chain (Fig. 30.18). As expected the OSNR drops more quickly for the larger split ratios because of the lower power at the input of the EDFAs. Considering the minimum OSNR for the soft packet of 13 dB, the maximum number of nodes supported by the 512 and 256 split ratios is 3 and 7 respectively, while the 128 split ratio can support more than 10 ANs. The total split factor supported by the network (ODN split multiplied by number of ANs), shows a very similar

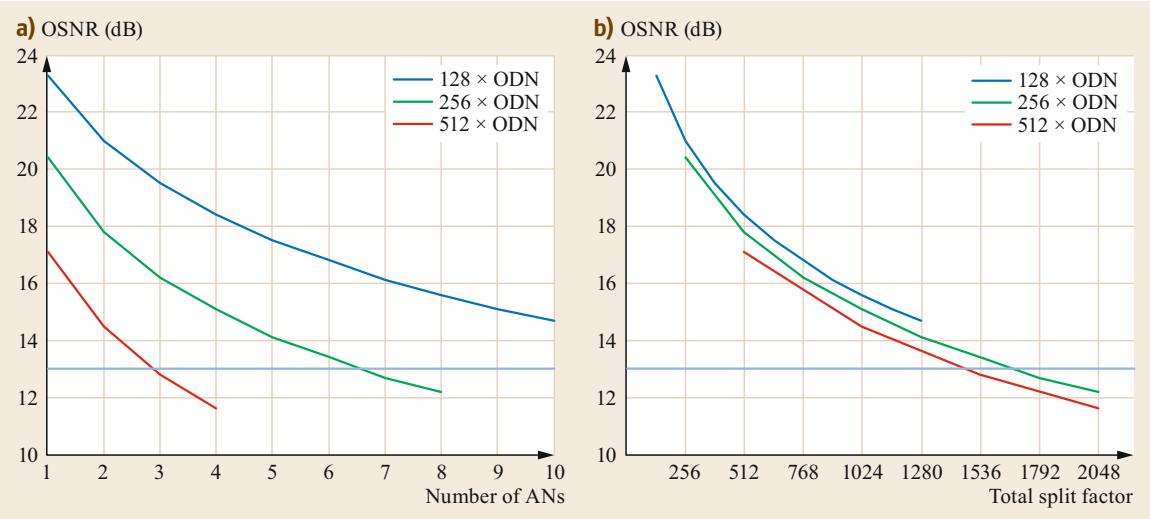


Fig. 30.18a,b Chain of ANs OSNR at the receiver in upstream direction as a function of the number of ANs **(a)** and for the total split ratio **(b)**

trend for the three ODN split ratios with only a small penalty for higher ODN split. The overall total split of the network is between 1536 and 2048. The lower ODN split can achieve a higher total split and is more flexible in supporting a wider geographical distribution of the users, but at the cost of a larger number of ANs and hence a larger capital cost. Because of the design rules utilized, the EDFA output power is higher for the lower ODN split ratios and decreases as the split increases. Considering the nominal channel power as the power in the center of the dynamic range, for each factor of 2 in the ODN split the EDFA output power decreases by 3 dB (Table 30.5).

Because the downstream and upstream path can be reversed for protection in case of equipment failure, the gain of the EDFAs in the downstream direction is the same as for the upstream direction and hence fixed by the span loss. The gain of the last EDFA before the ODN, which we will refer to as the $\text{EDFA}_{\text{drop}}$, is determined by the sensitivity of the ONU receiver and by the ODN loss. As for the *lollipop* case, we assume an ONU receiver sensitivity of −28 dBm at 1×10^{-3} BER. The $\text{EDFA}_{\text{drop}}$ output power can be calculated for the various maximum ODN losses and the receiver sensitivity, while the gain will depend on the input power. For similarity with the upstream, we fix the input power to the in-line EDFA in the AN to −24 dBm (the nominal input power for the 128-split ratio case). The $\text{EDFA}_{\text{drop}}$ gains for the 128, 256, and 512 split ratio are respectively 25, 28.5, and 32 dB. As the number of ANs increases the OSNR of the downstream signal is reduced. The OSNR is reduced to 18 dB for a chain of 10 ANs, which corresponds approximately to 1 dB power penalty at the 1×10^{-3} BER sensitivity level. This is important for the 128-split ratio case, where a larger number of ANs might be deployed. In this case, the power launched from $\text{EDFA}_{\text{drop}}$ should be increased to account for the OSNR power penalty. However, this does not significantly impact the system design, as the 128-split ratio case corresponds to the lowest power launched from the $\text{EDFA}_{\text{drop}}$ into the ODN.

Since the system targets multiwavelength operation, with up to 40 channels in each direction, the overall power of the EDFAs is 16 dB higher than the single channel power. In the proposed protection scheme, the upstream is also always transmitting in both directions of the chain, towards both the active and the protection CNs. This implies that the inline EDFAs in the AN have to be able to carry both the live downstream traffic and the copy of the upstream traffic directed towards

Table 30.5 Chain of ANs: Output power required for the AN EDFAs

Split	Upstream 1 channel (dBm)	Upstream 40 channels (dBm)	Downstream 1 channel (dBm)	Downstream 40 channels (dBm)	Total 40 channels (dBm)
128	4.5	20.5	4.5	20.5	23.51
256	1.5	17.5	4.5	20.5	22.26
512	−1.5	14.5	4.5	20.5	21.47

the protection CN. Since the traffic could be reversed in the chain in the event of a failure, all the inline EDFAs in the AN should be designed with the same specifications in terms of maximum output power. Table 30.5 summarizes the total power required by the EDFAs for the three ODN split ratios examined. There is only a marginal reduction in total power as the split ratio increases in the ODN because of the reduction in the upstream power, while the downstream power remains the same.

30.5 Experimental Results for LR-PON Architectures in End-to-End Networks

Some of the network concepts described in previous sections have been recently demonstrated experimentally both at the physical layer and in end-to-end system demonstration [30.54, 56]. The results reported here describe the end-to-end network demonstration performed using an SDN control plane implemented on the LR-PON and a simplified core network. The LR-PON configuration used in the end-to-end network demonstration is the tree-like, *lollipop* architecture using an adaptation of the XGPON protocol on the 10 Gb/s channels. The performance evaluation of different LR-PON architectures was also performed at the physical layer in order to study the limits in terms of achievable split ratio for the different channel types targeted: 10 Gb/s PON, 100 Gb/s coherent and fronthaul emulation. Two variants of the LR-PON architecture are analyzed experimentally, the dense area (*lollipop*) architecture and the sparse rural amplified chain design. The design criteria for the two architectures have been described in the previous sections, including a theoretical analysis of the scaling achievable for these systems over a large parameter space (i.e., split ratio, total reach, number of ANs in the amplified chain solution). For practical reasons, the experimental demonstration and analysis in [30.54, 56] were limited to varying the split ratio, while maintaining the same overall reach and, in the case of the amplified chain, the same number of ANs.

30.5.1 Demonstration of Protection and Dynamic Wavelength Services

The network architecture scenario implemented in [30.54] incorporates a flexible DWDM-TDMA LR-PON connected to a primary MC-node with a protection link to a secondary MC-node and a simplified core network, which is emulated by three nodes, implemented using electronic switches, interconnected by fiber links supported by commercial transceivers (Figs. 30.18 and 30.21). The network is controlled through a software-defined network (SDN) control plane, which enables dynamic service and capacity provision over the LR-PON in response to changing demand [30.25]. The overall SDN architecture follows the open network foundation (ONF) architecture and details of the implementations can be found in [30.25, 26, 62–64]. This testbed setup was used to demonstrate two main features:

1. The first is fast protection of a dual-parented LR-PON, which shows the ability of the network to support a level of resiliency typically required for enterprise and mobile backhaul applications and also allows implementation of failure recovery and load-balancing schemes, which substantially reduces cost in both IP and PON backup resources [30.65], by increasing the ability to share protection equipment across the network.
2. The second is the ability of the PON system to allocate new wavelengths dynamically to offer additional capacity to ONUs. Agile provisioning of wavelength channels is essential to support heterogeneous applications in the network, which can support highly differentiated quality of service, as they allow ONUs to switch dynamically between dedicated and shared channels, to suit their current requirements.

Protection Experiment

The protection experiment demonstrates a dual-parented LR-PON protection mechanism where backup OLTs are shared among PONs in an $N:1$ scheme [30.65], and the service restoration is provided over an end-to-end SDN-controlled topology. The first step of the protection scenario is the configuration phase where the controller carries out an initial phase of path-precomputation, setting up a backup path associated with the failure of a specific PON. The precalculation considers the input and output ports at the optical switch, the flow table configuration of the Openflow SDN switch (both access and core), and the configuration of the OLT flow table. For additional details of the PON activation mechanism operated by the OLTs and ONU, the reader can refer to [30.64]. This is followed by a failure event, triggered by using the optical switch to emulate a fiber cut in the backhaul fiber link between the primary OLT and the AN. Loss of signal in the upstream activates a countdown timer

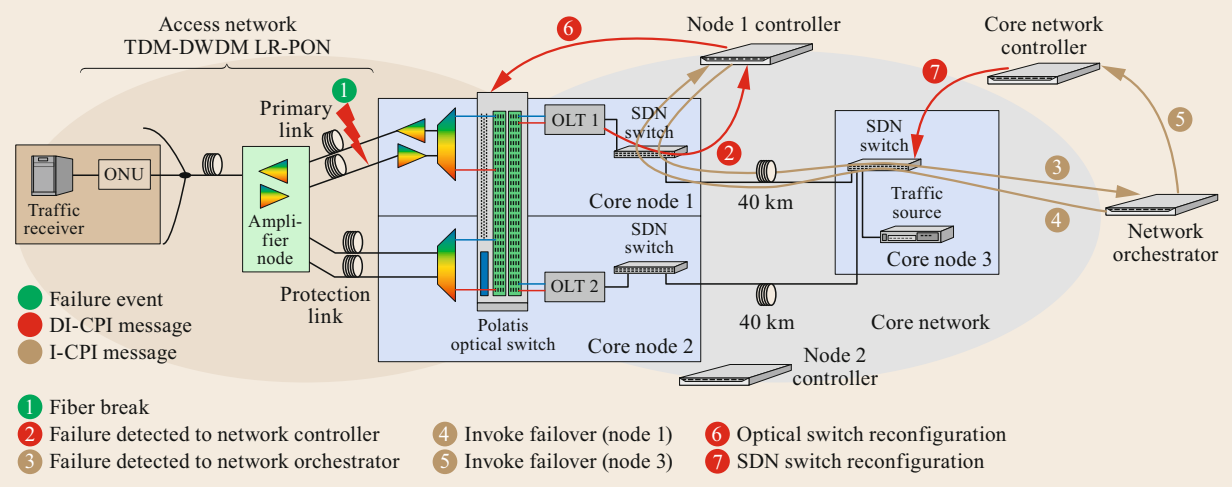


Fig. 30.19 Message exchange overview for the protection experiment

in the primary OLT, which on expiry generates a failure detection and an in-band alarm to the controller of node 1. The duration of this timer takes into account all normal silences on the PON due to the 1.25 ms quiet windows and 1.25 ms RTT over maximum distance supported by the protocol of 125 km, for a total of 2.5 ms. Node 1 controller informs the overarching NetO, which calculates a path to restore services to the ONUs. This is required because of the dual-parented nature of the protection, where the ONUs will be served by a different MC-node once protection is activated. Thus, data needs to be rerouted across MC-nodes and directed towards the protection OLT. Figure 30.19 gives a logical view of the messages exchanged by the control plane during the protection process.

The results of service restoration time for the SDN-control-plane-based protection mechanism are shown in Fig. 30.20. The average restoration time over 70 measurements was 41 ms. In order to understand the effect of centralizing both the NetO and the node controllers (NCs), the above results are compared with the case where the orchestrator and controllers are collocated within the MC-node by setting the emulated intracontrol plane latencies at zero. These results are shown in Fig. 30.20 as the *basic protection* line, which can be accomplished within 27.8 ms.

Figure 30.21 shows the breakdown of the various timings that comprise the 41 ms protection figure. The hardware monitoring at the OLT can detect a failure in the network in about 2.5 ms. A further 1 ms is taken for the alarm packet to be created and sent to the MC-node switch. The time needed by the protocol to reestablish downstream synchronization is between 2 and 3 ms, although additional time might be required for

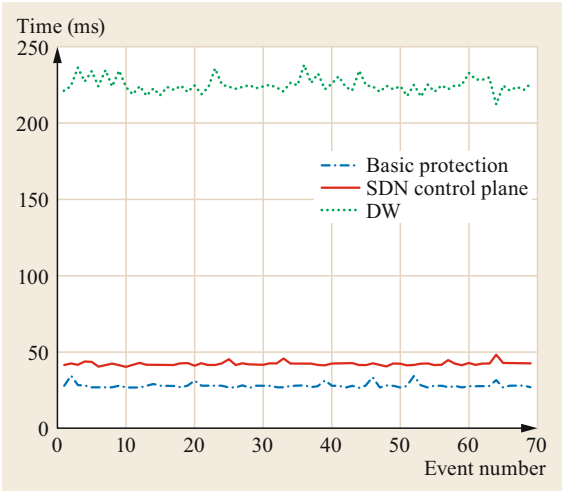


Fig. 30.20 Service restoration time for the protection mechanism and the DWA through the implemented SDN control plane

reranging [30.66]. Intra-control plane communication is carried out through a dedicated network which emulates latencies, of 4 ms each between OLT and NetO, as well as between NC and NetO. The latency and the processing times for both the NCs is also emulated as 5 ms each. The core network recovery happens in parallel to the access network recovery time. Within 15 ms of the failure, the optical and electronic switch components and the backup OLT have been instructed to reconfigure their protection paths and within 33 ms after the failure, the electronic switch components within the core and access are configured, and by 38 ms, the optical switch component is configured.

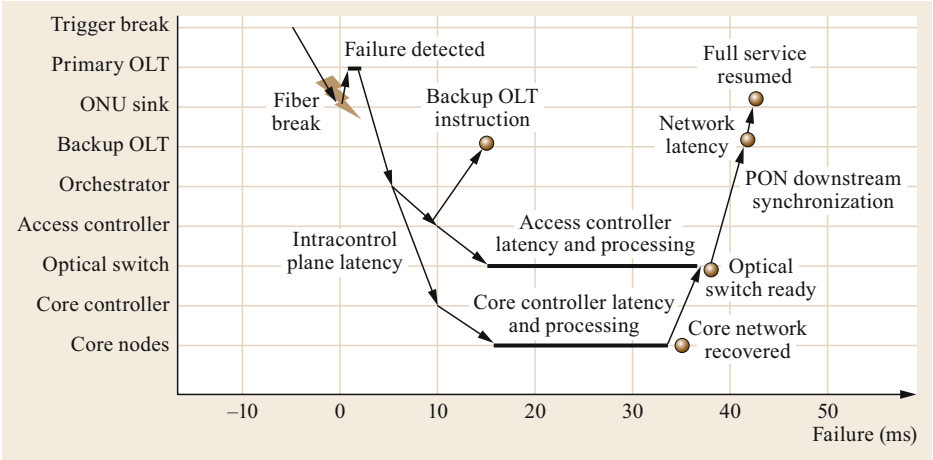


Fig. 30.21 Breakdown of SDN-control-plane protection timings

DWA Experiment

The DWA use case exemplifies how capacity constraints in one PON wavelength channel may be overcome by reallocating dynamically one or more end user ONUs to a different wavelength channel in order to assure quality of service or provide greater capacity. This could also be used for the opportunistic provision of high-bandwidth services (on-demand video and big data transfers), to specific PON users on a dynamic basis. Since the DWA use case is aimed at capacity provision, the wavelength and service reconfiguration times targeted are in the region of a few hundred milliseconds. It should be noted that although the wavelength assignment is not carried out at the granularity of individual burst transmission, the system is still referred to as a dynamic wavelength assignment as the change in wavelength is dynamically and automatically allocated by the controller as a response to an increase in user capacity, rather than being statically assigned by the network management plane.

The experiment setup for the control-plane message exchange is shown in Fig. 30.22. As in the protection use case, the NetO orchestrates the provisioning of the new path, according to its knowledge of the full end-to-end topology, by instructing the NC. The NC instructs both the optical switch and the secondary OLT to provision a new wavelength. The OLT provides an interface, through which the control plane can tune the OLT transceivers to a given wavelength. Since the ONU is remote from the control plane, the request to tune to a different channel is also performed through the OLT interface by the invocation of a custom LR-PON protocol message. Using a custom implemented physical layer operations, administration and maintenance (PLOAM) message, the primary OLT requests the ONU to tune to the wavelength provisioned by

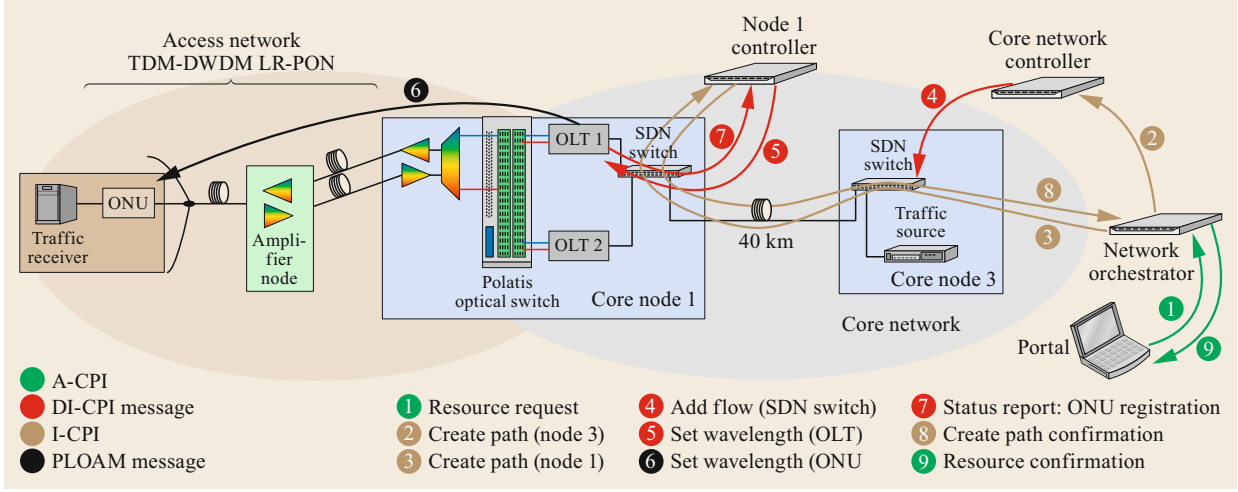


Fig. 30.22 Message exchange overview for the DWA experiment

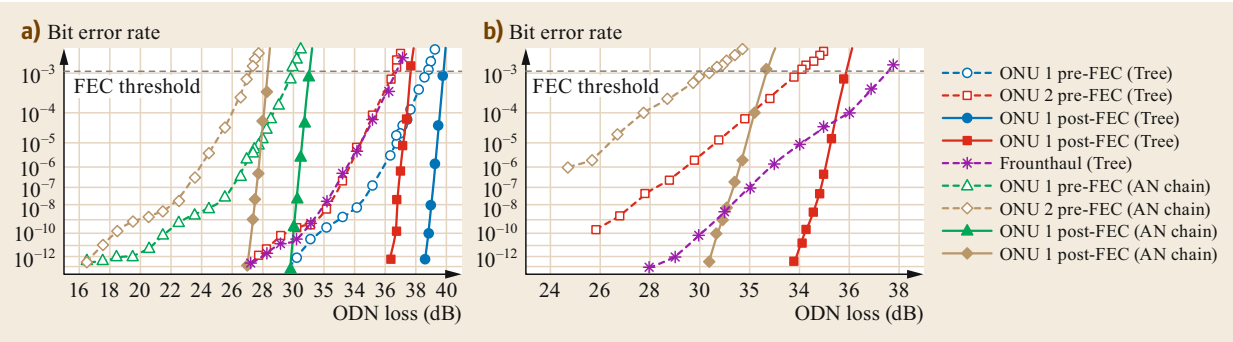


Fig. 30.23a,b BER versus ODN loss: **(a)** downstream, **(b)** upstream

the secondary OLT. The NC then acknowledges that it has completed the provisioning of the path through the access portion of the network. In the ONU, DWA is implemented by controlling the transceiver tuneable laser using an I2C (inter-integrated circuit) bus and controlling the tuneable filter through a universal asynchronous receiver-transmitter (UART) interface.

The DWA results shown in Fig. 30.20 refer to the service restoration time when, in response to an increase in traffic demand, the NetO instructs the core network controller and the NC to provision the new path and the ONU traffic is moved to a different PON channel, i.e., from event 4 to event 6. The average measured provisioning time of 225 ms is sufficient for service reconfiguration, but it could be further reduced by an optimized design of communication interfaces between the ONU FPGA and the tuneable components.

30.5.2 Demonstration of Multiservice Coexistence

The LR-PON architectures described in the previous sections are inherently capable of carrying simultaneously different channel types at the physical layer by utilizing DWA. Recent experimental demonstrations evaluated the limits in terms of achievable split ratio for the different channel types targeted for LR-PONs: 10 Gb/s PON, 100 Gb/s coherent Pt-Pt link and fronthaul emulation [30.54, 56]. Two variants of the LR-PON architecture were analyzed experimentally, the dense area (*lollipop*) architecture and the sparse rural amplified chain design. Commercially available gain-stabilized EDFAs are employed as in-line amplifiers in the AN of both architectures to reduce the impact of gain transients caused by the bursty nature of the upstream traffic [30.56]. The setups and the details of the implementations for both alternatives demonstrated can be found in [30.54, 56] but it is worth highlighting that the PON upstream channels were operated in realistic burst mode with FEC implemented on FPGAs using standard Reed–Solomon (RS) encoders and decoders, based on RS(248,216) [30.30, 67].

The results presented in [30.54, 56] and reported here in Fig. 30.23 reconfirm the design specifications derived in Sect. 30.4.2 and demonstrate that with today's commercially available components a 1024-way LR-PON with 100 km to 125 km reach can be achieved. Figure 30.23a shows the downstream bit error rate (BER) before and after FEC measured as a function of the ODN loss for the two architectures demonstrated using EDFAs. In both cases the two ONU's receiver sensitivities present a difference of less than 3 dB, because of variations in the receiver's performance. The dense area *lollipop* network design supported an ODN loss of at least 36 dB, which corresponds to a 256 split plus 20 km of fiber with a system margin of 2 dB and an overall split ratio of 1024 (256×4 due to the additional 4×4 AN split). The amplified chain design supported an ODN loss of at least 27 dB after FEC, corresponding to a 64 split plus 20 km of fiber in the ODN and an overall split ratio of 256 (64×4) per AN. Assuming a fully implemented network with four ANs this would correspond to a total split of 1024 (256×4).

Figure 30.23b, from the results in [30.54, 56], shows the upstream BER measured at the OLT in burst-mode operation on 2 µs bursts. The authors varied the ODN loss from 15 to 35 dB only for ONU 2 to simulate the dynamic range (DR) of the burst powers reaching the linear BMRx in a nonpower leveled LR-PON system. The burst power of ONU 1 was maintained constant and close to the linear BMRx overload power (loud ONU) with an ODN loss of 16 and 15 dB, respectively for the dense area *lollipop* and the amplified chain networks. The loud burst from ONU 1 therefore acted as a worst-case interferer for ONU 2 in terms of the linear BMRx operation. The results reported show for both cases that the post-FEC BER was always below 1×10^{-12} for ONU 1 and hence it is not reported in Fig. 30.23b. The ONU 2 BER shows that the dense area *lollipop* design supported ODN losses of up to 34 dB, corresponding to

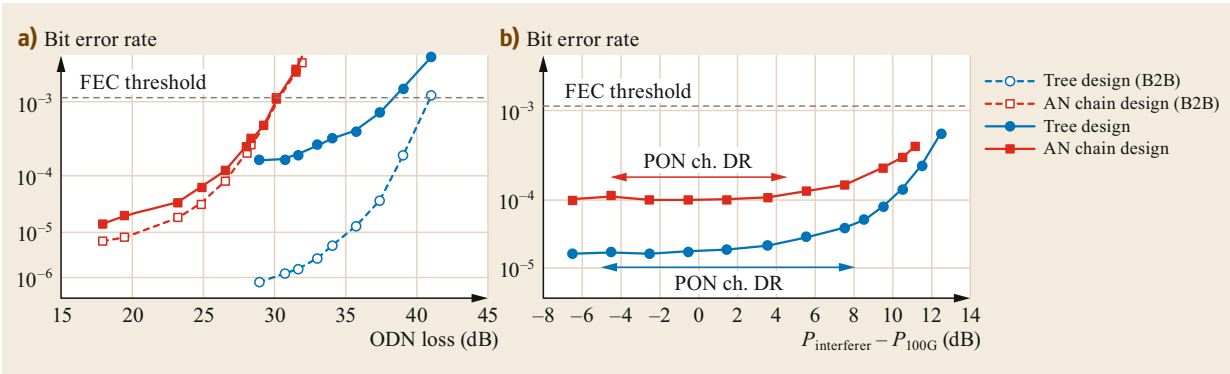


Fig. 30.24a,b Performance of the 100G link in downstream as a function of the ODN loss **(a)**; performance of the 100G link in upstream as a function of the power of the neighboring NRZ PON channels **(b)**

a dynamic range of 18 dB, a 256 split plus 20 km of fiber with no system margin and an overall split ratio of 1024 (256×4). For the AN chain design ONU 2 operated post-FEC with BER below 1×10^{-12} for ODN losses of up to 30 dB, corresponding to a dynamic range of 15 dB, an overall split ratio of 256 (64×4) per AN with margin and a total network split of 1024 (256×4). The measured FEC performance closely followed the theoretical threshold of 1.1×10^{-3} for a post-FEC error rate of 1×10^{-12} in both the downstream and upstream links.

A fronthaul channel terminated in the AN was also demonstrated in [30.54] for the dense area *lollipop* case and it was emulated by a 10 Gb/s continuous mode link driven by FPGAs located in the AN using one of the ODN output ports. Architecturally, the backhauling part of the wireless signal can be carried over one of the PON channels and terminated in the AN using an ONU connected to a 1% power tap on one of the ODN 4×4 splitter ports. The fronthaul channel operated error free without FEC (BER $< 1 \times 10^{-12}$) in both directions for an ODN loss of up to 28 dB (Fig. 30.23). While the downstream is effectively showing the same performance as the downstream PON channel, in the upstream direction the fronthaul channel shows better performance than the PON upstream channel because it is not transmitted through the link towards the MC-node and hence maintains a higher OSNR. It should also be noted that the fronthaul upstream channel is a continuous mode link and hence does not require the use of a burst-mode receiver.

Standard commercial solutions for high-capacity 100G links use coherent modulation formats such as dual-polarization quadrature-phase-shift-keying signal (DP-QPSK), which are potentially susceptible to linear and nonlinear impairments caused by interference from copropagating OOK channels. The performance of a 100G Pt-Pt downstream link, realized using a commercial transponder to generate and receive a DP-QPSK signal, was reported in [30.54, 56]. It was characterized in the presence of emulated traffic for a system fully loaded with 10G OOK PON channels and in particular with two interfering channels 50 GHz on each side of the 100G channel. The results of the 100G downstream link from [30.54, 56] are shown in Fig. 30.24a with the pre-FEC BER plotted as a function of the ODN loss using an EDFA preamplifier at the receiver with 5.5 dB noise figure (NF). The dense area *lollipop* and chain designs are both shown to be capable of supporting the 100G channel with ODN loss higher than 36 and 30 dB respectively (equivalent overall to 1024 network users). Figure 30.24a also shows the BER measured as a function of the ODN loss in a back-to-back (B2B) configuration, where the 100G transmitter is connected directly to the EDFA preamplifier at the receiver and the ODN loss is emulated by a variable optical attenuator at the input of the EDFA preamplifier. Assuming a negligible impact from other sources of impairments, such as for example chromatic dispersion, the effect of the nonlinearities in the ODN and backhaul fibers can be clearly seen for both designs by comparing the B2B curve with the one obtained with the complete network. In the upstream direction the OOK PON traffic, which is interfering with the 100G channel, presents bursts with different power. In [30.54, 56] the 100G channel was hence characterized in terms of the penalty caused by the nonlinear crosstalk from the two neighboring 10 Gb/s NRZ (non-return-to-zero) PON channels bursting with high power. The BER of the 100G upstream link was measured as a function of the power of two 50 GHz-spaced interfering channels operated with 2 µs bursts and 2 µs gaps overlapped in time to provide also a worst case for the nonlinearity. The experimental results in Fig. 30.24b show that the 100G channel can work below FEC threshold even when the neighboring bursty channels present dynamic ranges larger than the 13 dB (dense area *lollipop* design) and 9 dB (AN chain design) caused by the nonuniform ODN loss.

30.6 Summary

In this chapter, we described the history and evolution of LR-PON which could enable major restructuring of communications networks to massively increase capacity to end users while enabling the networks to be affordable and economically viable and also massively reducing energy use by eliminating large quantities of electronics processing and minimizing the number of disparate transmission systems. The flexibility of the topology and architecture of the LR-PON enables it to provide services to all geotypes from dense city to sparse rural and we have described two alternative LR-PON designs that illustrate this flexibility for this range of customer densities. The chapter has also described experimental results showing that these network designs are fully capable of implementation in real network environments today and could pave the wave for the network restructuring that will be necessary as bandwidth demands continue to grow and current networks struggle to remain financially viable and energy efficient.

References

30.1 D.B. Payne, M. Ruffini: *Local Loop Unbundling (LLU) regulation: Is it a barrier to FTTH deployment?* DISCUS White Paper (Univ. Dublin, Dublin 2005), available from: http://img.lightreading.com/downloads/Local-Loop-Unbundling-regulation-is-it-a-barrier-to-FTTH-deployment.pdf

30.2 A.M. Hill, R. Wyatt, J.F. Massicott, K. Blyth, D.S. Forrester, R.A. Lobbett, P.J. Smith, D.B. Payne: 39.5 million-way WDM broadcast network employing two stages of erbium-doped fibre amplifiers, Electron. Lett. **26**(22), 1882–1884 (1990)

30.3 D.S. Forrester, A.M. Hill, R.A. Lobbett, R. Wyatt, S.F. Carter: 39.81 Gbit/s 43.8 million-way WDM broadcast network with 527 km range, Electron. Lett. **27**(22), 2051–2052 (1991)

30.4 M.O. van Deventer, Y.M. van Dam, P.J.M. Peters, F. Vermaerke, A.J. Phillips: Evolution phases to an ultra-broadband access network: results from ACTS-PLANET, IEEE Commun. Mag. **35**(12), 72–77 (1997)

30.5 I. Van de Voorde, C.M. Martin, J. Vandewege, X.Z. Qiu: The superPON demonstrator: an exploration of possible evolution paths for optical access networks, IEEE Commun. Mag. **38**(2), 74–82 (2000)

30.6 N.P. Caponio, A.M. Hill, F. Neri, R. Sabella: Single-layer optical platform based on WDM/TDM multiple access for large-scale 'switchless' networks, Eur. Trans. Telecommun. **11**(1), 73–82 (2000)

30.7 D.B. Payne, R.P. Davey: The future of fibre access systems?, BT Technol. J. **20**(4), 104–114 (2002)

30.8 D.B. Payne, J.R. Stern: Wavelength switched, passively coupled, single mode optical networks. In: *Proc. ECOC, Venice* (1985)

30.9 G. Talli, P.D. Townsend: Hybrid DWDM–TDM Long-Reach PON for Next-Generation Optical Access, J. Lightwave Technol. **24**(7), 2827 (2006)

30.10 P. Ossieur, C. Antony, A. Clarke, A. Naughton, H. Krimmel, Y. Chang, C. Ford, A. Borghesani, D. Moodie, A. Poustie, R. Wyatt, B. Harmon, I. Lealman, G. Maxwell, D. Rogers, D. Smith, D. Nesset, R. Davey, P. Townsend: A 135-km 8192-Split Carrier Distributed DWDM-TDMA PON With 2 32 10 Gb/s Capacity, J. Lightwave Technol. **29**(4), 463–474 (2011)

30.11 P. Ossieur, C. Antony, A. Naughton, A. Clarke, H. Krimmel, X. Yin, X. Qiu, C. Ford, A. Borghesani, D. Moodie, A. Poustie, R. Wyatt, B. Harmon, I. Lealman, G. Maxwell, D. Rogers, D. Smith, S. Smolorz, H. Rohde, D. Nesset, R. Davey, P. Townsend: Demonstration of a 32 x 512 split, 100 km reach, 2 x 32 x 10 Gb/s hybrid DWDM-TDMA PON using tuneable external cavity lasers in the ONUs, J. Lightwave Technol. **29**(24), 3705–3718 (2011)

30.12 T.H. Wood: Multiple quantum well (MQW) waveguide modulators, J. Lightwave Technol. **6**(6), 743–757 (1988)

30.13 G. Maier, M. Martinelli, A. Pattavina, E. Salvadori: Design and cost performance of the multistage WDM-PON access networks, J. Lightwave Technol. **18**(2), 125 (2000)

30.14 H. Rohde, E. Gottwald, P. Alves, C. Oliveira, I. Dedic, T. Drenski: Digital multi-wavelength generation and real time video transmission in a coherent ultra dense WDM PON. In: *Proc. Opt. Fiber Commun.* (2013), Paper OM3H.3

30.15 J.B. Jensen, R. Rodes, A. Caballero, N. Cheng, D. Zibar, I.T. Monroy: VCSEL based coherent PONs, J. Lightwave Technol. **32**(8), 1423 (2014)

30.16 C. Raack, R. Wessally, D. Payne, M. Ruffini: Hierarchical versus flat optical metro/core networks: a systematic cost and migration study. In: *Proc. ONDM* (2016), https://doi.org/10.1109/ONDM.2016.7494075

30.17 S. McGettrick, F. Slyne, N. Kitsuwan, D.B. Payne, M. Ruffini: Experimental end-to-end demonstration of shared N:1 dual homed protection in long reach PON and SDN-controlled core. In: *Proc. Opt. Fiber Commun* (2015), Paper Tu2E.5

30.18 EU FP7 project DISCUS, http://cordis.europa.eu/project/rcn/106294_en.html (2012 to 2015)

30.19 M. Ruffini, D. Mehta, B. O'Sullivan, L. Quesada, L. Doyle, D.B. Payne: Deployment strategies for protected long-reach PON, J. Opt. Commun. Netw. **4**(2), 118–129 (2012)

30.20 U. Doetsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, P. Sehier: Quantitative analysis of split base station processing and determination of advantageous ar-

chitectures for LTE, Bell Labs Tech. J. **18**(1), 105–128 (2013)
30.21 NGMN Alliance: NGMN overview on 5G RAN functional decomposition, White Paper, https://www.ngmn.org/publications/ngmn-overview-on-5g-ran-functional-decomposition.html (2018)
30.22 S. Kuwano, J. Terada, T. Kawamura, N. Tanaka, S. Shigematsu, N. Yoshimoto: A novel DBA scheme for TDM-PON based mobile fronthaul. In: *Proc. Opt. Fiber Commun.* (2014), Paper Tu3F.3
30.23 P. Rost, C.J. Bernardos, A. De Domenico, M. Di Girolamo, M. Lalam, A. Maeder, D. Sabella, D. Wübben: Cloud technologies for flexible 5G radio access networks, IEEE Commun. Mag. **52**(5), 68–76 (2014)
30.24 NGMN Alliance: Suggestions on potential solutions to C-RAN, version 4.0, https://www.ngmn.org/publications/suggestions-on-potential-solutions-to-c-ran.html (2013)
30.25 M. Ruffini, F. Slyne, C. Bluemm, N. Kitsuwan, S. McGettrick: Software Defined Networking for next generation converged metro-access networks, Opt. Fiber Technol. (2015), https://doi.org/10.1016/j.yofte.2015.08.008
30.26 V. Lopez, J.M. Gran Josa, V. Uceda, F. Slyne, M. Ruffini, R. Vilalta, A. Mayoral, R. Muñoz, R. Casellas, R. Martínez: End-to-end service orchestration from access to backbone, J. Commun. Netw. **9**(6), B137–B147 (2017)
30.27 B. Skubic, B. Chen, J. Chen, J. Ahmed, L. Wosinska: Improved scheme for estimating T-CONT bandwidth demand in status reporting DBA for NG-PON. In: *Proc. Asia Commun. Photonics Conf. Exhib., Shanghai, 2009* (Opt. Soc. America, Washington, DC 2009), Paper TuT2
30.28 H. Leligou, C. Linardakis, K. Kanonakis, J. Angelopoulos, T. Orphanoudakis: Efficient medium arbitration of FSAN-compliant GPONs, Int. J. Commun. Syst. **19**, 603–617 (2006)
30.29 D. Shea, A. Ellis, D. Payne, R. Davey, J. Mitchell: 10 Gb/s PON with 100 km reach and 1024 split. In: *Eur. Conf. Opt. Commun. (ECOC), Rimini, Italy* (2003) pp. 850–851
30.30 ITU-T Recommendation G.987.3: 10-Gigabit-capable passive optical networks (XG-PON): transmission convergence (TC) layer specification (2014)
30.31 B. Kantarci, H. Mouftah: Bandwidth distribution solutions for performance enhancement in long-reach passive optical networks, IEEE Commun. Surv. Tutor. **14**, 714–733 (2012)
30.32 B. Skubic, J. Chen, J. Ahmed, B. Chen, L. Wosinska, B. Mukherjee: Dynamic bandwidth allocation for Long-Reach PON: overcoming performance degradation, IEEE Commun. Mag. **48**(11), 100–108 (2010)
30.33 H. Song, B.-W. Kim, B. Mukherjee: Multi-thread polling: a dynamic bandwidth distribution scheme in long-reach PON, IEEE J. Sel. Areas Commun. **27**, 134–142 (2009)
30.34 P. Alvarez, A. Hill, N. Marchetti, D. Payne, M. Ruffini: Analysis of the maximum balanced load in long-reach PONs. In: *2016 Int. Conf. Opt. Netw. Des. Model. (ONDM), Cartagena, Spain* (IEEE, Piscataway 2016)
30.35 ITU-T Recommendation G.983.4: A broadband optical access system with increased service capability using dynamic bandwidth assignment (2001)
30.36 ITU-T Recommendation G.987.2: 10-Gigabit-capable passive optical networks (XG-PON): physical media dependent (PMD) layer specification (2010)
30.37 R. Bonk, W. Poehlmann, D. van Veen, J. Galaro, R. Farah, H. Schmuck, T. Pfeiffer: The underestimated challenges of burst-mode WDM transmission in TWDM-PON, Opt. Fiber Technol. **26**, 59–70 (2015)
30.38 A. Hill, S. McGettrick, P. Alvarez, G. Talli, M. Achouche, R. Bonk: Design and implementation of a dynamic bandwidth protocol for TWDM-PON, and FPGA-based LR-PON OLT and ONU prototypes, EU 7th Framework Project DISCUS, Deliverable Report D4.12 (2016) https://cordis.europa.eu/project/id/318137/reporting
30.39 N. Cheng, F. Effenberger: Automatic ONU wavelength control in TWDM PONs. In: *Proc. Opt. Fiber Commun. Conf. 2014* (Opt. Soc. America, Washington, DC 2014), Paper W1D.4
30.40 ITU-T Recommendation G.989.2 Amendment 1: 40-Gigabit-capable passive optical networks 2 (NG-PON2): physical media dependent (PMD) layer specification (2016)
30.41 F. Liu, C. Rasmussen, R. Pedersen: Experimental verification of a new model describing the influence of incomplete signal extinction ratio on the sensitivity degradation due to multiple interferometric crosstalk, IEEE Photonics Technol. Lett. **11**(1), 137–139 (1999)
30.42 ITU-T Recommendation G.989.3: 40-Gigabit-capable passive optical networks (NG-PON2): transmission convergence layer specification (2015)
30.43 D. Carey, P. Ramaswamy, G. Talli, C. Antony, B. Roycroft, B. Corbett, P.D. Townsend: Characterization of a low-cost, monolithically integrated, tunable 10G transmitter for wavelength agile PONs, IEEE J. Quantum Electron. **54**(6), 1–12 (2018)
30.44 C. Antony, G. Talli, P.D. Townsend, J. Bauwelinck, D.W. Smith, I. Lealman: High extinction switching of SOAs for in-band crosstalk reduction in PON, Electron. Lett. **44**(14), 872–873 (2008)
30.45 Y.A. Akulova, G.A. Fish, P.-C. Koh, C.L. Schow, P. Kozodoy, A.P. Dahl, S. Nakagawa, M.C. Larson, M.P. Mack, T.A. Strand, C.W. Coldren, E. Hegblom, S.K. Penniman, T. Wipiejewski, L.A. Coldren: Widely tuneable electroabsorption-modulated sampled-grating DBR laser transmitter, IEEE J. Sel. Top. Quantum Electron. **8**(6), 1349–1357 (2002)
30.46 H. Ishii, H. Tanobe, F. Kano, Y. Tohmori, Y. Kondo, Y. Yoshikuni: Quasicontinuous wavelength tuning in super-structure grating (SSG) DBR lasers, IEEE J. Quantum Electron. **32**(3), 433–440 (1996)
30.47 B. Roycroft, P. Lambkin, S. Riesner, B. Corbett, J.F. Donegan: Transition from perturbed to coupled-cavity behavior with asymmetric spectral

emission in ridge lasers emitting at 1.55 μm, IEEE Photonics Technol. Lett. **19**(2), 58–60 (2007)
30.48 F. Smyth, E. Connolly, B. Roycroft, B. Corbett, P. Lambkin, L.P. Barry: Fast wavelength switching lasers using two-section slotted Fabry-Pérot structures, IEEE Photonics Technol. Lett. **18**(20), 2105–2107 (2006)
30.49 K. Shi, F. Smyth, D. Reid, B. Roycroft, B. Corbett, F.H. Peters, L.P. Barry: Characterization of a tuneable three-section slotted Fabry–Perot laser for advanced modulation format optical transmission, Opt. Commun. **284**(6), 1616–1621 (2011)
30.50 A. Densmore, P.E. Jessop: A quantum-well waveguide photodetector for high-precision wavelength monitoring about 1.55 μm, IEEE Photonics Technol. Lett. **11**(12), 1653–1655 (1999)
30.51 S. Pachnicke, J. Zhu, M. Lawin, M.H. Eiselt, S. Mayne, B. Quemeneur, D. Sayles, H. Schwuchow, A. Wonfor, P. Marx, M. Fellhofer, P. Neuber, M. Dietrich, M.J. Wale, R.V. Penty, I.H. White, J.-P. Elbers: Tunable WDM-PON system with centralized wavelength control, J. Lightwave Technol. **34**(2), 812–818 (2016)
30.52 Y. Luo, H. Roberts, K. Grobe, M. Valvo, D. Nesset, K. Asaka, H. Rohde, J. Smith, J.S. Wey, F. Effenberger: Physical layer aspects of NG-PON2 standards—part 2: system design and technology feasibility, J. Opt. Commun. Netw. **8**(1), 43–52 (2016)
30.53 H. Schmuck, R. Bonk, W. Poehlmann, C. Haslach, W. Kuebart, D. Karnick, J. Meyer, D. Fritzsche, E. Weis, J. Becker, W. Freude, T. Pfeiffer: Demonstration of SOA-assisted open metro-access infrastructure for heterogeneous services. In: *39th Eur. Conf. Exhib. Opt. Commun. (ECOC 2013), London* (2013), paper We.4.F.2
30.54 G. Talli, F. Slyne, S. Porto, D. Carey, N. Brandonisio, A. Naughton, P. Ossieur, S. McGettrick, C. Bleumm, M. Ruffini, D. Payne, R. Bonk, T. Pfeiffer, N. Parsons, P. Townsend: SDN enabled dynamically reconfigurable high capacity optical access architecture for converged services, J. Lightwave Technol. **35**(3), 550–560 (2017)
30.55 A. Kaszubowska-Anandarajah, R. Oberland, E. Bravi, A. Surpin, O. Aharoni, U. Ghera, R. Giller, E. Connolly, E.K. MacHale, M. Todd, G. Talli, D. McDonald: EDFA transient suppression in optical burst switching systems. In: *14th Int. Conf. Transpar. Opt. Netw. (ICTON)* (2012), https://doi.org/10.1109/ICTON.2012.6254384
30.56 D. Carey, N. Brandonisio, S. Porto, A. Naughton, P. Ossieur, N. Parsons, G. Talli, P. Townsend: Dynamically reconfigurable TDM-DWDM PON ring architecture for efficient rural deployment. In: *Proc. ECOC* (2016), W.3.E.4
30.57 S. Porto, C. Antony, G. Talli, P. Ossieur, P.D. Townsend: Requirements for adaptive electronic dispersion compensation in burst-mode systems. In: *Proc. Opt. Fiber Commun.* (2013), OTh3B.5
30.58 ITU-T Recommendation G.984.7: Gigabit-capable passive optical networks (GPON): long reach (2010)
30.59 P. Ossieur, C. Melange, T. De Ridder, J. Bauwelinck, B. Baekelandt, X.-Z. Qiu, J. Vandewege: Burst-mode electronic equalization for 10 Gb/s passive optical networks, Photonics Technol. Lett. **20**, 1706–1708 (2008)
30.60 S. Porto, A. Cleitus, J. Jain, D. Kelly, D. Carey, G. Talli, P. Ossieur, P.D. Townsend: Demonstration of 10 Gbit/s burst-mode transmission using a linear burst-mode receiver and burst-mode electronic equalization [invited], J. Opt. Commun. Netw. **7**, A118–A125 (2015)
30.61 P. Ossieur, N.A. Quadir, S. Porto, C. Antony, W. Han, M. Rensing, P. O'Brien, P.D. Townsend: A 10 Gb/s linear burst-mode receiver in 0.25 μm siGe:C BiCMOS, J. Solid-State Circuits **48**, 381–390 (2013)
30.62 ONF: SDN architecture, TR-502, Issue 1 (2014)
30.63 R. Muñoz, R. Vilalta, R. Casellas, R. Martínez, F. Francois, M. Channegowda, A. Hammad, S. Peng, R. Nejabati, D. Simeonidou, N. Yoshikane, T. Tsuritani, V. López, A. Autenrieth: Transport network orchestration for end-to-end multi-layer provisioning across heterogeneous SDN/openflow and GMPLS/PCE control domains, J. Lightwave Technol. **33**, 1540–1548 (2015)
30.64 S. McGettrick, F. Slyne, N. Kitsuwan, D.B. Payne, M. Ruffini: Experimental end-to-end demonstration of shared N:M dual homed protection in SDN-controlled long reach PON and pan-European core, J. Lightwave Technol. **34**(18), 4205–4213 (2016)
30.65 A. Nag, D.B. Payne, M. Ruffini: N:1 protection design for minimising OLTs in resilient dual-homed long-reach passive optical network. In: *Opt. Fiber Commun. Conf. (OFC)* (2014), Paper Tu2F.7
30.66 S. McGettrick, D.B. Payne, M. Ruffini: Improving hardware protection switching in 10 Gb/s symmetric long reach PONs. In: *Opt. Fiber Commun. Conf. (OFC)* (2013), Paper OW3G.2
30.67 N. Brandonisio, S. Porto, D. Carey, P. Ossieur, G. Talli, N. Parsons, P. Townsend: Forward error correction analysis for 10 Gb/s burst-mode transmission in TDM-DWDM PONs. In: *2017 Opt. Fiber Commun. Conf Exhib. (OFC), Los Angeles, CA* (2017), Paper Th2A.28

David Payne

Electrical Engineering
Trinity College Dublin
Dublin, Ireland
david.b.payne@btinternet.com

David Payne joined BT Labs as a Group Leader in 1978 and was Head of Optical and Broadband Research from 1999–2007. He was co-inventor of TPON (the first PON) in 1987 and was awarded the BT Martlesham Medal for his work on optical networks in 2005. He was Professor of Optical Networks at Trinity College Dublin from 2010 to 2018.

Giuseppe Talli

Photonic Systems Group
Tyndall National Institute, University College Cork
Cork, Ireland
giuseppe.talli@tyndall.ie

Giuseppe Talli received a Laurea degree in Electronic engineering from the University of Padova, Italy, and a PhD in Electronic Engineering from the University of Essex, UK. He worked at University College Cork and Intune Networks, Dublin and is currently with the Photonic Systems Group at the Tyndall National Institute, University College Cork.

Marco Ruffini

Trinity College, CONNECT telecommunications research centre,
Computer Science and Statistics
The University of Dublin
Dublin, Ireland
marco.ruffini@tcd.ie

Marco Ruffini received his PhD from Trinity College Dublin (TCD), Ireland, in 2007. He is Associate Professor at TCD. He is a Principal Investigator (PI) with both the CONNECT Telecommunications Research Centre and the IPIC Photonics Integration Centre. He carries out research on 5G optical networks, fixed-mobile convergence, access-metro networks, and network virtualization.

Alan Hill

Trinity College, CONNECT Telecommunications Research Centre,
School of Computer Science and Statistics
University of Dublin
Dublin, Ireland
alanhill@timewave.f9.co.uk

Alan Hill received his PhD from London University in 1976. He worked at BT Research Laboratories in the UK for most of his career. He has more recently worked for Trinity College Dublin in Ireland, on dynamic bandwidth assignment and dynamic wavelength assignment issues in time and wavelength division multiplexed long-reach passive optical networks.

Paul Townsend

Photonics
Tyndall National Institute, University College Cork
Cork, Ireland
paul.townsend@tyndall.ie

Paul Townsend received his PhD in Physics from the University of Cambridge in 1987. He worked at Bellcore (USA), BT (UK) and Corning (UK) before moving to University College Cork in Ireland in 2003, where he is currently Professor of Photonic Systems Research, Head of Photonics at the Tyndall National Institute and Director of the IPIC SFI Photonics Research Centre.

31. Digital Optical Front-Haul Technologies and Architectures

Fabio Cavaliere, David Larrabeiti

This chapter analyzes the evolution of fronthaul interfaces and networks, based on current standardization activities. Then, since fronthaul networks require, as any transport network, multiplexing and switching technologies, different options are discussed, highlighting their pros and cons and application space: multiplexing at physical layer, both in space (relying on dark fibers) and wavelength domain; circuit multiplexing, with focus on the widespread ITU-T Optical Transport Network standard; and packet switching, where satisfying strict timing constraints and maintaining the benefits of statistical multiplexing at the same time poses new design trade-offs, which are also discussed. The final part of the chapter addresses the evolution of the fronthaul network as a multilayer network that can efficiently exploit all the above multiplexing technologies, compatibly with the split options of the radio protocol stack defined at 3GPP for 5G and the introduction of new packet fronthaul interfaces.

The term fronthaul was introduced to indicate the interface between the two units of a split radio base station architecture, namely remote radio head and baseband unit. Several factors contributed to extend the original definition of fronthaul from a mere link specification to a new network domain having its peculiar needs, especially in terms of timing and synchronization requirements: the need to improve the coordination between adjacent cells in dense radio deployment scenarios; the advent of 5G with its demanding transport capacity requirements due to high bandwidth on air and beamforming; the centralization and virtualization of baseband processing functions in fewer nodes, to save operational costs; and the convergence of different services in the same fiber infrastructure.

31.1 Fronthaul Networks: Definitions and Terminology

In the RoF (radio-over-fiber) paradigm [31.1, 2], the radio signals coming from the radio equipment (RE) of the base stations (BS) in a mobile network are transported over optical fibers to a centralized location where a radio equipment controller (REC) is deployed, instead of being locally processed at the BS. This way, the base stations (now re-named as remote radio heads, RRH) become simpler, and the more complex elements processing the baseband signals (called baseband units, BBU) can be centralized. The purpose of this physical

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_31

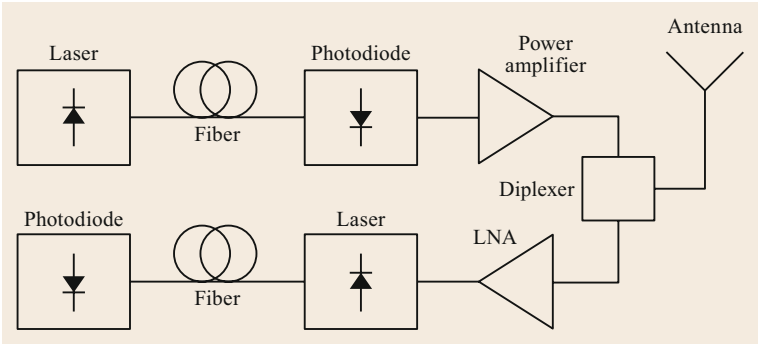


Fig. 31.1 Scheme of a basic A-RoF system

separation of BS functions is to reduce the costs in the outside plant, as RRHs are expected to have smaller installation footprint, energy consumption, and cooling requirements than a complete BS. In RoF, the interface between the RE and the REC is called fronthaul. It can convey the radio signal in analog format (A-RoF) or, more frequently, in digital format (D-RoF). In principle, an A-RoF scheme (Fig. 31.1) is appealing due to its simple and cost effective hardware configuration: a diplexer connected to the antenna separates received and transmitted wireless signals; the received signal is sent to a low-noise amplifier (LNA), modulated by a laser, transmitted into the optical fiber, and reconverted in the electrical domain by a photodiode; the signal to be transmitted on air is modulated by a laser, launched into the fiber, photo-detected and sent to the antenna through a power amplifier. However, the necessity to propagate an analog waveform undistorted through all the transmission chain requires the use of linear devices and penalizes the performance of A-RoF compared to D-RoF. Moreover, D-RoF can rely on standardized integrated circuits libraries and interfaces, making it the most adopted option in today's fronthaul networks. In this chapter, we will only discuss D-RoF.

The centralization of radio signal processing in a single site serving several RRHs enables the deployment of advanced interference reduction techniques, such as cooperative multipoint, enhanced forward error correction (FEC) techniques and faster radio processor upgrades. This scenario is referred to as centralized RAN (radio access network).

The cloud infrastructure radio access network (C-RAN) vision [31.3] goes one step beyond the mere concentration of baseband processing equipment in the same site, by advocating for the virtualization of shared BBU pools (sometimes called baseband hotel) in a computing cluster (the cloud) in such a way that BBUs can be shared and launched on demand according to the real traffic workload in the different RRHs. This way, antenna elements in cells with low traffic can be switched off and do not use BBU or fronthaul network resources.

The fronthaul network can be as simple as a pair of optical fibers or as complex as a time-sensitive

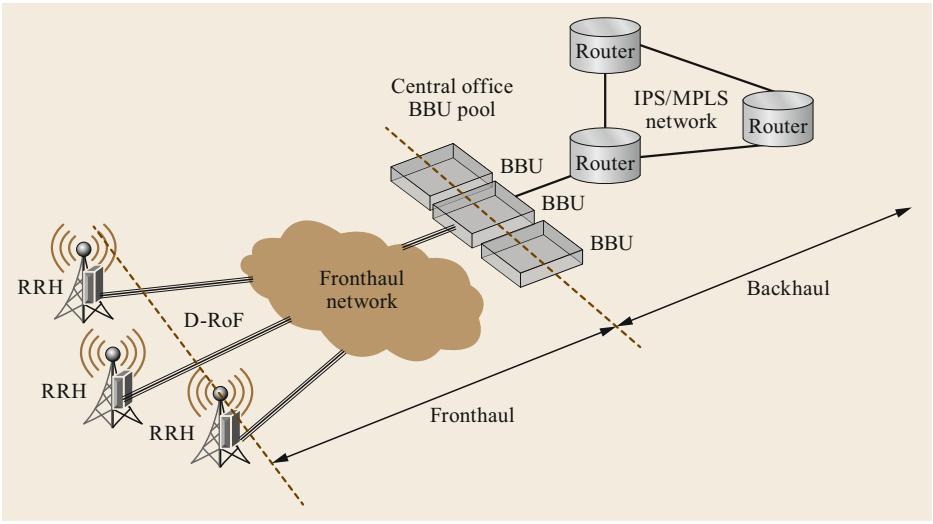


Fig. 31.2 Concept of the fronthaul network

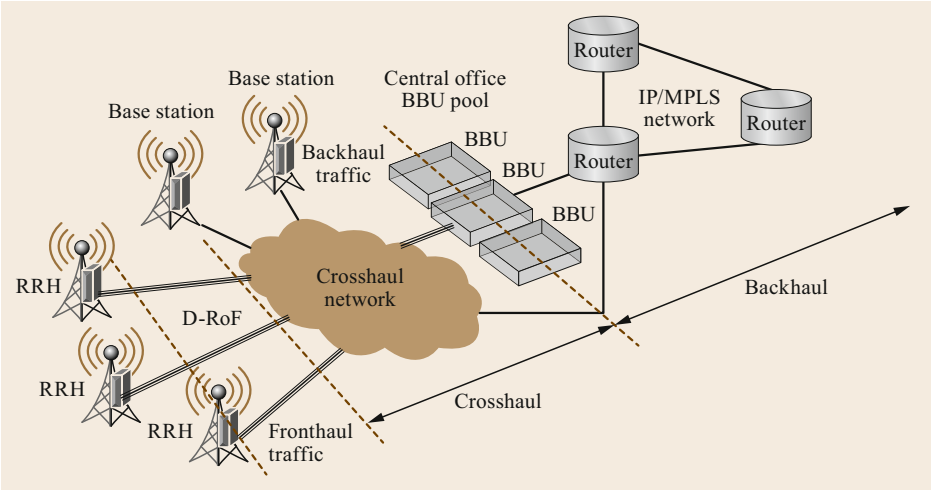


Fig. 31.3 Concept of the crosshaul network

packet-switched network over a dense wavelength division multiplexing (DWDM) optical transport network, as described later in this chapter. Figure 31.2 shows a fronthaul network interconnecting, through D-RoF links, several RRHs to a pool of BBUs, located in the same central office. The BBU pool is connected to a packet switched backhaul network, for example, based on IP/MPLS (internet protocol/multiprotocol label switching).

Since both RRHs and regular base stations must cohabitate and require a transport network, telecommunications operators and system vendors are developing the concept of a crosshaul (or Xhaul) network [31.4], as a network capable of transporting both fronthaul traffic and backhaul traffic coming from or going to regular base stations. Figure 31.3 shows an example of a crosshaul network. Compared to the fronthaul network of Fig. 31.2, this network is able to manage backhaul traffic from base stations, bypassing the BBU pool, concurrently to fronthaul traffic generated by the RRHs.

The MEF (Metro Ethernet Forum) also introduced the term midhaul to refer to the network used to interconnect base station sites, especially small cell base stations, and its master macrocell base station. However, the current prevalent use of the term midhaul by 3GPP is to refer to the segment between fronthaul and backhaul that transports radio traffic at a high-layer split (above the physical layer). Midhaul traffic has transport requirements similar to backhaul traffic.

31.2 Digital Fronthaul Standards

Base stations manufacturers have defined two main specifications for the digital transport of fronthaul data: the common public radio interface (CPRI) [31.5, 6] and the open base station architecture initiative (OBSAI) [31.7]:

- The CPRI specification was initially intended to allow the replacement of the local radio frequency (RF) coaxial cable carrying the analog radio signal between the radio equipment and the baseband unit, by an optical fiber carrying a digitized radio signal (user data), and control and management and synchronization signals (Fig. 31.4). Thus, the focus of CPRI is exclusively set on the interface between the RF stage (radio equipment in CPRI terminology) and the baseband unit (radio equipment controller). Furthermore, it only covers the physical and the link layers. In order to address most open issues, the ORI (open radio equipment interface) ETSI (European Telecommunications Standards Institute) Industry Specification Group (ISG) was chartered in 2010 to define an interface on top of the CPRI specification to facilitate interoperability across vendors, including higher layer functions and other IQ formats.
- The goal of OBSAI is to define a complete base station reference architecture to enable an open market of interoperable base station components. The OBSAI reference architecture defines four functional blocks (RF, baseband, control-and-clock, and transport), interfaces between them, and requirements for external interfaces. Unlike CPRI, the RF-baseband interface of OBSAI (reference point 3) defines the full protocol stack: physical, link, trans-

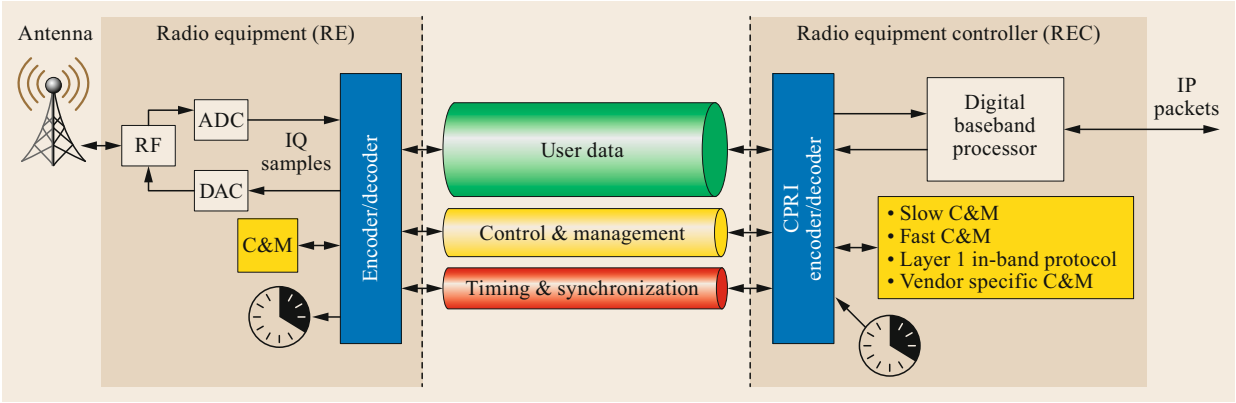


Fig. 31.4 Scheme of a D-RoF system, interface RE–REC in CPRI

Table 31.1 CPRI bit rate options

CPRI option	Bit rate (Mbit/s)	Line coding	Notes
1	614.4	8B/10B	1 × 491.52 × 10/8 Mbit/s
2	1228.8	8B/10B	2 × 491.52 × 10/8 Mbit/s
3	2457.6	8B/10B	4 × 491.52 × 10/8 Mbit/s
4	3072.0	8B/10B	5 × 491.52 × 10/8 Mbit/s
5	4915.2	8B/10B	8 × 491.52 × 10/8 Mbit/s
6	6144.0	8B/10B	10 × 491.52 × 10/8 Mbit/s
7	9830.4	8B/10B	16 × 491.52 × 10/8 Mbit/s
8	10 137.6	64B/66B	20 × 491.52 × 66/64 Mbit/s
9	12 165.12	64B/66B	24 × 491.52 × 66/64 Mbit/s
10	24 330.24	64B/66B	48 × 491.52 × 66/64 Mbit/s

port, and application. However, OBSAI has a fixed IQ sample envelope size, whereas CPRI has a programmable IQ sample size. This leads to a more efficient mapping for transmission which, together with market factors, may have led to a wider adoption of CPRI by BS vendors. Therefore, in the rest of the chapter we shall refer mainly to CPRI. Table 31.1 summarizes the bit rate options of CPRI.

Moving the baseband processing chain tasks far from the RE exhibits potential performance issues. For example, the latency between the RE and REC—CPRI terms for RRH and BBU respectively—has to be low enough in order to not negatively affect the retransmission mechanisms of the air interface. Moreover, in order to not affect the precision of the phase transfer, the imbalance between uplink and downlink delay should be very small. Moreover, the noise contributions introduced by the optical link (for example, the jitter) should be low enough to continue to guarantee the target bit error rate (BER). CPRI does not specify a complete set of performance requirements, which may also depend on the implementation of the system manufacturer. Moreover, the CPRI specification was conceived for RRH and BBU hosted in the same site and connected by short (from tens to hundreds of meters) links. Hence, it is not trivial to extend the specification to a network scenario like the one depicted in Fig. 31.2.

Table 31.2 reports an example of what the basic set of requirements for CPRI transport over an optical net-

Table 31.2 CPRI requirements

Fronthaul requirement	Value	Notes
Latency (RTT, round trip time)	≈ 200 µs	This parameter is not specified by CPRI for an optical network. Hence, the value of the most time sensitive class of service in the eCPRI transport specification [31.8], which has similar latency constraints, has been taken as reference.
Latency accuracy	≈ ±8.138 ns	In the CPRI specification it is the link delay accuracy in downlink between master and slave service access points, excluding the cable length. In optical networks, it is often interpreted as the maximum allowed delay imbalance among different time aligned RRHs connected to the same BBU. If the delay is estimated by means of a round trip measurement, it is also the allowed imbalance between downlink and uplink delay. However, no standard value has been agreed yet.
Frequency jitter	±2 ppb	It is the frequency jitter specified by CPRI for point to point link. As discussed later, additional jitter sources may exist in optical networks, which make difficult to respect the CPRI specification.
BER	10^{-12}	It is the bit error rate value usually taken as a reference by the ITU-T standards for the optical interfaces specification.

work could be, but no agreed set of parameters has been standardized yet.

Different multiplexing technologies can be used to aggregate and transport over optical network fronthaul signals generated by RRHs hosted at different termination nodes of the network. The following sections will discuss multiplexing technologies at physical, circuit, and packet layers.

31.3 Physical-Layer Multiplexed Optical Fronthaul

Multiplexing several fronthaul links at the physical layer allows us to minimize the latency introduced by the transport network, since no processing is required except basic functionalities like media conversion, forward error correction, and overhead for operation, administration and maintenance (OAM). Physical layer multiplexing is possible in the space domain using a dedicated optical fiber (dark fiber) for each fronthaul link or allocating different wavelengths in the same fiber. The latter method can adopt the coarse or dense wavelength division multiplexing standards (CWDM and DWDM, respectively) specified by the ITU-T.

31.3.1 Fronthaul Installations over Dark Fibers

In areas with plenty of installed and unused fibers (dark fibers), the BBUs can be directly connected to the RRHs by means of dedicated fiber links. In a C-RAN, individual dark fibers will connect all RRHs to the same baseband hotel, with a hub-and-spoke fiber topology (Fig. 31.5).

The IEEE 802.3 Ethernet Working Group [31.9] developed a number of standards for transceivers used in local area networks (LANs), whose technology also apply to point-to-point fronthaul connections between radio and baseband units. Gigabit Ethernet was introduced in 1998 with Amendment IEEE 802.3z, followed in 2000 by Amendment IEEE 802.3ad, adding the capability to aggregate multiple full-duplex point-to-point links into a single logical link. In 2002, Amendment IEEE 802.3ae added operation at 10 Gb/s (10 gigabit Ethernet), followed in 2010 by Amendment IEEE 802.3ba (40 gigabit Ethernet and 100 gigabit Ethernet).

Figure 31.6 shows bit rates and distance for various optical interfaces over single-mode fibers (SMF) supported by IEEE 802.3. Other speed values, for example 25 Gbit/s, are under study at the time of writing this book.

There are several aspects that make appealing the deployment of fronthaul links over dark fiber: the transceivers are based on commodity technology, developed for large production volumes and low unit cost for LANs and fiber-to-the home (FTTH) applications; pluggable optical interfaces, such as small form-factor pluggable (SFP) modules, allow to gradually upgrade the system capacity, according to a pay-as-you-grow policy; transceivers working over the so-called extended temperature range (−40 to +85 °C) are available for outdoor operation in rough environments.

31.3.2 Fronthaul over CWDM

Using dedicated dark fibers is surely the simplest way to connect RRHs to the baseband hotel. However, the number of optical fiber cables is not always sufficient to connect all the RRH sites, which may be numerous and scattered over a wide area. In other cases, the cable

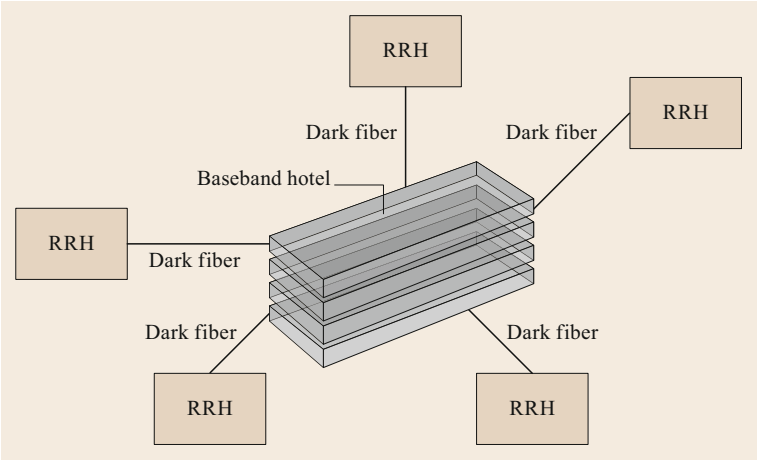


Fig. 31.5 CRAN based on dark fibers

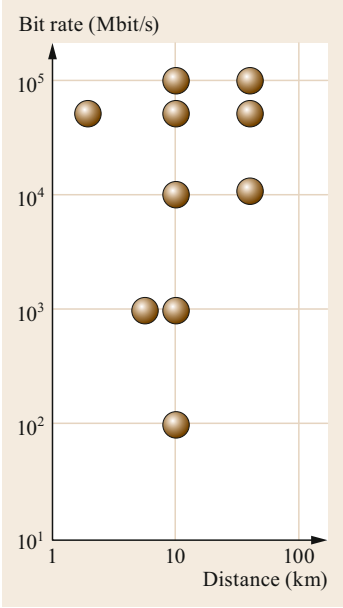


Fig. 31.6 Bit rates and distance of IEEE 802.3 optical interfaces

needs to be leased at a cost that may be not affordable, especially for small network operators (an analysis is reported in [31.10]). In these situations, more RRHs share the same optical fiber towards the baseband processing site by means of multiplexing techniques.

Coarse wavelength division multiplexing (CWDM) is an optical technology suitable for networks with low or moderate aggregate capacity, since it can realize cost-effective WDM systems through a combination of uncooled lasers and wide pass-band filters. There are 18 CWDM channels, from 1270 to 1610 nm, separated by 20 nm, defined by ITU-T [31.11]. For example, using CWDM transceivers working with CPRI Option 9 (12.16512 Gbit/s), the total capacity would be 219 Gbit/s. CWDM allows a more efficient use of the deployed fiber cables. Figure 31.7 shows the same C-RAN of Fig. 31.5. With CWDM, only a pair of fibers (one downstream, the other upstream), running through all the RRH sites, is now necessary.

This is possible due to the use of optical add drop multiplexers (OADM). The working principle of an OADM is illustrated in Fig. 31.8 for the 1290 nm channel (just one direction is shown for the sake of simplicity).

A 3-port optical filter receives the CWDM channel's comb at port 1, drops the selected wavelength at port 2 (which has a pass-band filtering response) and lets the other ones pass at port 3 (which has a stopband filtering response). A second filter, equal to the previous one, performs the add operation, as illustrated in Fig. 31.8. CWDM transceivers, usually SFPs, are plugged in the RRH and connected to the OADM. The modules commercially available guarantee outdoor operation, at least for bit rates up to 10 Gbit/s.

It has already been mentioned that a widespread fronthaul protocol like CPRI requires almost equal propagation delays in upstream and downstream. The maximum tolerated delay difference is in the order of 10 ns, approximately corresponding to 2 m of fiber.

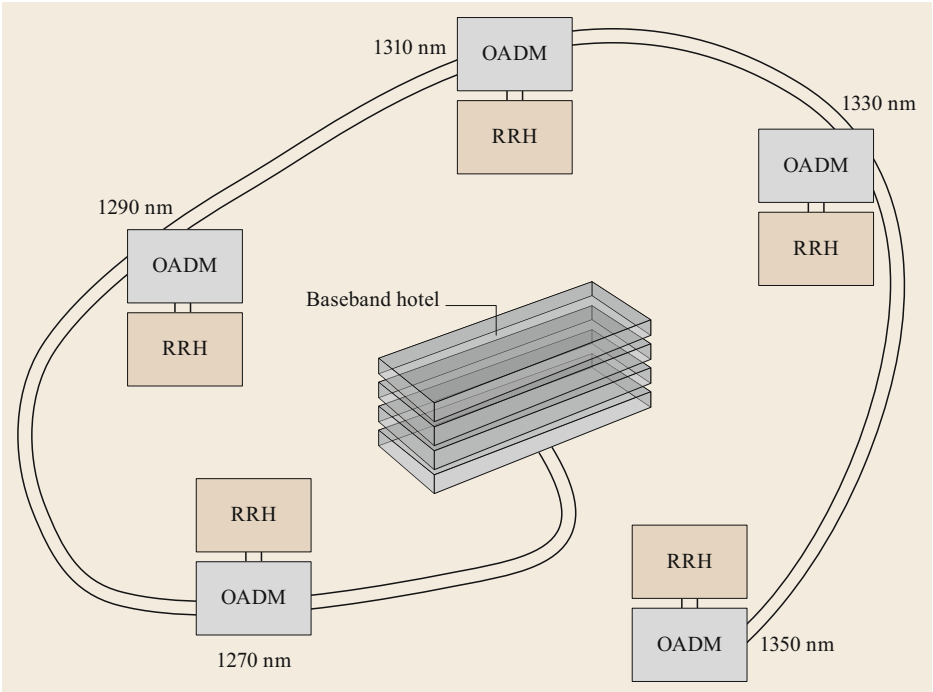


Fig. 31.7 C-RAN based on CWDM

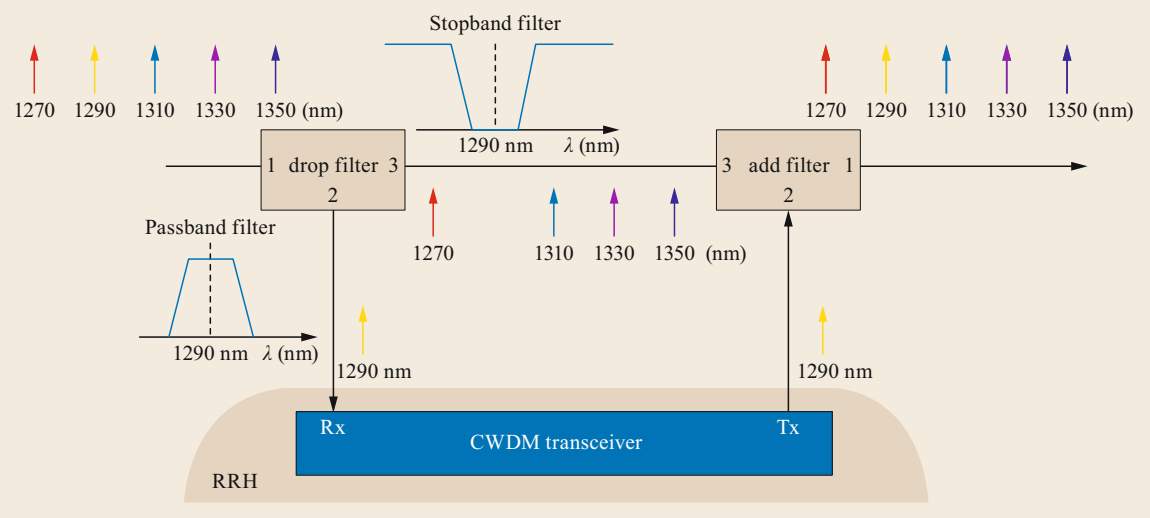


Fig. 31.8 OADM working principle

Hence, in transmission systems that use two fibers, like CWDM, it is required either to compensate for any difference of cable length in the two directions (for example, by using patch cords) or to use buffers after the signal has been photodetected. Due to the wide channel spacing, in CWDM systems it is also recommended to use the same wavelength in the downstream and upstream directions to avoid any delay difference induced by fiber chromatic dispersion. To estimate the delay difference, it is possible to use the upper bound in [31.12] for the fiber chromatic dispersion coefficient (31.1)

$$D_{\max}(\lambda) = \frac{\lambda S_{0\max}}{4}\left[1-\left(\frac{\lambda_{0\min}}{\lambda}\right)^4\right], \qquad (31.1)$$

where $D_{\max}$ is the chromatic dispersion coefficient in (ps/nm)/km, λ is the wavelength in nm, $\lambda_{0\min}$ is the minimum value of the zero dispersion wavelength (1300 nm for the standard single-mode fiber specified in [31.12]), and $S_{0\max}$ is the maximum slope at λ_0 (0.092 (ps/nm^2) km for the standard single-mode fiber specified in [31.12]).

The maximum delay coefficient in ps/km can be calculated by integrating (31.1),

$$\text{Delay}(\lambda)_{\max} = A + \frac{\lambda^2 S_{0\max}}{8}\left[1+\left(\frac{\lambda_{0\min}}{\lambda}\right)^4\right]. \qquad (31.2)$$

A is an undetermined integration constant that disappears in the calculation of the delay difference ΔT between two wavelengths λ_1 and λ_2 calculated as

$$\Delta T = |\text{Delay}(\lambda_1)_{\max} - \text{Delay}(\lambda_2)_{\max}|\, L\,, \qquad (31.3)$$

where L is the fiber length in km.

For example, the delay difference between two contiguous CWDM channels at 1590 and 1610 nm is 8.3 ns for 20 km of fiber, which is very close to the limit value tolerated by CPRI.

A practical issue of CWDM systems is that it is necessary to keep the inventory of the wavelengths installed at any site of the network in order to make it possible to replace the transceivers in the case of failure and avoid that equal wavelengths are used at different sites, causing traffic disruption. For the first reason, spare parts should be made available at any site for any installed wavelength. These are all issues that increase the operational costs of the network.

31.3.3 Fronthaul over DWDM

Today, DWDM systems are not widely used in fronthaul networks due to higher cost compared to dark fiber or CWDM systems. However, the next generation mobile network, known as 5G, is required to support 10–100 times higher user data rates and 1000 times higher mobile data volume per geographical area than the current mobile network. Even if the introduction of more bandwidth-efficient fronthaul interfaces than CPRI is under study, the increase of capacity is still expected to be significant, making DWDM a suitable candidate for 5G fronthaul.

The maximum capacity and distance of DWDM systems are far beyond those needed for fronthaul applications, due to the narrow channel frequency spacing (100 or 50 GHz) and the availability of erbium-doped fiber amplifiers in the 1550 nm wavelength region, which is the operating range of DWDM systems. Commercial DWDM systems transmit 96 optical channels,

Table 31.3 Link dimensioning of a 10 Gbit/s DWDM centralized network

	Optical link parameters	Value
A	Available link attenuation	22 dB
B	Link distance	20 km
C	Fiber attenuation coefficient	0.25 dB/km
D	Fiber attenuation ($= B \times C$)	5 dB
E	OADM add/drop loss	2 dB
F	OADM pass-through loss	2 dB
G	Add/drop loss at the baseband hotel	5 dB
H	Number of OADM nodes, where $A = D + E + (H-1) \times F + G$	6

modulated at 100 Gbit/s, over thousands of kilometers. However, the availability of many types of compact and pluggable DWDM transceivers makes possible to trade off cost with distance and capacity performance. For example, using cost-effective SFPs it is possible to transmit over 80 km of fiber at a maximum bit rate of 11.3 Gbit/s, supporting attenuation budgets higher than 22 dB. These distance and attenuation values are suitable for centralized scenarios like the one shown in Fig. 31.7 for CWDM. An example of link dimensioning using DWDM is reported in Table 31.3.

As it has already been mentioned, due to the concurrent increase of user data rate and number of users, 5G fronthaul will require bit rate values higher than 10 Gbit/s. Using 25, 50, or 100 Gbit/s would allow to exploit technologies already developed for IEEE 802.3 Ethernet interfaces. However, current 100 Gbit/s DWDM transceivers based on coherent dual polarization quadrature shift keying (DP-QPSK) are designed for long distances, in the order of 1000 km, and are too expensive for the fronthaul segment. This is the reason why industry, standardization bodies and the research community are studying cost-effective solutions working on shorter distances, typically a few tens of kilometers. Spectrally-efficient modulation formats suitable for direct detection, like pulse amplitude modulation-4 (PAM-4) or discrete multitone (DMT) are promising but have a limited link budget (less than 10 dB) and low tolerance to fiber chromatic dispersion. These issues may be solved in future by means of integrated photonic devices, like semiconductor optical amplifiers (SOA) and optical dispersion compensators based on microring resonators.

High capacity is not the only advantage of a DWDM design. Using different wavelengths between different transmitting and receiving nodes, DWDM makes it possible to design complex networks, for example having a mesh topology, by maintaining a point-to-point logical connectivity between each RRH and BBU node, without any intermediate electronic signal processing and switching, which would impair latency and energy efficiency.

31.4 Circuit Multiplexing of Fronthaul Links

Time division multiplexing (TDM) allows us to efficiently exploit the fiber bandwidth, allocating a certain number of low and constant bit rate (CBR) signals, like CPRI, in a single aggregate higher bit rate line signal, while ensuring deterministic and predictable delays, a key feature in fronthaul links.

Different TDM schemes can be implemented. One is proposed in the CPRI specification [31.5], making it possible, for example, to allocate two CPRI Option 1 tributary signals in one CPRI Option 2 frame.

CPRI transport over the optical transport network (OTN), described in the ITU-T supplement [31.13], is another multiplexing scheme. OTN is an optical transport standard developed by ITU-T [31.14], which specifies payload encapsulation, operation administration and maintenance (OAM) overhead, FEC, and a multiplexing hierarchy. According to [31.13] and using the OTN terminology in [31.14], CPRI is mapped into an optical channel payload unit-k (OPUk). Two CPRI options 1 and 2 signals are transported via OPU0 (whose bit rate is about 1.25 Gbit/s), CPRI option 3 is transported via OPU1 (with a bit rate of about 2.5 Gbit/s), and CPRI options 4–10 are transported via OPUflex (a flexible transport container, whose bit rate can be adapted to that of the client signal).

The generic mapping procedure is used for CPRI options 1–3, and the bit-synchronous mapping procedure is used for CPRI options 4–10. The generic mapping procedure (GMP) is a mechanism described in the OTN standard [31.14], used to accommodate nominal bit rate and clock difference between the client and OTN frame. According to the GMP, the OTN frame is divided into a certain number of GMP words, where each word may contain either data or stuffing bits. Using the generic mapping procedure [31.13], the relationship between the OTU2 (optical transport unit) bit rate and the CPRI client bit rate is given by

$$\text{CPRI\_BitRate} = k \times \text{OTU2\_BitRate} \times \left(\frac{79}{1377}\right) . \tag{31.4}$$

Table 31.4 Scaler k of CPRI options 1–6

CPRI Client	k
Option 1 (614.4 Mbps)	1
Option 2 (1228.8 Mbps)	2
Option 3 (2457.6 Mbps)	4
Option 4 (3072.0 Mbps)	5
Option 5 (4915.2 Mbps)	8
Option 6 (6144.0 Mbps)	10

The values of the scaler k for CPRI Options 1–6 are shown in Table 31.4.

When all the client signals are synchronized, i.e., use the same time reference, the OTN frame rate can be derived from the client rate, making the rate adaptation mechanism simpler. This is the case for the bit-synchronous mapping procedure [31.13]. It is applicable within a single administrative domain with a single clock reference, which means that all the RRHs present similar characteristics; for example, they are all provided by the same vendor. In the bit-synchronous mapping procedure, the native line coding of CPRI is sometimes replaced by a more bandwidth-efficient code, according to a procedure named transcoding. For example, CPRI Option 3 uses 8B10B as the line code, which maps 8 bits of real data into a 10-bit word. This would lead to a coding overhead equal to 20% in the absence of transcoding. A more efficient code is 64B66B, used in CPRI Options 8–10, which gives only 3% of overhead and does not need to be transcoded. After transcoding, the CPRI client signals are multiplexed into an optical payload unit by interleaving, i.e., each client CPRI frame is divided into time slots, and consecutive positions in the OTN payload are assigned to time slots belonging to different clients. Both the optical channel data unit and the optical transport unit (ODU2r (optical data unit), and OTU2r, respectively) are overclocked compared to the client signals to absorb small differences of the client signals clock from the nominal value. Nominal bit rate and tolerance values resulting from the bit-synchronous mapping procedure are summarized in Table 31.5. The tolerance values are usually given in parts per million (ppm), i.e., 10^6 times the ratio between deviation from the nominal value and nominal value itself. For very accurate timing references, a similar unit, parts per billion (ppb), is sometimes used.

CPRI signals require special consideration when carried as a CBR client signal over OTN, since the mapping procedure does not guarantee that the clock frequency deviation specified by CPRI, i.e., 2 ppb (Table 31.2), is respected. Simulation setups to check the compliance of the OTN mapping procedure with the timing requirements of CPRI are reported in [31.13]. The clock frequency deviation obtained ranges from 29 to 317 ppb, which are values that are much higher than

Table 31.5 CPRI over OTN bit rate and bit-rate tolerances

Signal type	Nominal bit rate (kbit/s)	Tolerance (ppm)
OTU2r	$255/238 \times 128 \times 24 \times 3840$	±100
OTU2r no FEC	$239/238 \times 128 \times 24 \times 3840$	±100
ODU2r	$239/238 \times 128 \times 24 \times 3840$	±100
OPU2r	$128 \times 24 \times 3840$	±100

The nominal OTU2r rate is approximately 12 639 085.714 kbit/s
The nominal OTU2r without FEC and ODU2r rates are approximately 11 846 045.042 kbit/s
The nominal OPU2r rate is 11 796 480 kbit/s

2 ppb. This means that to interwork with an OTN-based transport network, the RRHs would need a special design that is able to filter the frequency noise added by the OTN equipment.

Another issue of mapping CPRI over OTN is that any path of the OTN network should be designed so that uplink and downlink delays are equal within a tolerance of approximately 8 ns, according to the CPRI specification (Table 31.2). In real networks, unbalanced delay times can be caused by:

- Different cable lengths when two optical fibers are used in upstream and downstream (7 m approximatively corresponds to 34 ns of delay).
- Difference of propagation delay when downstream and upstream transmission wavelengths are different (for example, using 1.3 and 1.55 µm as transmission wavelengths causes about 33 ns of time difference over 20 km of standard single mode fiber).
- Difference of signal processing time (including functions such as time multiplexing, encapsulation, compression).

Any of the mentioned time differences needs to be compensated by using buffers, dimensioned based on the latency measured in the two directions.

The performance issues discussed above, together with the evolution from fronthaul to crosshaul networks (Sect. 31.1), are leading to the introduction of TDM schemes alternative to OTN and more suitable for client signals having strict timing requirements. Independently of the specific solution, any new framing scheme shall be synchronous to the most time sensitive client signal, i.e., the fronthaul one, so as not to degrade the accuracy of the clock frequency during the mapping procedure. A simple example of a framing mechanism supporting both fronthaul and backhaul and using the clock of the fronthaul signal (e.g., CPRI) as a reference, is reported in Fig. 31.9. The CPRI clock is used to de-

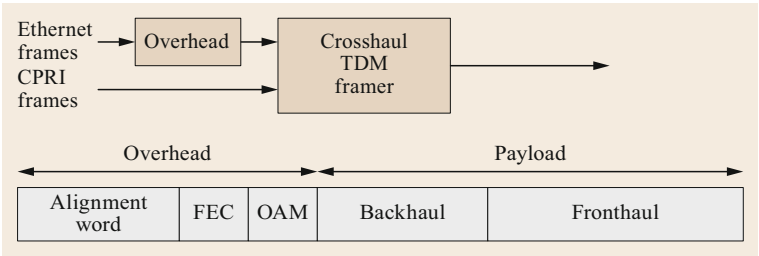


Fig. 31.9 CPRI and Ethernet over the same optical channel

rive the clock of the crosshaul frame, while the clock of other types of client (e.g., Ethernet frames used for backhaul links) is adapted by means of a buffer, which absorbs any difference between the nominal value and the accuracy of CPRI and Ethernet clocks. In the example of Fig. 31.9, two separate portions of the frame are allocated for fronthaul and backhaul but other multiplexing strategies, e.g., bit interleaving, could be used.

31.5 Fronthaul over Packet

Another alternative to transport fronthaul traffic is the use of a packet switching network. The purpose of this approach is to get a more cost-effective solution to multiplex a higher number of fronthaul signals than optical wavelength or TDM multiplexing and also to leverage statistical multiplexing gains. Like with DWDM or TDM, it is possible to merge backhaul and fronthaul traffic on the same links and use a single switching and multiplexing infrastructure (the aforementioned crosshaul concept), provided that the delays of the packets are kept under control. The distance range of reference for fronthaul traffic transport (< 20 km) allows the use of standard Ethernet transceivers without regeneration. Additional wavelength-division multiplexing is possible if there is a lack of deployed fibers.

31.5.1 Packet Fronthaul Technologies and Standards

The packet switching technologies suitable for fronthaul transport are the ones already deployed by telecom operators in metropolitan and aggregation networks, namely: IP; MPLS-TP (MPLS Transport Profile) pseudowires [31.15, 16]; Metro-Ethernet Mac-in-Mac (IEEE 802.1ah – Provider Backbone Bridges), later integrated into IEEE 802.1Q-2014 [31.17].

All of them have enough quality of service (QoS) support to carry CPRI and eCPRI, but additional time synchronization support must be provided by other means, as none of these technologies provide this feature. An outline of the key features required to select a suitable packet transport technology, as identified in [31.18], is given in Table 31.6.

As a conclusion, IP lacks the carrier features available in MPLS-TP and Metro Ethernet, like protection switching or OAM protocols, that make it less adequate for carrier-grade inter-connection of RRH and BBUs. Both MPLS-TP (or equivalently, conventional IP/MPLS) and Ethernet are suitable for fronthaul and backhaul transport, although fronthaul over Ethernet is getting standardized faster. Particularly relevant is the work carried out in the *time-sensitive networking for fronthaul* IEEE 802.1cm standard [31.19], which aims to enable the transport of time-sensitive fronthaul streams over Ethernet bridged networks.

Finally, among the standards for fronthaul over packet, the IEEE 1914.3 standard for radio over Ethernet encapsulations and mappings specifies the encapsulation of digitized radio payload and control data into an Ethernet frame payload field. The header format can be structure-aware or structure-agnostic: the structure-aware encapsulation has a detailed knowledge of the format of the client signal; the structure-agnostic encapsulation is a generic container for digitized radio transport frames.

31.5.2 Delay Control in Fronthaul Packet Networks

A fronthaul packet network is expected to supply a service as close as possible to CBR. No specification of maximum jitter is specified by 802.1cm. Only a target maximum end-to-end delay and loss for each type of traffic has been specified. This means that the emulation of the CBR service needs not be synchronous over a packet network, and that the transport jitter can be compensated by a jitter compensation buffer at the BBU and at the RRH. The jitter compensation buffer is out of the scope of 802.1cm and adds on delay to the end-to-end network delay budget (100 µs). Therefore, although the dominating delay factor in the target network is propagation, it is convenient to use switches with low

Table 31.6 Comparison of IP, MPLS-TP, and Ethernet as technologies for fronthaul–backhaul integration

Requirement for a packet fronthaul transport network	IP	MPLS-TP	Ethernet
Standard for fronthaul transport	No Work in progress in IETF DetNet (deterministic networking) working group (WG) applicable; it relies on IEEE TSN. eCPRI UDP/IP/MPLS stack	No Patent for proprietary encapsulation of CPRI exists. Radio over Ethernet/MPLS or eCPRI UDP/IP/MPLS stacks required	Yes IEEE 802.1cm from IEEE TSN (time sensitive networking)
CBR service • Rate guaranteed • Low delay bound	Not native Emulated through network planning and traffic control Guaranteed service scheduling based on priority queuing (PQ) or weighted fair queuing (WFQ)	Not native Emulated through network planning and traffic control Guaranteed service through PQ or WFQ	Not native Emulated through network planning and traffic control 802.1cm: PQ based
Service survivability • Fast re-routing (FRR) • OAM fault management	Slow recovery: IP routing based Limited fast re-routing over tunnels No service OAM protocols	Fast recovery: < 50 ms MPLS facility backup and pseudowire redundancy RFC6718 Protection mechanisms do not need signaling OAM support request for comments (RFCs)	Fast recovery: < 50 ms provider backbone bridges – traffic engineering (PBB-TE) (IEEE 802.1Qay-2009) required for FRR Full OAM support (IEEE 802.1ag for service OAM, 802.3ah for link OAM)
Multiservice capability • Multiple functional splits support: eCPRI, backhaul • Multiplex of many traffic types preserving QoS • Differentiated forwarding behavior	Yes QoS granularity: fine It depends on router support through the path: • Differentiated services code point(DSCP)-based Diffserv classes in the core • Flow (multifield discrimination): unlimited number of classes and policy-based routing in the edge	Yes QoS granularity: Very fine throughout the path. It depends on label switching routing (LSR) support • Per hop behavior (PHB) scheduling and drop preference determined by label switched path (LSP) label traffic class (TC): 3 bits. 8 classes differentiated in core. Pseudowire (PW) label TC: 8 classes differentiated in the edge LSR • Label-only-inferred PHB scheduling class LSP (unlimited)	Yes QoS granularity: coarse Port, virtual local area network (VLAN) and Ethertype-based differentiation Basic scheme: PQ 802.1p: 3 bits 8 classes in the core and 8 classes in the edge switch Other packet scheduling schemes on top
Multitenancy • Tenant identification • Traffic isolation (guaranteed QoS) • Traffic separation (tenant privacy)	No tenant identification in data plane. Administrative labelling of tunnels Per-class or per-flow differentiation	Pseudowire label allows to identify tenant Up to per-LSP differentiation in the core Full isolation requires separate LSPs to carry pseudowires	Service instance ID allows to identify tenant Per-service VLAN differentiation in the core

Table 31.6 (Continued)

Requirement for a packet fronthaul transport network	IP	MPLS-TP	Ethernet
Synchronization support	Lower precision: network time protocol (NTP), precision time protocol (PTP) over IP	None in MPLS-TP Possible support of SyncE Ethernet at Ethernet layer for frequency sync	Higher precision IEEE1588v2, IEEE802.1AS ITU-T G.8261
Security	High if connectivity based on tunnels network management system (NMS) and control plane security: peer-to-peer authentication, encryption, monitoring and logging	High: only edge nodes visible from user–network interface (UNI) No reliance of MPLS-TP on a signalling protocol adds on security with respect to MPLS NMS access security	High: only edge visible from UNI NMS interaction security
Transport efficiency • Low overhead • Multipath support • Per-flow differentiation • Per-class differentiation	High overhead Equal cost multi path (ECMP) multipath support in data plane	Medium overhead ECMP is not available in MPLS-TP, only in MPLS, to fulfil OAM requirements	Low overhead No multipath support in data plane
Service management • OAM protocol	No sophisticated OAM protocol. RFC5880 *bidirectional forwarding detection* (*BFD*) can be setup on tunnels or link-by-link	Yes In-band OAM ITU-T G.8113.1 (based on ITU-T Y.1731 and 802.1ag) and ITU-T G.8113.2 (based on BFD from IETF MPLS OAM framework)	Yes In-band OAM. 802.1ag and Y.1731, fault detection and performance monitoring
Energy efficiency • Energy consumption proportional to traffic	No Shutdown of linecards by control plane	No Shutdown of linecards by NMS	No Energy efficient Ethernet standard only in copper transceivers. Shutdown of linecards by NMS or control plane
Store-and-forward latency	CRC needs recalculation on each hop Cut-through not feasible Longest-prefix-match lookup requires route cache or ternary content-addressable memory (TCAM) for efficient lookup	L2 FCS needs update Cut-through not feasible Exact-match lookup	FCS needs no update on forwarding Cut-through feasible Exact-match lookup

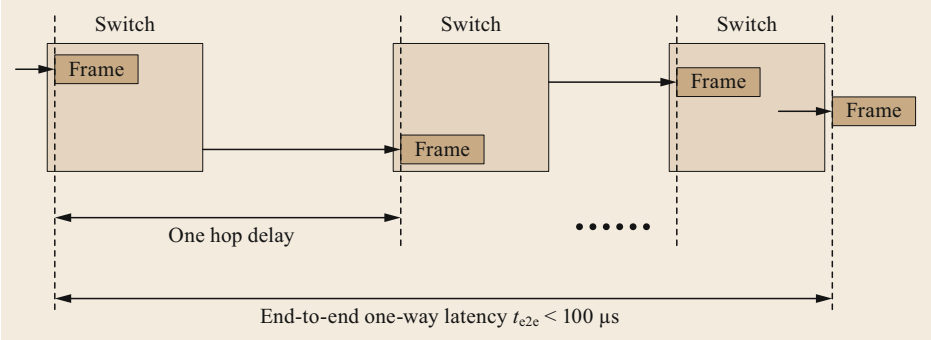


Fig. 31.10 Definition of end-to-end one-way latency according to IEEE802.1Q and maximum value allowed by 802.1cm for fronthaul traffic

and predictable latency, since a low switching delay allows us to extend the RRH–BBU distance and a low jitter allows to reduce the jitter compensation delay.

Figure 31.10 is a pictorial representation of the definition of one hop delay and end-to-end one-way latency according to the IEEE802.1Q standard. One hop delay between two ports of two consecutive bridges is defined as the time interval from the arrival of the last bit of the frame at the port of the first bridge to the arrival of the last bit of the frame at the port of the second bridge. The end-to-end one-way latency is measured from the arrival of the last bit of the frame at port n of the ingress bridge to the transmission of the last bit of the frame by the egress bridge. If the RE and REC are locally attached to the ingress and egress bridges, respectively, no additional propagation delay needs to be taken into account. This will be commonplace as the switch in the RE will aggregate CPRI flows from all the antennas in the base station. Thus, the end-to-end one-way latency is given by the addition of all one-hop delays plus an additional store-and-forward operation, including fabric switching and queuing. Regarding queuing delay, 802.1cm proposes to use the lowest latency mechanism available in Ethernet: fronthaul data flows should be transported with high 801.1p priority, if possible the highest, in order to guarantee bounds. CPRI control and management data should be transported separately (the CPRI frame is not transported as is) in a lower priority class. Synchronization and timing information is not transported; the task is delegated to existing protocols and devices specialized in this issue such as ITU-T Rec. G.8261, G.8262, G.8264 Synchronous Ethernet, and IEEE1588 Precision Time Protocol.

Coming back to the latency question, there are other components to take into account besides queuing delay. The delay components of a complete Ethernet hop and their worst-case values are outlined in Table 31.7. The sum of all these terms gives the value of the worst-case one hop delay. The first column of the table specifies the latency component; the second column is the name of the variable to be used in the formulation below; the third column is the name for the latency component given in the 801.1 standard; the fourth column gives expressions for the minimum and maximum latencies for the component; and the last column provides an explanation for the latency range.

The maximum allowed frame size is 2000 bytes as specified by 802.1cm following IEEE 802.3's maximum frame size. Terms M and M_f must take into account the real duration of the frame on the link, which includes the preamble, start of frame delimiter, and inter-packet gap as well as the frame size.

The maximum delay caused by other classes' frames being transmitted $t_{\text{QueuingOtherClasses}}$ can be reduced if frame preemption is configured (IEEE Std 802.3br-2016), but its effect is usually negligible at the link rates required for fronthaul traffic.

The most architecture-dependent term in the table is the maximum internal switching time t_{IS}. It should be noted that the latency budget in the 802.1cm standard counts the delay from the arrival of the last bit of the frame. Therefore, in theory, the next-hop lookup may have been performed before the last bit enters the switch, and the rest of the frame may have already traversed the fabric. Only the checksum validation and the transfer of the last bit remain. Thus, the theoretical minimum is about a fabric clock cycle. However, normally the transfer of information across interfaces takes place in units larger than a single bit, e.g., in cells of 64 bytes, and hence the minimum time is given by the cell transfer time. If the architecture does not start to switch the frame before the FCS (frame check sequence) has been validated, at least M_f/R_{fabric} is incurred. In practice, the maximum ranges between 1 and 5 µs. This parameter should not be mixed up with the *cut-through switching* latency usually displayed by switch vendors in technical specifications, which is usually measured as the minimum time from the moment the first bit of the frame enters the switch to the time it departs the switch in cut-through mode. Cut-through means that

Table 31.7 Latency components and worst-case bounds for a fronthaul packet if fronthaul traffic is configured to maximum priority

Latency component	Variable	Term used in IEEE 801.1 standard	Latency range with maximum priority 802.1p for fronthaul packets	Explanation
Input queuing	t_{iq}	Input queuing delay	0	Output queued switch architectures do not have input queuing delay. Input queued architectures need to take this term into account
Internal switching	t_{IS}	Store-and-forward delay	0.. hardware-specific maximum Typical maximum values: 1–5 µs	Delay caused by the internal processing of the switch, after the last bit was received Factors: Fabric architecture and speed SAR (segmentation and reassembly) delay (if SAR) Other contending cells/packets, except in a few architectures
Output queuing	$t_{oq} = t_{QueuingOtherClasses} + t_{QueuingThisClass}$	Interference delay		Output queuing delay depends on link occupation. Two components are distinguished $t_{QueuingOtherClasses}$ and $t_{QueuingThisClass}$
	$t_{QueuingOtherClasses}$	Queuing	0.. M/R_{out}	Waiting for other classes to be served is bound to M/R_{out} in maximum priority M = Maximum link frame size (bits) R_{out} = Output link rate (bits/s)
	$t_{QueuingThisClass}$	Self-queuing	0.. b/R_{out}	Waiting for other frames in the same class to be served before the one under consideration (in the worst case all other fronthaul bursts from other input ports are to served before this packet) b = maximum feasible burst generable by all the flows in the class in this port (bits) R_{out} = Output link rate (bits/s)
Transmission	t_{tx}	Frame transmission delay	M_f/R_{out}	Service time M_f = Maximum frame size of the flow (bits) R_{out} = Link rate (bits/s)
Propagation	t_{prop}	LAN propagation delay	$\delta_{medprop}\, L$	$\delta_{medprop}$ = Propagation delay of the link per km (optical fiber (5 µs/km)) L = Link length (km)

the frame is forwarded as the first bits of the frame arrive at the switch without waiting for the end of the frame (FCS). Cut-through latencies for state-of-the-art low-latency data center switches start at 300 ns [31.20]. Two conditions must hold to perform cut-through: a) $R_{in} = R_{out}$ and b) the output line must be idle when the frame arrives. The first condition makes cut-through switching unlikely in fronthaul scenarios given that the aggregation of traffic toward the core requires increasingly higher line rates.

Table 31.7 assumes:

1. that the traffic pattern of each ingress fronthaul flow is deterministic and consists of periodic fixed-size bursts of packets (as generated by CPRI and eCPRI split E)
2. that fronthaul frames have maximum priority (IEEE802.1p priority 7) and the network manager has not oversubscribed it in the output link (i.e., all class 7 is fronthaul traffic, and the link's capacity is not exceeded), and
3. that the switch has a distributed output queued architecture (see the next section).

In this scenario, the total worst-case end-to-end latency for IQ data is given by

$$t_{e2e} = \sum_{j \in \text{BridgeOutputPorts}} t_{\text{bridge}}(j) + \sum_{i \in \text{Links}} t_{\text{prop}}(i)\,, \tag{31.5}$$

where BridgeOutputPorts is the set of output ports used through the network by the frame under consideration, Links is the set of crossed links, and $t_{\text{bridge}}(j)$ is the time spent by the frame in one bridge if the frame is for-

Table 31.8 Example of worst-case packet delay calculation

Use case data	Value	Units	Value	Units	Comment	
T_{ofdm}	0.0000667	s	66.7	µs		
CPRI rate for LTE 2 × 2 MIMO channel of 20 MHz bandwidth	9 830 400 000	b/s	9830.4	Mb/s	CPRI option 7	
burst_size = CPRI rate × T_{ofdm}	655 688	bits	81 961	bytes		
payload_size	12 000	bits	1500	bytes		
burst_size (packets) = burst_size/payload_size	55	packets	55	packets		
RoE and Mac-in-mac overhead per packet	368	bits	46	bytes		
M (maximum frame size)	12 368	bits	1546	bytes		
R (link rate)	1×10^{11}	b/s	100	Gb/s		
t_{iq} (distributed output queuing is assumed)	0	s	0	µs		
t_{IS} (maximum internal processing delay of the employed switch)	0.000003	s	3	µs		
Worst-case delay breakdown **Bridge 1**	**Value**	**Units**	**Value**	**Units**	**Accumulated switching latency**	**Units**
t_{iq}	0	s	0	ns		
$t_{QueuingOtherClasses} = M/R$	1.2368×10^{-7}	s	123.68	ns		
b, packets contending with the observed one	0	bits	0	bytes		
$t_{QueuingThisClass} = b/R$	0	s	0	ns		
$t_{tx} = M/R$	1.2368×10^{-7}	s	123.68	ns		
t_{IS}	0.000003	s	3000	ns		
Total worst-case delay switch 1 ($t_{bridge1}$)	3.24736×10^{-6}	s	3247.36	ns	3.2474	µs
Bridge 2						
t_{iq}	0.000003	s	0	ns		
$t_{QueuingOtherClasses} = M/R$	1.2368×10^{-7}	s	123.68	ns		
b, packets contending with the observed one	655 688	bits	81 961	bytes		
$t_{QueuingThisClass} = b/R$	6.55688×10^{-6}	s	6556.88	ns		
$t_{tx} = M/R$	1.2368×10^{-7}	s	123.68	ns		
t_{IS}	0.000003	s	3000	ns		
Total worst-case delay switch 2 ($t_{bridge2}$)	9.80424×10^{-6}	s	9804.24	ns	13.052	µs
Bridge 3						
t_{iq}	0	s	0	ns		
$t_{QueuingOtherClasses} = M/R$	1.2368×10^{-7}	s	123.68	ns		
b, packets contending with the observed one	1 311 376	bits	163 922	bytes		
$t_{QueuingThisClass} = b/R$	1.31138×10^{-5}	s	13 113.76	ns		
$t_{tx} = M/R$	1.2368×10^{-7}	s	123.68	ns		
t_{IS}	0.000003	s	3000	ns		
Total worst-case delay switch 3 ($t_{bridge3}$)	1.63611×10^{-5}	s	16 361.12	ns	29.413	µs
Bridge 4						
t_{iq}	0	s	0	ns		
$t_{QueuingOtherClasses} = M/R$	1.2368×10^{-7}	s	123.68	ns		
b, packets contending with the observed one	2 622 752	bits	327 844	bytes		
$t_{QueuingThisClass} = b/R$	2.62275×10^{-5}	s	26 227.52	ns		
$t_{tx} = M/R$	1.2368×10^{-7}	s	123.68	ns		
t_{IS}	0.000003	s	3000	ns		
Total worst-case delay switch 4 ($t_{bridge4}$)	2.94749×10^{-5}	s	29 474.88	ns	58.888	µs
Available budget for propagation delay (100 µs – max delay)					41.112	µs
Equivalent maximum end-to-end distance in fiber					8.222	km
Required de-jitter buffer (max queuing delay)					46.3931	µs
Maximum packetization delay (before last packet of burst is sent, b/R)			6556.88	ns	6.5571	µs
Total maximum effective REC–RE latency for OFDM symbol for maximum distance					152.950	µs

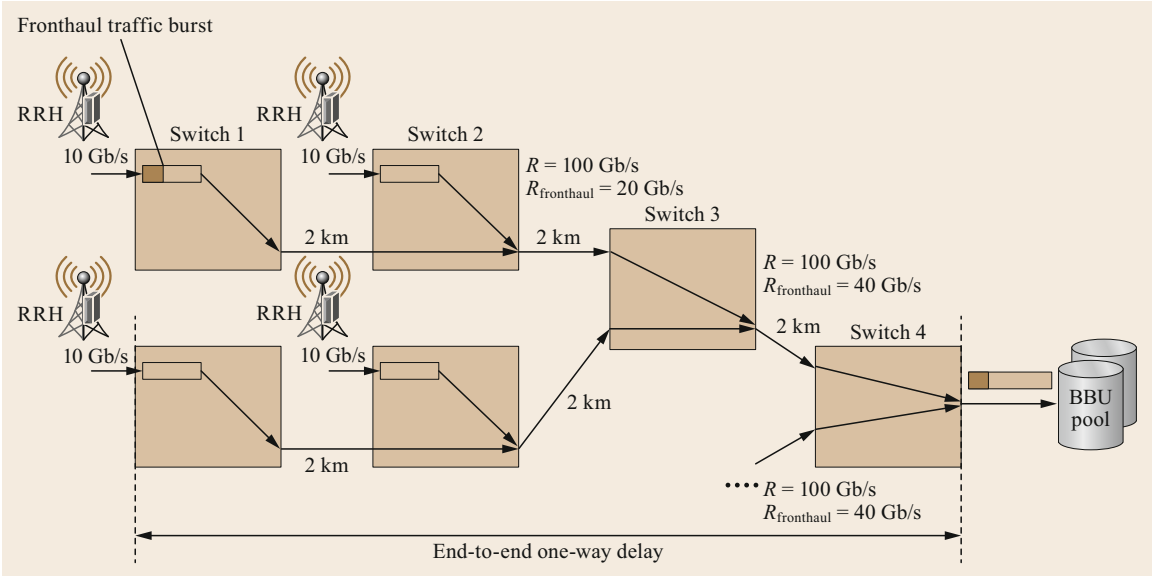


Fig. 31.11 An example of hierarchical aggregation of eight CPRI option 7 flows over a 100 Gb/s Ethernet

warded to output port j. This time is

$$t_{\text{bridge}}(j) = t_{\text{IS}}(j) + t_{\text{tx}}(j) + t_{\text{QueuingThisClass}}(j) + t_{\text{QueuingOtherClasses}}(j) \,. \quad (31.6)$$

Table 31.8 provides bounds for each term for each bridge's output port j traversed. The only nonstraightforward value is $t_{\text{QueuingThisClass}}(j)$, which is given by

$$t_{\text{QueuingThisClass}}(j) = \frac{\sum_{k \in \text{CFFFOP}(j)} G_k M_{\text{f}}(k)}{R(j)} \,, \quad (31.7)$$

where CFFFOP(j) is the set of fronthaul flows incoming from other ports than the one under observation, contending with the packet for output port j. The summation of burst lengths of all these flows makes $b(j)$, $M_{\text{f}}(k)$ is the maximum frame size of fronthaul flow k, G_k is the maximum burst of back-to-back frames (in number of packets) injected by the RRH into the ingress switch (normally, the IQ samples making up an orthogonal frequency division multiplexing (OFDM) symbol) by fronthaul flow k.

The traffic generated by other eCPRI functional splits (eCPRI will be introduced later in this chapter)—namely splits D, I_D/II_D, and I_U in Fig. 31.15—depends on the current radio resource utilization in the cells. In these cases, the formulas above still hold, but the maximum latencies are variable as are the burst sizes. Two main approaches are applicable: either provisioning the network for simultaneous maximum cell utilization in the area or designing the network with a maximum target load based on past statistics, considering that it is very unlikely that all fronthaul sources work at full rate at the same time.

In the case of eCPRI split D and, in general, for splits above the MAC (medium access control) function, the real time requirements of the traffic are much less stringent (1 ms end-to-end). Thus, 802.1cm recommends to transport this fronthaul traffic as medium priority fronthaul.

31.5.3 Example of Worst-Case Delay Calculation

As an example, we analyze whether the setup described in Fig. 31.11 fulfills the maximum latency requirement established for fronthaul: 100 µs. This is the case for a hierarchical aggregation of 8 CPRI option 7 flows (corresponding, for example, to a four-antenna setting per RRH operating an LTE (long-term evolution) 2 × 2 MIMO (multi-input multi-output) channel of 20 MHz bandwidth) over a bridged network of 100 G Ethernet interfaces configured with maximum priority for fronthaul traffic as recommended by the 802.1cm WG. Frame preemption is disabled, and the switching fabric is assumed to be a distributed output-queued architecture (more discussion on this architecture will be provided later in the chapter) that works at line speed.

To perform this analysis (31.5) and (31.6) are applied. Given the 100 µs end-to-end latency budget and

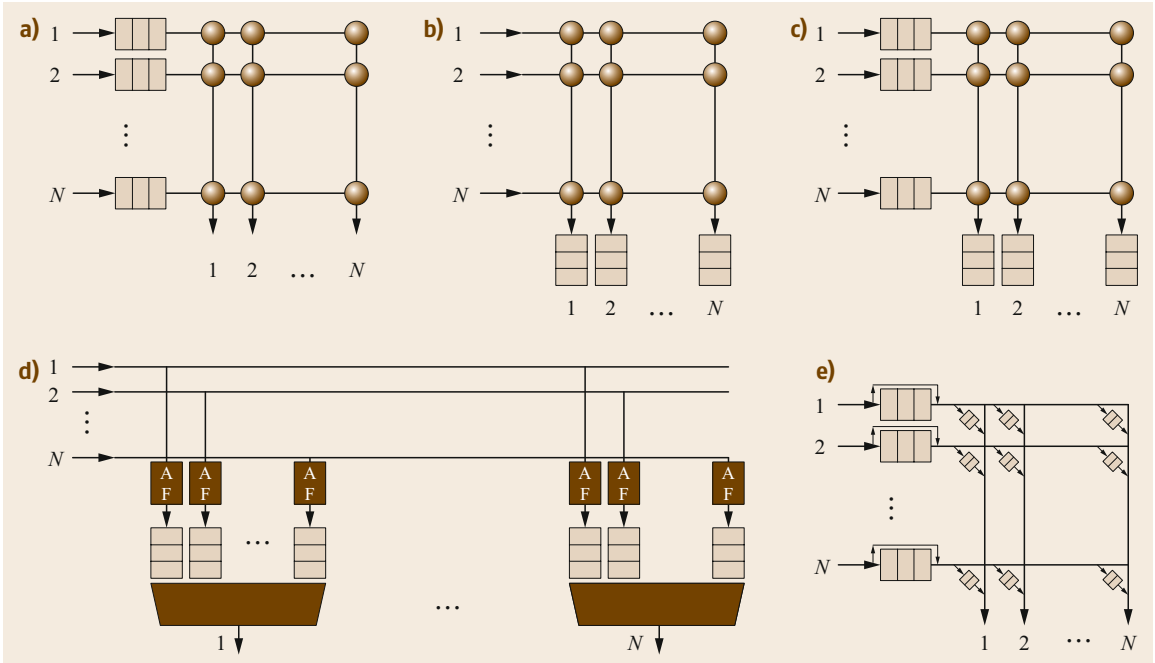


Fig. 31.12a–e Classic parallel architectures for packet switches: **(a)** input queued (IQ), **(b)** output queued (OQ), **(c)** combined input–output queued (CIOQ), **(d)** distributed output queued (DOQ), and **(e)** combined input–crosspoint queued (CIOQ)

in order to obtain more general results, we extract the propagation term from (31.5) as

$$\sum_{i\in\text{Links}} t_{\text{prop}}(i) = 100\,\mu\text{s} - \sum_{j\in\text{BridgeOutputPorts}} t_{\text{bridge}}(j)\,, \tag{31.8}$$

and we focus the calculus on the worst-case time spent on switching and forwarding. The calculations are shown in Table 31.8.

As Table 31.8 shows, the margin for propagation delay is 41 µs, which means a maximum RE–REC distance of 8.2 km. Since in the scenario of Fig. 31.11 the maximum end-to-end distance is 6 km, we conclude that the setup is viable. If the scenario had distances longer than 8 km, a WDM-based solution should be used instead. Packet switching leaves smaller margins than WDM to other latency sources such as propagation delay. It should be considered that the 58.9 µs worst-case switching time calculated in Table 31.8 only includes the networking component, but an additional de-jitter buffer at the receiver is required to emulate a constant bit rate service (CBR). This induces an additional latency at user level of at least 46.4 µs, which corresponds to the total maximum queuing delay terms shown in the table.

31.5.4 Overview of Packet Switch Architectures for Fronthaul

A packet switch for fronthaul transport must be predictable in terms of latency and throughput in order to design a network that fulfils the 100 µs budget established for packetized CPRI and eCPRI splits under the MAC function (splits D, I_D, II_D and I_U in Fig. 31.16). There are a number of switch architectures that can achieve 100% of switching throughput. However, not all are appropriate to get a low and predictable latency. This feature usually comes at the price of more memory units. The most popular parallel switch architectures for a crossbar or crossbar-like fabric are depicted in Fig. 31.12. Other architectural options can be found in [31.21].

Let us review the suitability for fronthaul transport of each architecture. The conclusions are summarized in Table 31.9, where the red color denotes a negative feature, orange denotes not desirable but acceptable, and green means a positive feature for fronthaul transport. In Table 31.9, speed-up is the times the switching fabric needs to be higher than the line rate:

- The input-queued switch (IQ) (Fig. 31.12a) is an inexpensive architecture in terms of buffer re-

Table 31.9 Comparison of parallel packet switch architectures in terms of latency and memory requirements using a 10 Gb/s 32-port design

	Input queued (IQ)	Input queued (IQ)	Output queued (OQ)	Combined input–output queued (CIOQ) = IQ with speedup	Distributed output queued (DOQ)	Knock-out switch (DOQ with KO)	Combined input–crosspoint queued (CICQ)
Scheduling algorithm	None (FIFO)	VOQ + ISLIP	No	VOQ + MUCFA	No	No	VOQ + push-back
SpeedUp (fabric and memory)	1	1	32	4	1	1	1
$r + w$ time budget (ns) = cell_period/speedup	51.2	51.2	1.6	12.8	51.2	51.2	51.2
Throughput (%)	58.60	100	100	100	100	100	100
Memory (# of buffers)	$N = 32$	$N = 32$	$N = 32$	$N = 32$	$N \times N = 1024$	f (target loss)	N regular buffers + $N \times N$ small SRAM buffers
Delay sources	Head-of-line blocking (HOL), SAR, transfer	Scheduling, SAR, transfer, input (or output) queuing	SAR, transfer, output queuing	Scheduling, SAR, transfer, output queuing	Transfer, output queuing	Knock-out process, output queuing, transfer	Pushback under high loads, transfer, crosspoint queuing
Delay control of individual packets	No	No	Yes	Yes	Yes	Yes	Yes
Efficient cut-through	No	No	Yes	No	Yes	Yes	Yes

quirements capable of supporting very high-speed ports. However, it does not provide delay control of individual packets due to the input–output matching scheduler required to solve contention and achieve 100% throughput (Table 31.9). Virtual output queueing (VOQ) and an algorithm like iSLIP [31.21] are the most common approaches. Moreover, achieving a high throughput requires a synchronous crossbar that works with fixed-size data chunks, usually called *cells*. This requires segmentation and reassembly (SAR) of packets into cells, which adds to the delay.

- The *output-queued (OQ)* architecture (Fig. 31.12b) does have delay control, as packet chunks are written to the output buffers as they arrive at the linecard, but the required speed-up (equal to the number of ports) makes it unsuitable for very high-speed ports. The design used as reference in the table, is a 32-port × 10 Gb/s switch with a 64-byte cell size. With a speed-up of $N = 32$, the margin to read from the input port and write to the output buffer a cell is too low.
- *Combined input–output queuing (CIOQ)* (Fig. 31.12c) is a good alternative, as it can emulate an OQ switch with a smaller speed-up if a specific scheduling algorithm is used. A speed-up of 4 has been proven to suffice to emulate the performance of an OQ with the MUCFA algorithm (most urgent cell first algorithm) [31.22] and virtual output queueing. However, CIOQ, as well as IQ and OQ, require segmentation and reassembly (SAR) to achieve 100% throughput irrespective of packet sizes or to take advantage of the speed-up, in the case of OQ and CIOQ. Moreover, the scheduling computation time itself adds an extra delay (except for OQ).
- The *Distributed output-queued architecture (DOQ)* (Fig. 31.12d) is the ideal architecture in terms of latency, as packets do not need segmentation to traverse the switching fabric and are written straight to the dedicated output buffer that each input has at each linecard, without the need for SAR or an input–output scheduling mechanism to traverse the fabric. The problem is its high cost in terms of memory (N^2 buffers, where N is the number of ports). Two main alternatives emulating DOQ are described in the literature. One is the knock-out switch (KO) [31.24], which reduces the number of parallel buffers available at each output under the assumption than not all ports statistically need to write simultaneously on the same output port. If this happens, some of the contending packets are discarded (knocked-out) by a concentrator stage. The other alternative is CICQ.

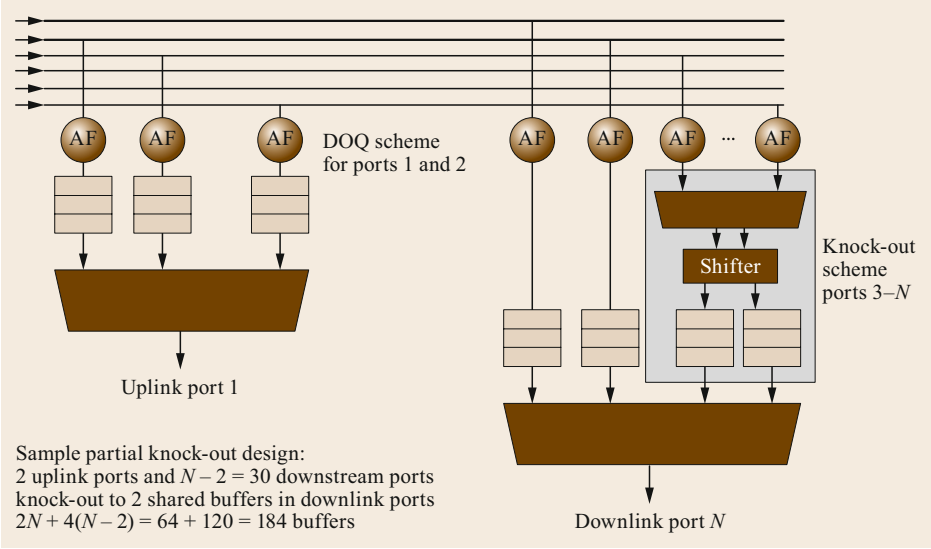


Fig. 31.13 A partial knock-out switch for fronthaul aggregation/distribution scenarios (after [31.23])

- *Combined input-crosspoint queuing (CICQ)* is an interesting alternative to emulate a DOQ switch and consists of a crossbar fabric with integrated static random access memory (SRAM) buffers at its crosspoints. This is, in principle, equivalent to a DOQ switch with $N \times N$ buffers, one per input–output pair. However, in order to reduce cost, the capacity of the crosspoint buffers is small, and the excess of packets is stored at an input buffer (which is used only if the target crosspoint buffer is full). This architecture provides a trade-off between latency and memory requirements, as the input buffer memory is shared by all crosspoints of that input, and no scheduling or SAR is required.

In practice, hybrid approaches to the basic alternatives described here are employed.

In the case of fronthaul traffic, the line rates are high 10, 40, or 100 G, and there is a hierarchy of ports from the cloud (uplink ports) to the RRHs. Unlike other contexts where latency is also essential (e.g., data center high performance computing switches), all fronthaul traffic goes to or comes from the BBU, and all backhaul traffic goes to or comes from the evolved packet core. Thus, a number of simplifications can be made to reduce costs. For example, a partial knock-out switch [31.24] can be suitable for aggregation switches, as depicted in Fig. 31.13. In this switch, uplink ports have all N buffers required in DOQ. In addition, each downlink port has a buffer dedicated to each uplink port (in Fig. 31.13, two buffers) so as not to lose any fronthaul packet. However, since no fronthaul traffic is expected from a downlink port to another downlink port, a knock-out concentration stage can be used to make it possible to share a small set of buffers by many input ports and save memory.

An alternative popular centralized architecture that behaves like a DOQ switch is the distributed shared memory switch. This architecture relies on a central memory manager that drives one (or two) $N \times M$ crossbars in such a way that the linecards can access in parallel a number of $M \geq N$ shared memory banks both for writing and reading packets. This requires the aggregate memory bandwidth to be $2NR$, which can be accomplished by setting an appropriate value for M. The architecture is classified as centralized, because a single device—the memory manager—has to deal also with the virtualization of the distributed memory as a number of independent output buffers, including per-output-port per-class packet scheduling. That means that the memory manager must handle pointers to all the packets in the switch, which is an intensive operation that requires complex coordinated parallel processing itself. This architecture does not scale well with the number of ports and line rate but can also be optimized for aggregation/distribution, provides latency guarantees, can support cut-through operations, and can switch full frames at once.

31.6 Evolution of Fronthaul Networks

The 3GPP studies different functional splits between remote and baseband unit ([31.23], Fig. 31.14). Option 8, which has the split point close to the antenna, is compatible with both analog and digital radios over fiber. All other options correspond to different D-RoF formats, leading to different fronthaul interfaces and requirements.

31.6.1 Functional Split Between Remote and Baseband Unit

A detailed analysis of the different split options is out of the scope of this chapter, as is an exact calculation of bandwidth and latency values, which may depend on use case and implementation. Thus, we will only provide approximate values of bit rate and latency, useful to dimensioning a transport network. The tolerable latency is low for the highest split options, from Options 5 to 8, where the hybrid automatic repeat request (HARQ) protocol is performed in the BBU. This is critical in a transport network, which introduces latency sources due to propagation in fiber and packet processing.

For Options 1 to 4, the tolerated latency is in the order of several milliseconds, a value compatible with common transport network deployments. The D-RoF implementation of Option 8 corresponds to CPRI, which has already been discussed in this chapter, so we will focus on Options 5 to 7.

In Option 5, the latency requirement is in the order of few milliseconds, and the transport capacity is only around 10% higher than the backhaul. Option 6 imposes stricter latency requirements ($< 250\,\mu s$), but the capacity is similar to Option 5. Current packet or fiber access technologies might not be able to guarantee this low latency value, so that a multilayer optical transport network may be required, as discussed later in this chapter. Similar considerations hold for Option 7, which has similar latency requirements ($< 250\,\mu s$) but higher peak capacity (tens or hundreds of Gbit/s). The use of 100 Gbit/s optical transceivers may be needed to support this option, with significant cost implications for the transport network.

In parallel to the 3GPPP work, the CPRI consortium recently introduced a new packet based interface, known as eCPRI, for bandwidth-efficient fronthaul, based on split points internal to the physical layer. Details are provided in Sect. 31.6.2.

31.6.2 eCPRI: The New Packet Fronthaul Interface

CPRI cooperation has released the specification of a new interface, known as eCPRI [31.25], which presents increased bandwidth efficiency compared to CPRI and is based on new functional partitioning of the base station functions, positioning the split point inside the physical layer (Option 7 split in Fig. 31.15). The required bandwidth scales according to the user-plane traffic and the use of packet-based transport technologies, like Ethernet, is enabled. This paves the way to carrying eCPRI traffic and other kinds of traffic concurrently in the same packet-switched network, according to the crosshaul concept introduced in Sect. 31.1. The physical layer adopts the IEEE 802.3 standards for optical interfaces implementable in small form pluggable formats. The bit rate depends on the split point: eCPRI specifies three different reference splits, as illustrated in Fig. 31.15, two splits in downlink (Split I_D and II_D) and one split in uplink (Split I_U). Splits E and D in Fig. 31.15 correspond to traditional fronthaul and backhaul, respectively. The user data bit rates (Gbit/s) estimated in [31.25] for the different split options are shown in Table 31.10. The estimation corresponds to a scenario with 3/1.5 Gb/s end-user DL/UL rates, 100 MHz air bandwidth, 8/4 downlink/uplink MIMO layers, no MU-MIMO, 256 QAM, re-

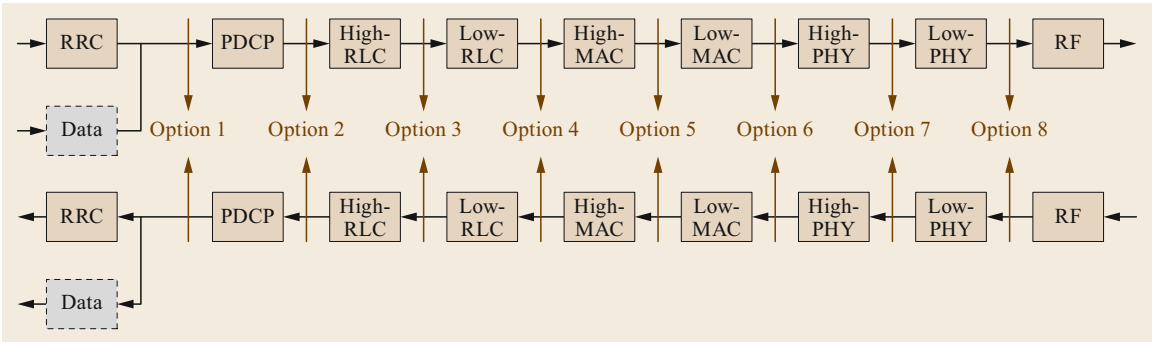


Fig. 31.14 Functional split between remote and baseband unit

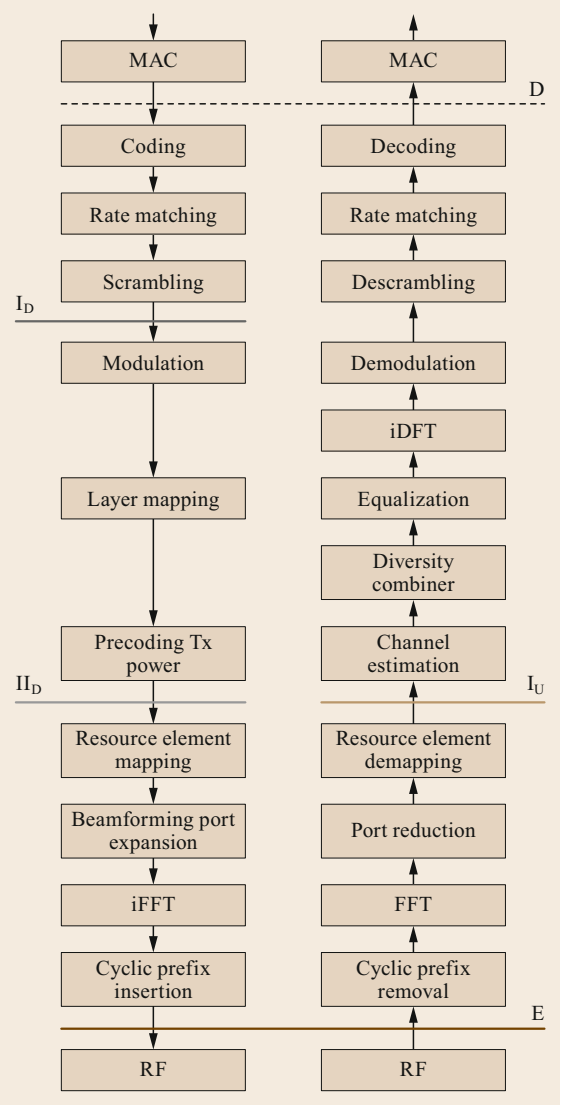


Fig. 31.15 Different RAN split points in the eCPRI specification

Table 31.10 User data bit rates (Gb/s) for the different splits specified for eCPRI [31.25]

	D	I_D	II_D	E
Downlink	3	4	20	236
Uplink	1.5	≈ 20	20	236

mote beam forming calculation, 64 antennas, 15 kHz subcarrier spacing, 30 bit/sample × 122.88 Msample/s, and code rate matching of 0.8.

It is evident that the newly introduced functional splits significantly reduce the bit rate compared to CPRI. In downlink, three processes mostly determine the bit rate: modulation, beam-forming, and IFFT (inverse fast Fourier transform). By moving the split upwards the bit rate is lowered, and vice versa.

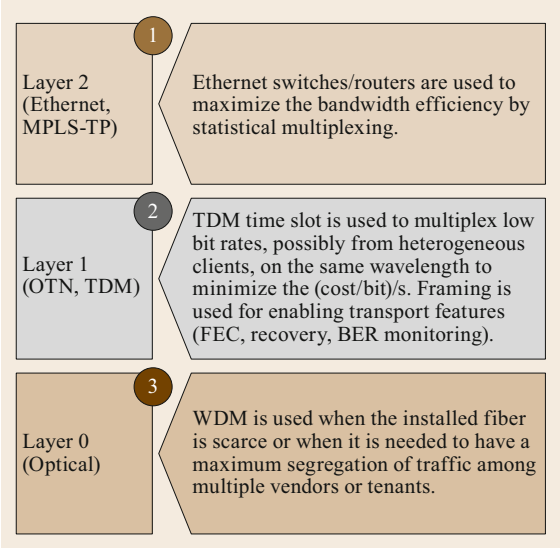


Fig. 31.16 A three-layer network

In combination with the fronthaul interface specification, the CPRI consortium also released a document with preliminary transport requirements [31.8]. In the document, three classes of services are defined (high, medium, and low), tolerating 100 µs, 1 ms and 100 ms as maximum one-way packet delay. Categories are also defined for the maximum timing error that can be provided by the transport network relative to an absolute time reference. Depending on the category, the error spans from 65 ns to 3 ms.

31.6.3 Multilayer Fronthaul Networks

5G transport networks will be able to concurrently transport all kinds of fronthaul interfaces. Moreover, with the advent of 5G, the amount of fronthaul traffic to be carried will become very high. These two factors lead to a need for fully flexible capacity allocation mechanisms in the network.

For this purpose, it is possible to use a multilayer architecture, where all or some of the multiplexing mechanisms discussed in this chapter (physical layer, circuit, and packet multiplexing) coexist to meet different requirements, for example due to different split options coexisting in the same transport network. Classic multilayer networks include three layers (Fig. 31.16):

- Layer 0 (physical layer, i.e., wavelength multiplexing), based on a ROADMs (reconfigurable add-drop multiplexers), providing circuits with wavelength

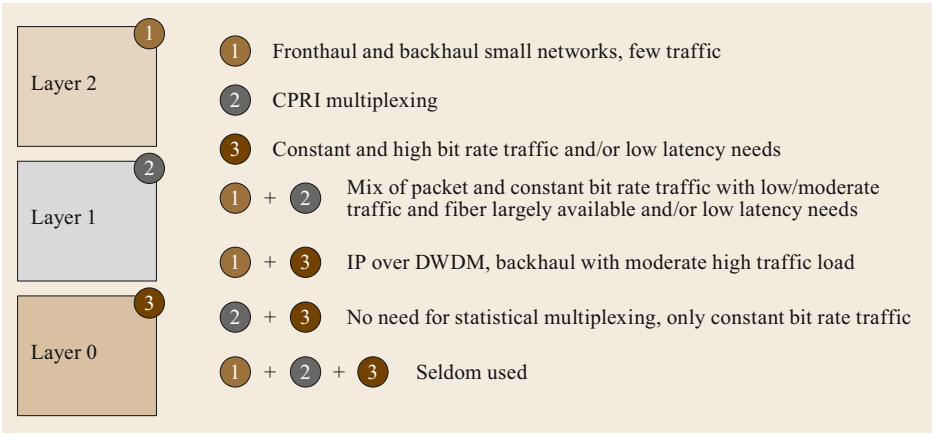


Fig. 31.17 Multi-layer network use cases

granularity. This is the option used for high bit rate fronthaul signal or when the latency requirement is very tight.

- Layer 1 (circuit multiplexing, i.e., OTN or other TDM technology) providing circuits with subwavelength granularity and a possible time and frequency synchronization mechanism. This is the option used when the latency requirement is demanding, so that it is not possible to rely on statistical multiplexing, but the bit rate is low, so that it is convenient to share the cost of the optical interface among different fronthaul signals, time-multiplexing them on the same optical channel.
- Layer 2 (Ethernet, MPLS-TP) providing a packet switching level to take advantage of multiplexing gain, transmitting at a bit rate proportional to the actual data load. This is the option for networks with low or moderate traffic load so that packet switches or routers can be interconnected with a low probability of experiencing congestion situations.

Three layers may result in excessive cost, and a switching node design based on a subset of layers suffices in many cases. Figure 31.17 illustrates possible combination of the layers to fulfil different deployment needs or operator use cases.

31.7 Conclusions

5G will drive a deep transformation of the optical transport network for several concurrent causes: increase of the user capacity, densification of the mobile network, and introduction of several functional split options that may coexist on the same transport network. The latter issue implies a blurring of the traditional distinction between fronthaul and backhaul networks. Since it is impossible, for cost and operational reasons, to design dedicated networks for all types of signal to be transported, the definition of a shared network infrastructure and common transmission and switching technologies is unavoidable.

In this chapter, several candidate technologies for D-RoF have been presented. All of them show pros and cons, and, as of today, it is not possible to foresee a winner. For example, packet-based fronthaul aggregation is likely to be implemented in scenarios with very variable demands such as those of small cells, where statistical multiplexing gains are very relevant. On the contrary, the high fronthaul traffic rates implied by 5G new radio and massive MIMO through large access networks with sustained load levels seem to justify the deployment of WDM-based optical circuit solutions.

References

31.1 B.R. Ballal, D. Nema: Performance comparison of analog and digital radio over fiber link, Int. J. Comput. Sci. Eng. Technol. **3**(6), 193 (2012)

31.2 S. Kuwano, Y. Suzuki, Y. Yamada, K. Watanabe: Digitized radio-over-fiber (DROF) system for wide-area ubiquitous wireless network. In: *MWP '06, Int. Top. Meet. Microw. Photonics* (2006), https://doi.org/10.1109/MWP.2006.346570

31.3 China Mobile: The road towards green RAN, http://labs.chinamobile.com/cran/wp-content/uploads/2014/06/20140613-C-RAN-WP-3.0.pdf (2014)

31.4 A. de la Oliva, X.C. Perez, A. Azcorra, A. Di Giglio, F. Cavaliere, D. Tiegelbekkersk, J. Lessmann, T. Haustein, A. Mourad, P. Iovanna: Xhaul: towards an integrated fronthaul/backhaul architecture in 5G networks, IEEE Wirel. Commun. **22**(5), 32–40 (2015)

31.5 CPRI cooperation: CPRI specifications version 7.0, http://www.cpri.info/spec.html

31.6 A. de la Oliva, J.A. Hernández, D. Larrabeiti, A. Azcorra: An overview of the CPRI specification and its application to C-RAN-based LTE scenarios, IEEE Commun. Mag. **54**(2), 152–159 (2016)

31.7 Open Base Station Architecture Initiative: BTS system reference document version 2.0, http://www.obsai.com/ (2006)

31.8 CPRI Consortium: eCPRI transport network D0.1 (2017-08-30), common public radio interface: requirements for the eCPRI transport network, http://www.cpri.info/ (2017)

31.9 M. Hajduczenia, S. B. Carlson, D. Dove, M. Laubach, D. Law, G. A. Zimmerman: Evolution of ethernet standards in IEEE 802.3 Working Group, IEEE Standards University, https://www.standardsuniversity.org/e-magazine/august-2016-volume-6/evolution-ethernet-standards-ieee-802-3-working-group/ (2016)

31.10 CTC Technology: Dark fiber lease considerations, http://www.ctcnet.us/DarkFiberLease.pdf (2012)

31.11 ITU-T: Recommendation ITU-T G.694.2, spectral grids for WDM applications: CWDM wavelength grid, http://www.itu.int/rec/T-REC-G.694.2-200206-S (2002)

31.12 ITU-T: Recommendation ITU-T G.652, characteristics of a single-mode optical fibre and cable, http://www.itu.int/itu-t/recommendations/rec.aspx?rec=13076 (2016)

31.13 ITU-T: Supplement ITU-T G Suppl. 56, OTN transport of CPRI signals, http://www.itu.int/itu-t/recommendations/rec.aspx?rec=12838 (2016)

31.14 ITU-T: Recommendation ITU-T G.709/Y.1331, interfaces for the optical transport network, http://www.itu.int/itu-t/recommendations/rec.aspx?rec=12789 (2016)

31.15 M. Bocci, S. Bryant, D. Frost, L. Levrau, L. Berger (eds.): A framework for MPLS in transport networks, by IETF, RFC 5921, https://tools.ietf.org/html/rfc5921 (2010)

31.16 S. Bryant, L. Martini, G. Swallow, A. Malis: Packet pseudowire encapsulation over an MPLS PSN, by IETF Standards Track, RFC 6658, https://tools.ietf.org/html/rfc6658 (2012)

31.17 IEEE Standards Association: IEEE 802.1Q-2014—IEEE standard for local and metropolitan area networks—bridges and bridged networks, https://standards.ieee.org/findstds/standard/802.1Q-2014.html (2014)

31.18 5G Crosshaul the integrated fronthaul/backhaul H2020-ICT-2014-2, Project ID: 671598, http://5g-crosshaul.eu/ (2015)

31.19 IEEE time-sensitive networking for fronthaul—IEEE Std 802.1CM-2018, http://www.ieee802.org/1/pages/802.1cm.htm (2018)

31.20 Mellanox: SN3000 series, http://www.mellanox.com/related-docs/prod_eth_switches/BR_SN3000_Series.pdf (2019)

31.21 H.J. Chao, B. Liu: *High Performance Switches and Routers* (Wiley-IEEE, New York 2007)

31.22 B. Prabhakar, N. McKeown: On the speedup required for combined input and output queued switching. In: *Proc. 1998 IEEE Int. Symp. Inf. Theory* (IEEE, Cambridge 1998) p. 165, https://doi.org/10.1109/ISIT.1998.708760

31.23 3GPP: RAN3 technical report TR38.803v1.1.0, http://www.3gpp.org/ftp/Specs/archive/38_series/38.801/38801-100.zip (2017)

31.24 Y.-S. Yeh, M. Hluchyj, A. Acampora: The knockout switch: a simple, modular architecture for high-performance packet switching, IEEE J. Sel. Areas Commun. **5**(8), 1274–1283 (1987), https://doi.org/10.1109/JSAC.1987.1146645

31.25 CPRI Consortium: eCPRI specification V1.0 (2017-08-22)—common public radio interface: CPRI interface specification, http://www.cpri.info/ (2017)

Fabio Cavaliere

Ericsson
Pisa, Italy
fabio.cavaliere@ericsson.com

Fabio Cavaliere is Expert in Photonic Systems and Technologies at Ericsson, having the goal to contribute to develop the technical and standardization strategy in this area. In 20 years of professional experience in optical transport, his research activity encompassed mobile fronthaul and backhaul networks, fixed fiber access, WDM systems, high speed optical transmission, and integrated photonics. Fabio is editor of the ITU-T recommendation G.698.4 (ex G.metro) and member of the Board of Stakeholders of Photonics 21, the European Technology Platform (ETP) representing the European Photonics Community, forming a Photonics Public Private Partnership back in 2012. He was also a member of the High-Level Steering Committee appointed by the European Commission to setup research agenda, implementation model and governance model of the Quantum Flagship. He co-authored 129 filed patent applications and more than 50 publications on optical networks.

David Larrabeiti

Universidad Carlos III de Madrid
Leganés, Spain
dlarra@it.uc3m.es

Dr David Larrabeiti has been with the Telematics Department of UC3M since 1998, where he holds the position of Full Professor in the Chair of Switching and Network Architectures. In 2005, he was technical manager of REDIMadrid. He has participated in EU research projects on next generation broadband networks and has co-authored two books on optical networks. In 2009–2010 he was visiting scholar at Stanford University.

32. Analog Optical Front-Haul Technologies and Architectures

John E. Mitchell

New traffic-generating technologies such as the Internet of Things and continued growth in mobile traffic are requiring mobile networks to rapidly expand their capacities. However, these increased capacity demands are not coupled to increased revenue, so the efficiencies and new architectures of the access network are required to keep pace. Optical fiber is now the medium of choice for radio access network backhaul due to the range of advantages it offers and the new architectural configurations it enables. These advantages are becoming particularly attractive as radio access networks shift, as predicted, to an increased number of small cells (micro- and picocells) to support 5G coverage. With increased bandwidths and higher-frequency systems, the case for radio-over-fiber systems grows. Although digital techniques currently dominate for fronthaul applications, the increased bandwidth requirements of digital interfaces mean that analog techniques may become favorable as capacity grows. In this chapter, we review the requirements of analog radiofrequency (RF)–optical transport systems and provide an overview of the different optical configurations that are used or have been proposed for radio-over-fiber networks.

In addition to the continued growth in mobile traffic, new traffic-generating technologies such as the Internet of Things are driving a rapid expansion in the capacities required in mobile networks. To serve this demand, the complexity and capacity requirements of radio access networks (RANs) have increased. This mandates the use of optical fiber technology to distribute radio signals due to the range of advantages it offers and the new architectural configurations it enables. These advantages are becoming particularly attractive as RANs shift, as predicted, to an increased number of small cells (micro- and picocells) to support 5G coverage. However, we must recognize that these increased capacity demands do not deliver increased revenue, so cost reductions in the access network are required to keep pace. In this context, these networks will form part of a much more integrated public wireless system that will be complementary to WiFi hotspots and other network technologies, forming a seamless and ubiquitous user experience.

To date, radio-over-fiber and distributed antenna techniques have mostly been deployed where the frequencies involved are difficult to transport using coaxial cables or where there are dead zones in radio coverage (e.g., in buildings). However, this is starting to change with new developments in the network architecture of the radio access network. In this chapter, we will explore how optical techniques are being deployed to support radio access networks. We consider the techniques required and the shift in network architecture that is promoting the use of optical fiber technology.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_32

32.1 Developments in Radio Access Networks

In the traditional model of a radio access network (RAN), baseband data packets are transmitted to remote cell sites where the base transceiver station (BTS) provides all the processing, modulation, and amplification required before the signal is radiated by the antenna. In the uplink direction, the BTS performs a similar function, processing the received radio communication into framed baseband data. The connection to these cell sites, typically termed *backhaul*, may be based on microwave links, copper leased lines, or, in some cases, optical fiber. The ultimate capacity available to the cell is in part determined by the capacity of these links, but it is predominantly controlled by the amount of processing equipment at the cell site. The space and plant required varies, but $> 2\,m^2$ cabinets or even small buildings are not uncommon in GSM (Global System for Mobile Communications) or UMTS (Universal Mobile Telecommunications Service) networks. The basic structure of such a distribution architecture is shown in Fig. 32.1.

For a long time, the radio-over-fiber (RoF) community have argued for a fundamental change in this architecture. The proposal is to separate baseband processing from the antenna, with the radio signals being distributed by means of optical fibers to form a connection called *fronthaul*, as shown in Fig. 32.2. In this model, the equipment and functionality placed at the base station is significantly reduced to just the remote radio unit (RRU). In this approach, the RRU only performs optical/electronic (or electrical/optical) conversion; signal amplification and all other processing are carried out at a centralized site. This allows for the baseband unit (BBU)—which performs all of the digital processing in the BTS—to be centrally located, enabling increased equipment and support service sharing. In addition, the use of optical fiber is advantageous as it is a small and lightweight medium that is immune to electromagnetic interference (EMI). Such antenna remoting techniques also open up the possibility of wideband and protocol-independent antenna units delivering a number of radio standards distributed over a single fiber infrastructure.

Although it has been an active research area since the mid-1990s, this concept was largely ignored by industry for many years outside of the use of distributed antenna systems (DASs) in specialist environments. However, it has recently been adopted, albeit not in the analog format originally proposed. In *digital radio over fiber* (D-RoF), the concept of the remote processing and optical distribution of signals is maintained, but the signals that are distributed are digitized versions of the analog waveforms, not the analog waveforms themselves, as would be the case in what is now termed *analog radio over fiber* (A-RoF). This typically results in the network configuration shown in Fig. 32.3.

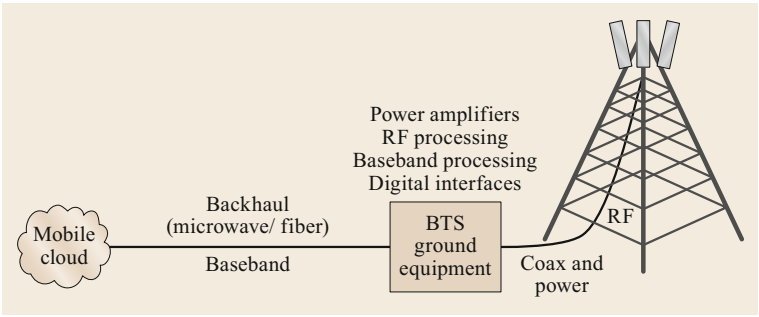


Fig. 32.1 A traditional radio access network

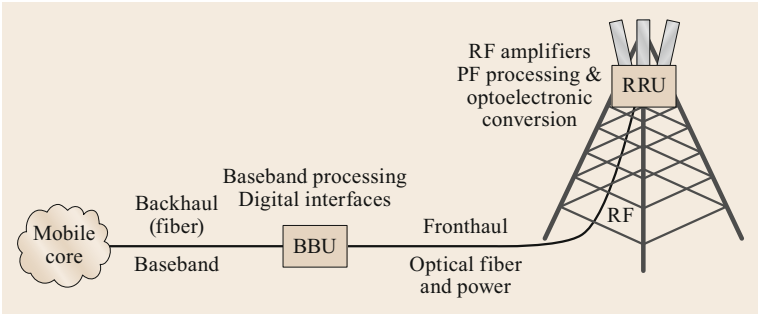


Fig. 32.2 Centralized vRAN architecture

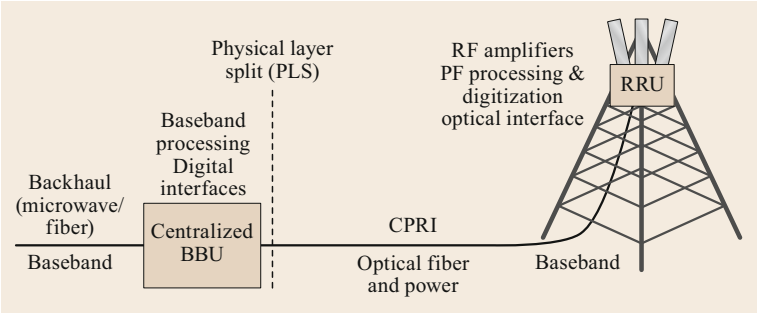


Fig. 32.3 Fronthaul network architecture using CPRI

This approach has advantages, as well as a number of constraints, as will be discussed below. The details of D-RoF and, in particular, the two industrial D-RoF standards OBSAI (Open Base Station Architecture Initiative) [32.1] and CPRI (Common Public Radio Interface) [32.2], which are now widely used, are not discussed here and will be covered elsewhere.

Once the requirement to have the baseband unit colocated at the antenna site is removed, a whole host of new possibilities emerge. This is now being exploited in Long-Term Evolution (LTE; also commonly termed 4G) systems as a distributed radio access network (D-RAN) or cloud radio access network (cloud RAN or C-RAN) architecture. C-RAN involves the pooling of BBU resources in a single location to enable faster service delivery and cost savings. In the longer term it will enable improved coordination of radio capabilities because capacity can be switched between RRUs rapidly, as well as virtualization of the equipment at the central station. This will provide significant cost reductions. Most currently deployed versions use a digital interface from the centralized BBU to the RRU, typically with high-capacity 10 or 100 Gbit ethernet fiber interfaces.

However, analog options to replace the digital interface offer a number of advantages. In the rest of this chapter, we will discuss the basic configuration of an analog radio-over-fiber network, consider its advantages compared to digital radio-over-fiber networks, and look at the networking technologies that will be required to integrate such technologies into wide-scale networks.

32.2 General Overview of Analog Radio-Over-Fiber Technologies

The basic structure of an analog radio-over-fiber system shares many of the same components that are typically witnessed in a digital system. However, there are significant differences in performance requirements when the signal to be transmitted is analog.

32.2.1 Basic Structure and Concept

Figure 32.4 shows the general structure of a radio-over-fiber link, with an optical transceiver located at the central office taking a feed from the output of the radio node. In the downstream direction (central office to base station), the radio signal—which may already be at the final radiofrequency (RF) for transmission, or may be at a lower intermediate frequency (IF) for upconversion at the antenna's site—is modulated onto a lightwave carrier by an optical to electrical converter. The optical transceiver can take a number of different forms, as will be discussed later. In the simplest variety, it may be a directly modulated laser such as a distributed feedback (DFB) laser, Fabry–Perot (FP) laser, or vertical cavity surface-emitting laser (VCSEL), or it may utilize an external modulator. Once converted to the optical domain, the radio signal is then transported across an optical network to the RRU located at the cell site. At this point, the signal is recovered from the optical carrier (typically with a photodiode) and amplified (if necessary), ready for wireless transmission. The losses in the optical fiber are significantly less than those in a coaxial cable, although, depending on the design, the electrical–optical–electrical conversion process can introduce loss into the link.

The return path is identical in terms of modulation onto an optical carrier, transport, and recovery, although as the input RF is typically considerably lower, this link direction is typically far more challenging from a circuit design viewpoint. A key aim of radio-over-fiber technology is to reduce the cost, size, and power requirements at the RRU, so considerable research effort has been directed into the development of low-cost devices for inclusion in the RRU.

In the research literature, there are examples of single optical devices that can both receive and transmit the optical signals, including some that demon-

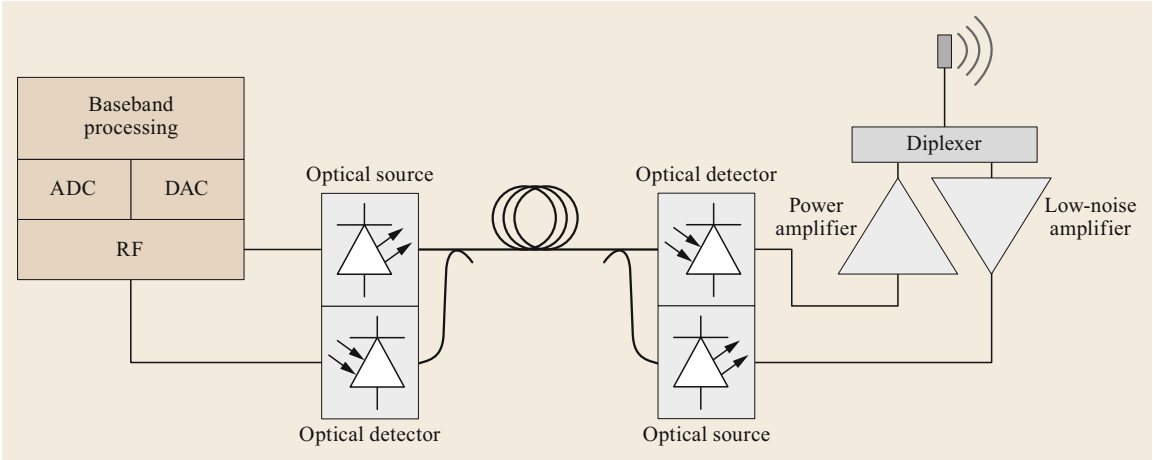


Fig. 32.4 Generic analog radio-over-fiber (A-RoF) configuration

strate 'powerless' operation—often termed *passive picocells* [32.3]. Based on electroabsorption transceiver (EAT) devices, they recover optical signals within a specific wavelength band and modulate data onto a carrier within a second wavelength band [32.4]. An alternative device with a similar operating principle is the asymmetric Fabry–Perot modulator (AFPM) [32.5], which has been employed for wireless transmission at 5.5 GHz [32.6]. In most cases, however, traditional optical techniques such as direct laser diode modulation or external modulators have been used. In Sect. 32.3, we will discuss the different technologies available for analog transmission in more detail.

32.2.2 Frequency Bands and Applications of A-RoF

In this section, we consider the radio standards that have typically been utilized for RoF systems. Table 32.1 provides a review of the most common radio standards required today. It is important to remember that, in most cases, the RoF system has little opportunity to interact or influence the standard in any way. The radio standards are usually determined without considering whether optical distribution networks of the sort proposed here are used in the link, so any optical link must support the standard exactly as set. In effect, what is required is transparent transportation of the radio signal, ideally without imparting any (or, at best, exceptionally small amounts of) loss, distortion, or delay. With careful system design, the first two of these ideals can be approached. However, it can be more difficult to deal with the impact of a delay in the transported signal. Although the signals travel at approximately two-thirds the speed of light in the fiber, and there is little processing delay with analog systems, some short-range systems are not designed to cope with the additional delay that transportation over any significant distance introduces. As an example, consider WiFi (WLAN). A number of systems have shown that at the physical layer, modulated WiFi signals can be transported via the direct modulation of a laser onto an unamplified optical fiber link in excess of 100 km. However, the link layer has been shown to be unable to cope with an additional delay, limiting transmission to < 10 km without modifying the media access and control (MAC) layer of the protocol [32.7].

Early commercial implementations of analog radio-over-fiber were typically focused on cellular radio systems (e.g., GSM and UMTS) and were mainly for in-building applications. These included single-mode fiber systems such as BriteCell from Andrew Corp. (now part of Commscope) [32.22] and multimode fiber solutions for in-building solutions, such as the systems available from Zinwave [32.23], LGC Wireless (now part of Commscope) [32.24], and ADC (now also part of Commscope, having originally purchased LGC Wireless) [32.25]. Outside of cellular, there are a number of providers of radio-over-fiber equipment (such as Foxcom [32.26]) who specialize in military and satellite systems.

As discussed above, different radio standards present different issues depending on the design of the radio formats at both the physical and MAC layers. However, in most cases, the primary issue of concern is the susceptibility to distortion that may be imparted by the optical components in the link. This is often related to the bandwidth requirement of the signals to be transported. For example, a number of the systems proposed to operate in the 60 GHz region require very wideband bandwidths (typically > 500 MHz) that are significantly greater than traditional radio standards. These systems introduce a number of new challenges

Table 32.1 Summary of current wireless modulation formats

Standard	Modulation formats	Bit rate (max.)	Frequencies (approx./typical)
GSM [32.8]	GMSK	14.4 kbit/s (GSM), 171.2 kbit/s (GPRS)	900, 1800, 1900 MHz
UMTS (3G) [32.9]	CDMA, HPSK (UL), QPSK (DL)	384 kbit/s	1.8–2.2 GHz, 800–900 MHz
802.11b [32.10] (WiFi)	Single carrier BPSK, QPSK, 16QAM, or 64QAM	11 Mbit/s	2.4 GHz
802.11a/g (WiFi) [32.11]/[32.12]	52-carrier OFDM, BPSK, QPSK, 16QAM, or 64QAM	54 Mbit/s	5 GHz/2.4 GHz
802.11n (WiFi) [32.13]	52-carrier OFDM, with MIMO	< 300 Mbit/s	5 GHz and/or 2.4 GHz
IEEE 802.15.3c [32.14]	Single carrier (BPSK, MSK, QPSK, 8PSK, 16-QAM) or OFDM (QPSK, 16-QAM, 64-QAM)	5.28 Gbit/s (single carrier), 5.775 Gbit/s (HSI OFDM), 3.8 Gbit/s (AV OFDM)	60 GHz
LTE (Long-Term Evolution) Release 8 [32.15]	Orthogonal frequency division multiple access (OFDMA) downlink, single-carrier frequency domain multiple access (SC-FDMA) uplink	Up to 300 Mbit/s downlink and 75 Mbit/s uplink	Various
3GPP LTE Advanced Release 10 [32.15]	Carrier aggregation (CA), MIMO 8×8 downlink and 4×4 uplink	Up to 1 Gbit/s downlink and 500 Mbit/s uplink	Various
802.11ad/WiGig [32.16]	Single carrier or OFDM	4.6 Gbit/s (single carrier), 7 Gbit/s (OFDM)	60 GHz
5G 3GPP Release 15 [32.17]	Unknown, although candidates include SEFDM [32.18], GFDM [32.19], FBMC [32.20], UFMC [32.20]	> 10 Gbit/s	Lower bands 3300–4200 MHz and 4400–4990 MHz
5G example [32.21]	4×4 MIMO, 64 (QAM)	35 Gbit/s	73 GHz

relating to linearity and induced distortion [32.27, 28]. In the research literature, WiFi (IEEE 802.11a/b/g, etc.) systems have been widely used in experimental demonstrations, in part due to the easy availability of access points and license-free spectrum usage. However, due to the low cost of the access points, it is hard to see fiber-based systems displacing electronic systems unless they form part of a wider, converged, multistandard network [32.29, 30].

While early research focused on the performance limits of single-standard systems, most recent research has concentrated on multichannel operation or the extreme performance limits of the systems, e.g., the interaction of second-generation (GSM) and third-generation (W-CDMA) wireless systems [32.31] or of W-CDMA and wireless LAN (WLAN) [32.32]. When multiple standards must be transported, distortion due to the nonlinearity of the optical components is a major consideration [32.33].

Out of all these standards, the one that is currently experiencing the most active research is that of 5G, which is—at the time of writing—the next mobile generation, so it is currently being researched, standardized, and trialed worldwide. There are very few confirmed details, with the exact frequency bands and bit rates to be delivered by 5G systems still to be decided. The key standards body in this area, the 3GPP [32.15], turned its attention to the first set of standards (called 3GPP Release 15) in the second half of 2017, with those standards released in September 2018 [32.34]. However, it is expected that 5G will demand the following features over and above those already offered by 4G systems [32.35]:

- 1000× higher data volume per unit area
- 100× more connecting devices
- 10× longer battery life
- 5× reduced latency.

These will all put significant pressure on the RAN due to the required increase in overall capacity as well as the likelihood of a significant reduction in the cell size (and therefore an increase in the number of cells) required to support such a high bandwidth air interface. It is also inevitable that increased use of techniques such as multiple input multiple output (MIMO) and cooperative multipoint (CoMP) will be required to further enhance coverage.

Furthermore, it is expected that frequency bands in the millimeter-wave region will have to be exploited.

Table 32.2 Frequency plan, maximum EIRP, and maximum transit power in the 60 GHz V band for various countries

Region	Frequency range (GHz)	Bandwidth (GHz)	Maximum EIRP (dBm)	Maximum transmit power (dBm)
US [32.36]/ Canada [32.37]	57–64	7.0	43.0 (peak), 40.0 (average)	27.0 (for bandwidth $\geq$ 100 MHz)
Europe [32.38]	57–66	9.0	40.0 (indoor only), 25.0 (indoor and outdoor)	Not defined
Australia [32.39]	57–66	9.0	43.0 (indoor only)	13.0
Japan [32.40]	59–66	7.0	57.0	10.0

Table 32.3 Frequency plan, maximum EIRP, and maximum transit power in the E band for various countries

Region	Frequency range (GHz)	Use	Maximum EIRP	Maximum transmit power
FCC [32.41]	71–76, 81–86, and 92–95	FDD or TDD	55 dBW, 43 dBi minimum antenna gain	5 dBW
Europe [32.42]	71–76, 81–86	Channel sizes ranging from 250 to 4750 MHz	55 dBW	30 dBm

Among the first bands of interest are the 26 and 28 GHz bands, which are already available to telecoms operators in many jurisdictions for fixed-wireless access systems [32.43]. Beyond this, others are advocating the use of the spectrum at both 37 and 39 GHz. The 60 GHz frequency range (often termed the *V band*) is subject to different regulations and frequency-band structures in different countries but is typically unlicensed. For some, this makes the V band attractive; however, operators typically prefer the safety of licensed bands, where usage is more tightly controlled. The 60 GHz frequency band plan, limits on the transmit power, and the maximum equivalent isotropically radiated power (EIRP) are shown for various countries in Table 32.2. The EIRP is important as it indicates how directional the antennas will need to be to realize the maximum transmit power. Above this frequency range, some early 5G experiments have looked at the E band, for which three bands are specified: 71–76, 81–86, and 92–95 GHz. The frequency plan, maximum EIRP, and maximum transit power in the E band are provided for various countries in Table 32.3.

32.2.3 Current Applications of Fronthaul Systems

Simplifying the equipment at the cell site has been seen to offer many benefits to mobile operators. It reduces costs associated with site-cell acquisition, installation, rental, and maintenance, allowing for a simpler plan and build at the base station that takes advantage of the smaller footprint. This in turn leads to lower energy consumption by reducing the quantity of electronics and air conditioning/environmental control needed at the cell site. Once pooled at a central location, the baseband processing can be virtualized. This allows for reduced hardware costs and simplified signaling, enabling improved deployment of advanced enhancements such as cooperative multipoint (CoMP) and interference coordination, which require high signaling data rates with low latency and high synchronization.

However, there are some issues that must be faced. In current digital solutions, the main issue is the bandwidth bloat that arises due to the digitization process and the associated protocols. For example, a current LTE-FDD (Long-Term Evolution frequency-division duplex) system that uses 2×20 MHz bands upstream and downstream will offer a peak data rate to customers of around 150 Mbit/s. However, to fronthaul this signal, the CPRI interface requirement is for 2.5 Gbit/s per sector. This greater than 15-fold bandwidth increase is common in digital fronthaul systems, and is likely to be the limiting factor in 5G given that estimates of 5G capacity suggest that it might drive the fronthaul requirement to > 500 Gbit/s if current techniques are maintained. This is hard to address, with data-rate requirements scaling linearly with increases in RF bandwidth and the number of sectors. It also places pressure on the number of fibers required or calls for more expensive WDM technologies to be used. In addition, although many have dreamed of very-low-cost systems that take advantage of high-volume (and therefore low-cost) ethernet equipment, the very tight timing requirements make the transport of CPRI over native ethernet impractical. Although work aimed at addressing all of these issues (e.g., by using CPRI compression techniques [32.44] or by directly using ethernet) is being carried out, it should be noted that delay and synchronization issues are still to be solved [32.45].

32.3 Analog Radio Over Fiber (A-RoF)

In the most basic optical link, shown in Fig. 32.4, a laser diode is biased in the linear region of its transfer characteristics and then directly modulated by the RF signal to be transported. This laser is then directly coupled into an optical fiber for transmission across some distance. At the receiving end, the light is detected using a biased photodiode that converts it back to an electrical form.

32.3.1 Microwave Signals on Optical Fibers

For frequencies of < 10 GHz, direct modulation is typically used (although some directly modulated lasers do exist above this [32.46]). Above this frequency, and if high levels of linearity are needed, external modulators based on either electro-optical effects (such as the Mach–Zehnder modulator) or electroabsorption-based modulators are preferred or required, as shown in Fig. 32.5. These modulators provide superior link performance, albeit at the expense of additional cost and size.

In Fig. 32.5, following the signal in the downstream direction (from the central office to the antenna), we see that the majority of the electronic processing—i.e., the baseband processing (framing, etc.) and the digital-to-analog conversion—take place at this central office. The signal is then upconverted to the required radiofrequency in the section marked RF. In some cases, an intermediate frequency (IF) may be transported. This reduces the frequency requirement of the optical components and link but introduces a requirement for further upconversion at the base station. As we will see later, this trade-off many be worthwhile at very high frequencies. The electrical signal is fed, possibly via an amplifier or driver circuit, to an optical modulator. Unlike for digital modulation, as the signal is analog and bipolar, the optical modulator is usually biased at its quadrature point (i.e., in the middle of its transfer function). The drive will need to be adjusted so that the maximum signal excursions remain within the linear region of the transfer function. The modulator imparts the RF signal onto a continuous-wave (CW) laser source, and the signal is then coupled into the optical fiber via either a wavelength-selective coupler or a circulator. Typically, this downstream wavelength is chosen to be different to the upstream wavelength to form a wavelength division duplex pair, thus avoiding crosstalk between the upstream and the downstream if any reflections are present in the link. At the end of the link, again using either a wavelength-selective coupler or a circulator, the signal is delivered to a photodiode of sufficient bandwidth to accommodate the RF modulation frequency. At this stage, the only processing that is required is power amplification (PA) before the RF signal is fed, via a diplexer, to the antenna for transmission.

In the reverse (upstream) direction, the signal collected from the antenna is first amplified using a low-noise amplifier (LNA) before being fed to a similar optical modulator to that used in the downstream direction. The main differences from the downstream direction are the wavelength of operation, as explained above, and the fact that the RF drive power received from the antenna is considerably lower than that produced by the RF equipment in the central office (which drives the link in the opposite direction). This typically creates the requirement for high-gain electronics at the base-station site.

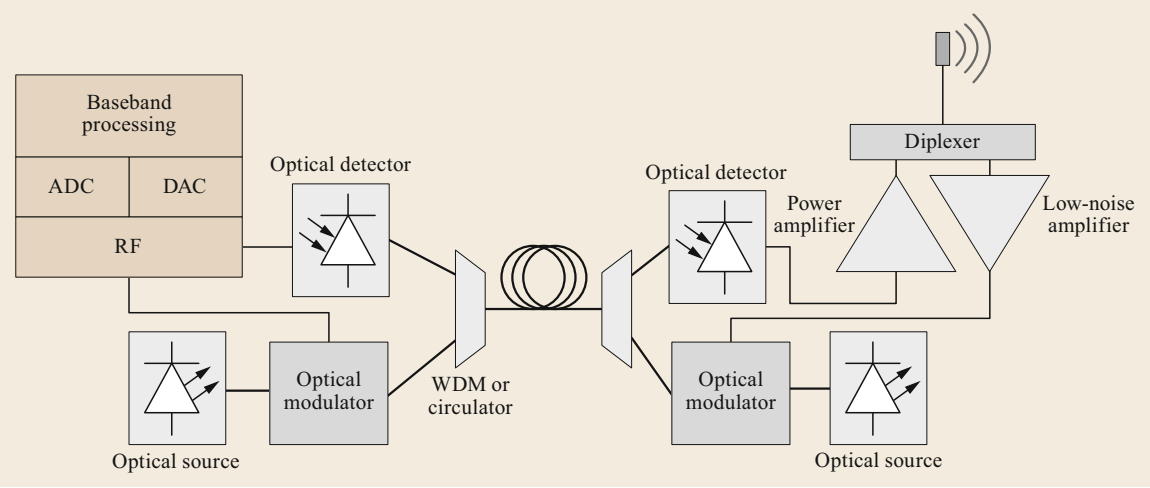


Fig. 32.5 Example of an externally modulated analog fiber link

A key issue in the transport of analog signals is the linearity of the transfer function of the electrical to optical device. The most common (and therefore inexpensive) devices are designed for digital systems where on-off keying dominates and hence linearity is less of an issue. Therefore, to achieve reasonable performance in terms of distortion, a low modulation depth is used at the laser or modulator to maintain operation in the linear region. If increased linearity is required, techniques to linearize the devices can be applied through either device design itself or external linearization techniques. This has been demonstrated for directly modulated devices [32.47, 48] as well as linearized external modulators [32.49], which are commonly used in cable TV (CaTV) networks. To enable the use of low-cost sources, more advanced techniques such as feedforward linearization have also been proposed [32.50].

In addition, an important performance measure in an analog link (although one often ignored in digital systems) is the relative intensity noise (RIN). Effects such as mode instability or spontaneous emission present in the laser diode will introduce some unwanted intensity variations that are usually characterized via the RIN parameter. The RINs of commercially available semiconductor laser diodes are typically in the range of $130-160\,\mathrm{dBHz}^{-1}$. The RIN peaks in the vicinity of the laser relaxation oscillation frequency and is a function of the modulating signal frequency and the bias current. In a directly modulated link, there is a trade-off between maintaining sufficient optical power at the photodiode to overcome the receiver thermal noise and minimizing the effect of RIN, which increases with the square of the optical power [32.51, p. 182]. The reflection of optical power back into the laser also causes an increase in RIN, so it must be tightly controlled in any system implementation.

To achieve the maximum link gain, appropriate impedance matching of the laser and photodiode is required, as both typically have quite low impedances (typically on the order of a few ohms). In certain circumstances, with correct matching, positive link gain can be achieved [32.52]. An excellent review of the state of the art of analog links can be found in [32.53], while those looking for a rigorous analysis of the design of such links should read [32.51].

32.3.2 Intensity Modulation Direct Detection of RF Signals

Making use of the basic structure shown in Fig. 32.4, full-duplex point-to-point links can be formed. For frequencies below around 6 GHz, which covers most of the current cellular and WiFi bands, this is typically achieved by repeating the link shown in Fig. 32.5. Although the loss in an optical fiber is approximately independent of the radio frequency, chromatic dispersion begins to have a significant influence on the transmission of a standard intensity modulation direct detection (IM-DD) signal for frequencies above around 10 GHz. To understand this better, consider a standard single-mode fiber with a chromatic dispersion $D = 17\,(\mathrm{ps/nm})\mathrm{km}$ at $\lambda = 1550\,\mathrm{nm}$. If we use a continuous-wave (CW) single-mode laser that is externally modulated with an ideal Mach–Zehnder modulator (MZM) biased for linear intensity, we see that the optical power spectrum of the signal produced is

$$S(f) = \frac{P_\mathrm{o}}{2}\left\{\sum_{n=-\infty}^{\infty} J_{2n+1}^2(m)\delta \times [f - f_\mathrm{o} - (2n+1)f_\mathrm{mm}]\right\},$$

where f_mm is the required mm-wave frequency, f_o is the mean optical frequency, P_o is the total optical power, J_n represents the Bessel function of the first kind of order n, and m is the modulation index, which is calculated as the ratio of the input voltage range to the switching voltage of the modulator (i.e., the voltage required to turn the modulator from on to off). From this, we can deduce that the RF signal is formed by both the lower and an upper side band ($n = -1$ and $n = 1$), which are offset by the frequency of the signal from the optical carrier ($n = 0$). This is known as a *three-term technique* because of the presence of three main terms in the optical spectrum (optical carrier and two sidebands). When transported across a fiber with chromatic dispersion, a differential delay is induced between the two sidebands which, when detected on a square-law photodetector, leads to an interference pattern between the terms generated by the beating of the upper side band with the carrier and the lower side band with the carrier. Therefore, for maximum power transfer, we require that the terms are in phase, which will only happen for certain delay values. For all other delays, there will be a reduction in the total RF power and therefore an associated reduction in the carrier-to-noise ratio of the resulting signal.

If the RF carrier is an unmodulated sinusoid, complete extinction of the received mm-wave carrier occurs in a cyclic fashion whenever the phase difference between the lower and upper sideband is π, resulting in repetitive, link-length-dependent nulls in the detected signal power. The first null occurs at a fiber length l (in m), given by

$$l = \frac{c}{2D\lambda^2 f^2},$$

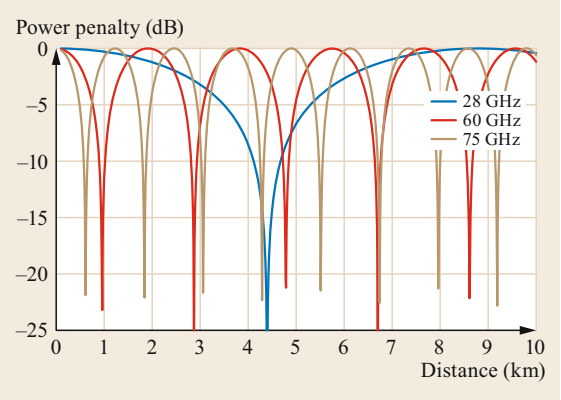


Fig. 32.6 Length-dependent fading of radiofrequency signals in an optical fiber

where D is the dispersion parameter, c is the vacuum velocity of light, λ is the mean source wavelength, and f is the modulation frequency [32.54]. As demonstrated in Fig. 32.6, at mm-wave frequencies, the transmission distances possible using three-term techniques are severely limited. This figure plots the theoretical results for a 1550 nm system operating with a standard single-mode fiber dispersion of 18 (ps/nm)km (fiber loss is neglected). It shows that the first power null occurs at around 4 km at 28 GHz and around 1 km at 60 GHz. As the period of the nulls decreases with the square of the frequency, shifting the carrier to higher mm-wave bands rapidly reduces the period.

Although reduced-dispersion fiber would be one solution to the problem, most research work has focused on developing techniques that generate two-term rather than three-term signals. In two-term techniques, the resultant RF signal at the photodiode is the product of the beating of only the carrier with a single sideband, so any dispersion-induced phase shift produces only a relative phase shift of the RF signal, which is of no significance to the final signal. There are a number of techniques that will generate two-term signals. A summary of some of the most common is given below.

As we have seen above, when operated in its linear region, the Mach–Zehnder modulator gives a three-term output. However, modifications to the drive of the modulator or the design can be used to generate two-term signals.

Single-Sideband Modulation

If a dual-arm (sometimes called a dual-drive) modulator (i.e., a modulator where modulation can be applied independently to both arms of the Mach–Zehnder structure) is used with each arm driven by signals that are identical but 90° out of phase, a single-sideband spectrum is produced at the output [32.55]. This is typically carried out using a hybrid coupler that creates in-phase and out-of-phase signals from a single RF input, as shown in Fig. 32.7 [32.56], although dual-parallel Mach–Zehnder modulators (MZMs) can also be used [32.57]. A single-sideband signal may also be generated by combining an external modulator and a phase modulator in a cascade [32.58], or using a monolithically integrated DFB laser with a multimode interferometer [32.59]. As shown in Fig. 32.7, the optical output of the single-sideband generator contains that carrier term plus one of the two sidebands that is modulated with the data at either the baseband or an intermediate frequency. At the photodiode, the mixing of the carrier with this single sideband results in a RF signal at the required frequency. The presence of only one sideband results in a 3 dB power reduction compared to the double-sideband signal, but the signal now has the advantage of being tolerant to chromatic-dispersion-induced fading.

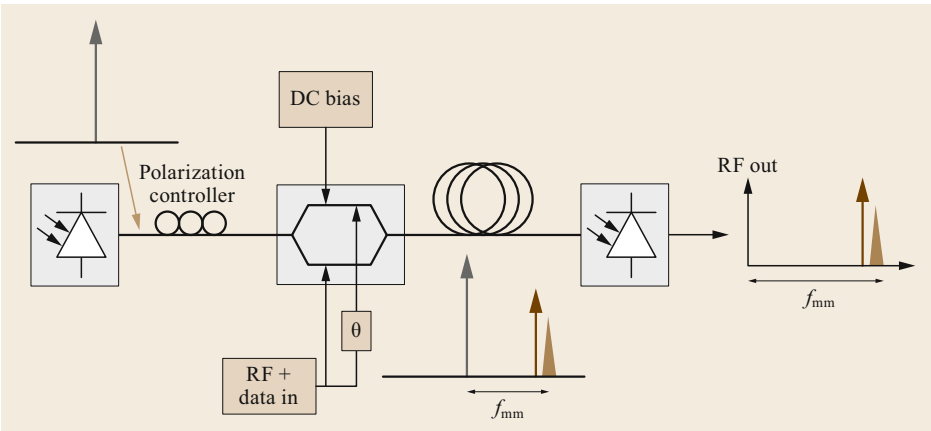


Fig. 32.7 Optical single-sideband setup

Double-Sideband Suppressed Carrier Modulation

An alternative option is to bias a standard Mach–Zehnder modulator at its null point [32.60] to give a double-sideband suppressed carrier signal, as shown in Fig. 32.8. Here, the two sidebands are present, with the optical carrier suppressed between them. One advantage of this technique is that a RF drive of only half the final transmission frequency is needed. It is also possible to bias the device at its maximum transmission point to give two tones at four times the frequency plus a carrier [32.61]. The main disadvantage is that due to the nonlinear nature of the modulation, data cannot be imposed with the same modulator as the RF tone in either of these techniques. In the example of Fig. 32.8, an optical fiber Bragg grating (FBG) is used to filter out either of the two sidebands, thereby selecting one of two frequencies for modulation with a data signal [32.62] using a separate modulator. Despite this requirement for an additional (albeit lower-speed) second modulator, this approach does have the advantage that a number of wavelengths can be upconverted using a single high-frequency device [32.63, 64].

Optical Filtering

Some have proposed the use of optical filtering with a device such as a fiber Bragg grating to remove one of the sidebands of a three-term technique. This technique is generally not preferred as it is not very power efficient [32.65] and requires high stability of both the grating and the transmitting laser.

32.3.3 RoF System Architectures

There are a number of candidate architectures that can be used to apply the techniques introduced in the section above to a system that delivers radio signals to users. Different implementations are developed to reduce the cost of the remote antenna unit by reducing the component count or by decreasing the number of high-frequency components (particularly high-frequency optical components). However, this usually comes at the expense of some flexibility.

RF Transmission Techniques

The architecture shown in Fig. 32.9 is the simplest configuration in that it transmits the signal at its final transmission frequency in both directions. At low frequencies, this can be achieved using intensity modulation direct detection (IM-DD) techniques with direct modulation of a laser diode or similar at the transmitter. However, as discussed above, two-term techniques such as those described in Sect. 32.2.1 will need to be deployed at higher frequencies (> 10 GHz). One issue with such a design is the need for the optics in the RRU to replicate those used at the central site. Ideally, we

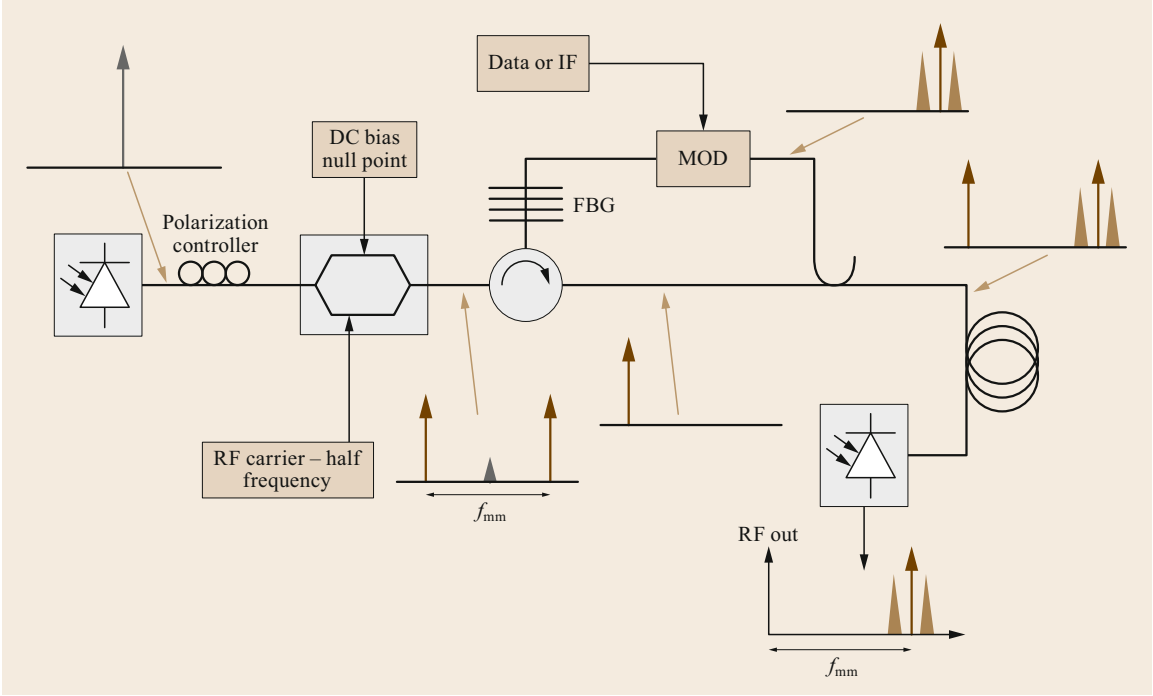


Fig. 32.8 Double-sideband suppressed carrier setup

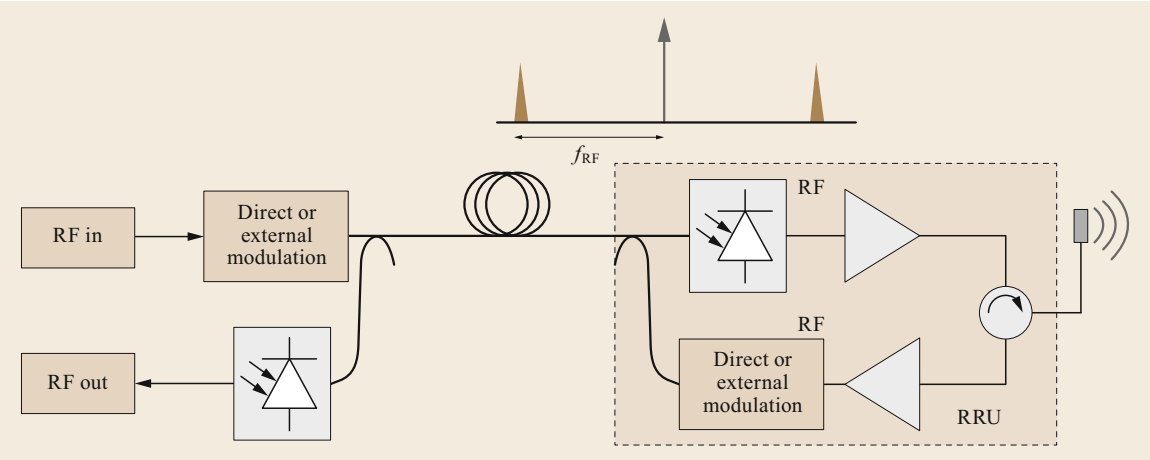


Fig. 32.9 RF downstream–RF upstream configuration

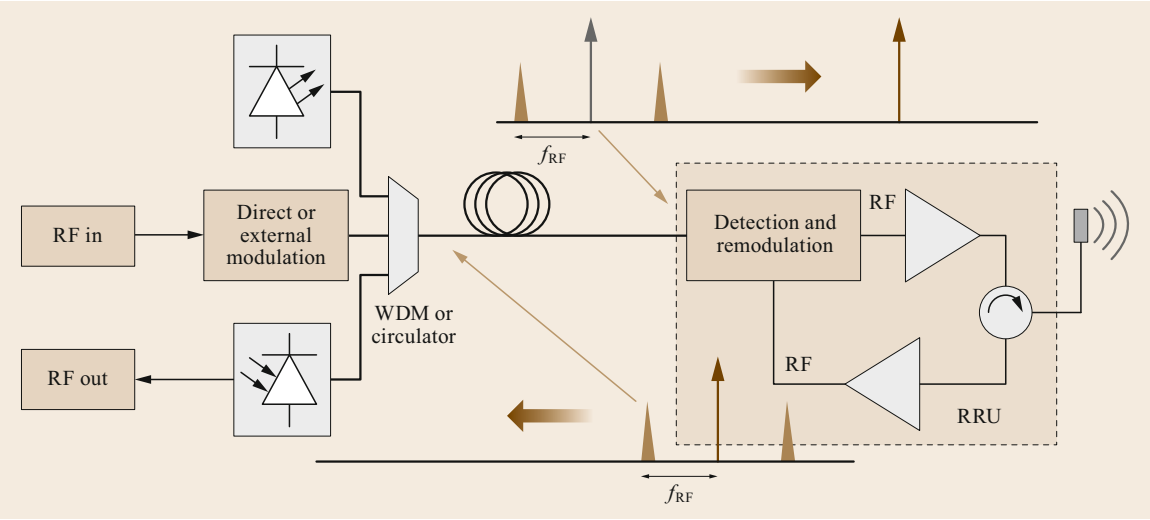


Fig. 32.10 RF downstream with loopback–RF upstream configuration

would like to simplify the RRU and move as much complexity to the central site as possible. One option is to remove the need for an optical source at the RRU by remotely generating the return carrier. This is commonly known as a loopback scheme. As shown in Fig. 32.10, two wavelengths are produced at the central site. One carries the downstream data (as in the standard case), while the second is an unmodulated light signal that will be used to carry the return signal. It is here that optical devices that are capable of both receiving and modulating an optical signal become attractive [32.4, 5]. This allows the RRU to be simplified, but may require dispersion compensation due to the limitation of IM-DD in the upstream (RRU to central office) direction. It is also possible for the optical carrier from a modulated downlink signal to be filtered and reused as the optical source on the uplink [32.66], although this technique may require the use of separate fibers for the downstream and upstream links.

IF Transmission Techniques

As the frequency of operation increases, the cost of the optical components that need to operate at high frequencies (typically the modulator and photodiode) can become the limiting factor in the cost of the RRU. As an alternative, schemes have been proposed that transport an intermediate frequency (IF) rather than the final transmission frequency over the fiber network. Such systems are usually deployed at high microwave or mm-wave frequencies using an IF of a few GHz where dispersion effects are minimal and direct modulation of lasers is possible. These techniques have been

demonstrated for a wide range of final transmission frequencies, including 60, 75 GHz, and even terahertz transmission. A number of mechanisms to up/downconvert the IF signal to the final frequency are possible. These are discussed below.

Electrical Upconversion. To upconvert the signals at the RRU, a stable local oscillator (LO) at around the carrier frequency ($\pm f_{RF}$) is required. In the configuration shown in Fig. 32.11, the LO is provided locally at the RRU. This has design implications in terms of the requirement for a temperature-stable LO source, and it reduces the range of frequencies available at the node.

To reduce the component requirement, and especially the need for a high-stability oscillator at the RRU, the LO can be distributed optically from the central office at either the required frequency or as a subharmonic to be used for subharmonic injection locking of an oscillator [32.40] or as a reference for a phase-locked loop (PLL) RF oscillator [32.67]. For a network with multiple nodes, a single LO can be distributed to a number of nodes, thereby sharing the high-frequency components used to generate the LO. For example, in [32.68], a single 40 GHz signal is distributed around a ring to a number of base transceiver stations. In this example, each node also receives an individual optical channel carrying a 2.4 GHz IF signal, which it combines with the 40 GHz oscillator to form the final channel. In this particular configuration, the LO is also used to downconvert the received signals at the RRU to allow the use of low-frequency optical devices for the return channel as well. Such a configuration is demonstrated in Fig. 32.12.

32.3.4 Optical Upconversion: Coherent Radio-Over-Fiber Techniques

Most recently, spurred on by the success of coherent techniques in baseband core networks, there has been renewed interest in a coherent system for radio over fiber. Although there are many parallels, there is one major problem. A typical RoF system rarely has the access needed to process the data itself at baseband—it is, in fact, transporting 'waveforms' for other services. This means that the types of digital signal processing that have enabled high-bit-rate coherent systems cannot be performed in the same way on the analog (as far as the RoF system is concerned) signals that the RoF system is transporting. However, some coherent operations are still possible.

The basic concept of a coherent system is the heterodyning of two laser sources, one of which is carrying data at an IF, to create a resultant RF signal at the required mm-wave frequency. To do this, the two laser sources used must be separated by the radiofrequency that is to be generated. As the laser linewidths of most sources are large compared to the signal bandwidths, free-running operation is not possible without significant impairment. This comes about because of the phase noise inherent to the laser, which manifests itself in the broadening of the spectral linewidth of the laser source. When two uncorrelated optical

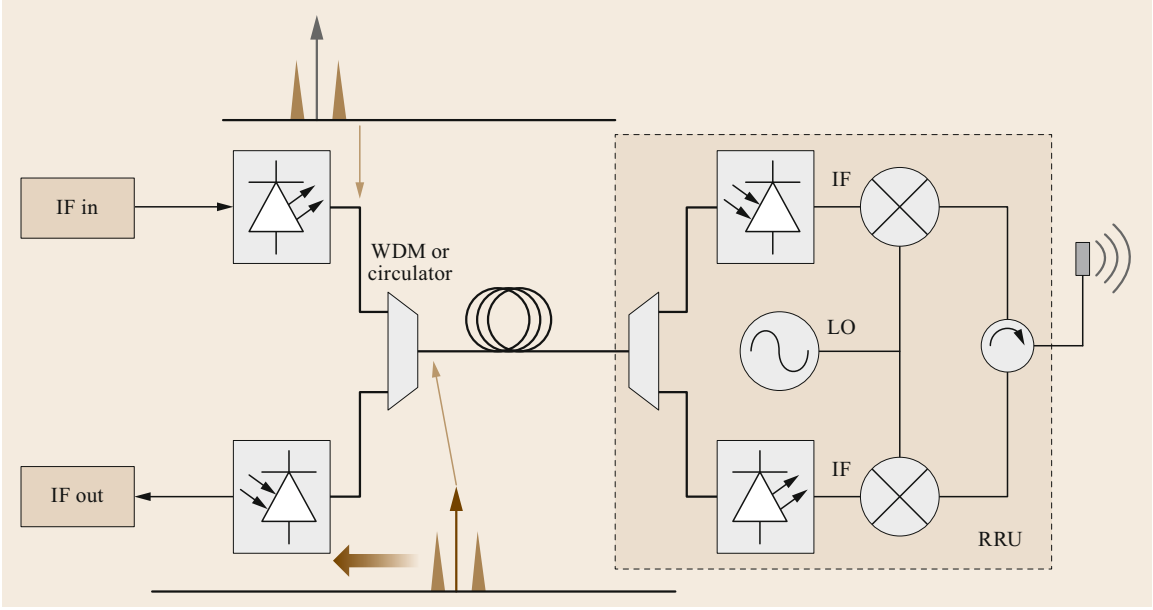


Fig. 32.11 IF downstream–IF upstream configuration

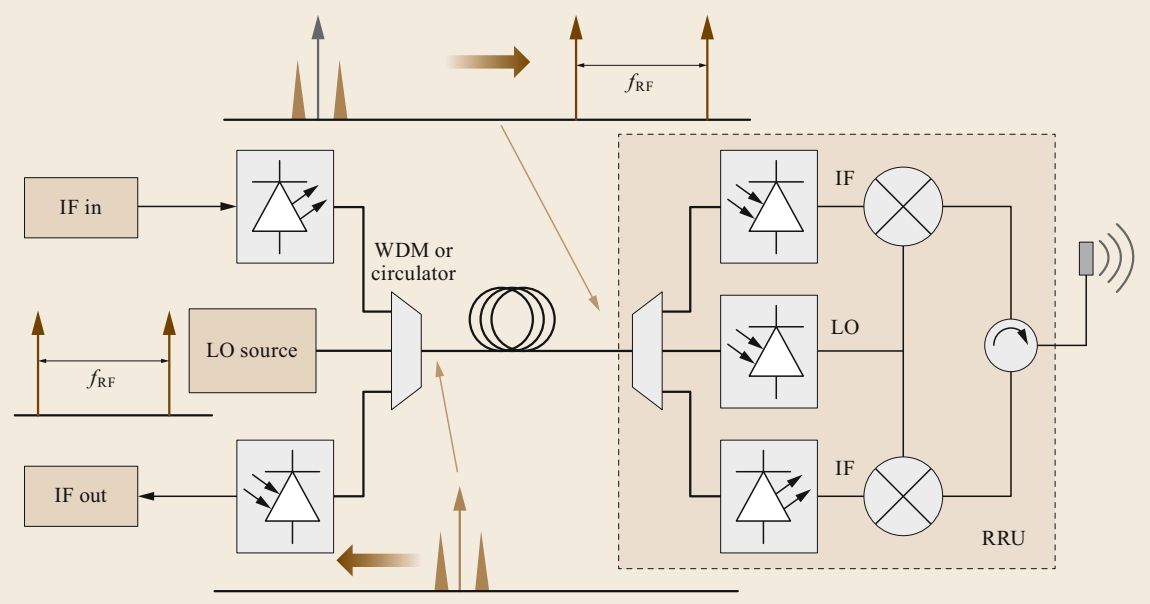


Fig. 32.12 IF and LO downstream–IF upstream configuration

sources beat together on a photodetector, this phase noise significantly degrades the detection performance of the transmission system. To overcome this, techniques to lock the two lasers together are usually required. These may take the form of an optical frequency locked loop (OFLL) [32.69], an optical phase locked loop (OPLL) [32.70], optical feedforward modulation [32.71], optical injection locking [32.72], or optical injection phase-locking [32.73].

There is, however, one configuration where two uncorrelated sources can be used for all-optical upconversion [32.74]. Such a system uses uncorrelated optical heterodyne mixing with a free-running CW laser to generate a high-frequency signal after a broadband photodiode. In contrast to the numerous heterodyne mixing implementations reported in the literature, which require optical injection locking or complex optical phase-locked loops, this proposed technique utilizes envelope detection to avoid laser phase noise effects at the baseband, thus obviating signal control techniques. This results in a very simple and cost-effective transmission system.

It has been shown [32.71] that if the data are on an appropriate IF, and a square-law envelope detector is used to downconvert the mm-wave wireless at the receiver, the phase noise penalty can be overcome. Envelope detection is preferred over self-homodyne reception because it is cheaper and has been shown to provide better sensitivity [32.75]. This allows for a much-simplified RRU, with the radiofrequency being generated by controlling the optical frequency different between the two optical tones. Such a system provides a very wide tuning range, although it does also impose some significant requirements on the frequency stability of the lasers.

This coherent reception technique is shown in Fig. 32.13, where one laser is used as the optical input to a low-bandwidth dual-drive Mach–Zehnder modulator (MZM) that is modulated with the IF signal. The phase fluctuations $\varphi_1(t)$ of this laser can be characterized as a Wiener–Lévy process with zero mean and a variance of $2\pi\beta|t|$, where β is the two-sided 3 dB laser linewidth of the Lorentzian power density spectrum [32.76]. At the output of the MZM, we can describe the optical field as

$$E_1(t) = \mathrm{e}^{\mathrm{j}[2\pi f_{\mathrm{L}1}t+\varphi_1(t)]} + m(t)\mathrm{e}^{\mathrm{j}[2\pi(f_{\mathrm{L}1}+f_{\mathrm{IF}})t+\varphi_1(t)]} ,$$

where $f_{\mathrm{L}1}$ and f_{IF} are the center frequency of the signal laser and the intermediate frequency, respectively, and $m(t)$ is the data carried on the IF. The output optical field of the second laser (the LO laser) with laser phase noise $\mathrm{e}^{\mathrm{j}\varphi_2(t)}$ can be expressed as

$$E_2(t) = \mathrm{e}^{\mathrm{j}[2\pi f_{\mathrm{L}2}t+\varphi_2(t)]} ,$$

where $f_{\mathrm{L}2}$ is the center frequency of the LO laser. When the two optical signals are combined in the optical coupler, the resulting optical signal, neglecting both the fiber and wireless channels in this analysis, produces a photocurrent $I(t)$ at the output of the square-law pho-

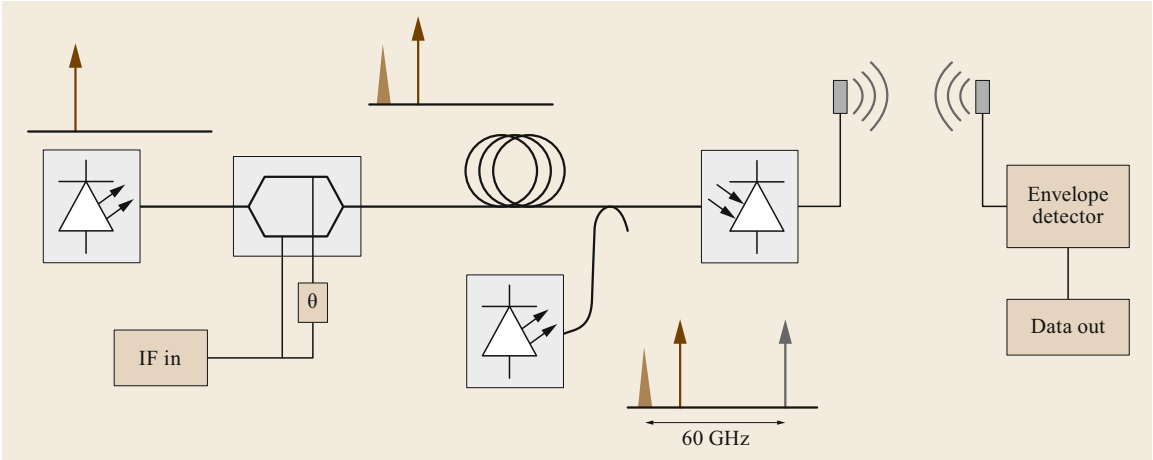


Fig. 32.13 Schematic of the coherent reception technique

todetector. This photocurrent is given by

$$I_1(t) = \mathcal{R}(\ldots + m(t)\cos[2\pi(f_{\mathrm{IF}})t] + 2\cos\{2\pi(f_{\mathrm{L1}} - f_{\mathrm{L2}})t + [\varphi_1(t) - \varphi_2(t)]\} + 2m(t)\cos\{2\pi f_{\mathrm{RF}}t + [\varphi_1(t) - \varphi_2(t)]\}) ,$$

where

$$f_{\mathrm{RF}} = f_{\mathrm{L1}} - f_{\mathrm{L2}} + f_{\mathrm{IF}} ,$$

and where $\mathcal{R}$ is the responsivity of the photodetector. Ignoring some direct current (DC) components and second-order intermodulation distortion products, the first term is a copy of the data signal centered at a frequency of f_{IF}. The second term is a RF carrier at a frequency of $(f_{\mathrm{L1}} - f_{\mathrm{L2}})$ that is obtained from the beating of the two optical carriers with themselves, while the final term is the desired message signal at a RF of f_{RF}. It is seen that the last two terms have random phase deviations of $[\varphi_1(t) - \varphi_2(t)]$. If $\varphi_1(t)$ and $\varphi_2(t)$ have spectral densities represented by η_1 and η_2, respectively, then it can be calculated that the last two terms will have spectral densities of $(\eta_1 + \eta_2)$ Hz. This will result in a 3 dB linewidth equal to the sum of the linewidths of the signal and LO lasers. If we assume that the signals are downconverted by passing them through an ideal square-law envelope detector, the downconverted signal $I_2(t)$ can be written as

$$I_2(t) = 4\mathcal{R}^2(\cos\{2\pi(f_{\mathrm{L1}} - f_{\mathrm{L2}})t + [\varphi_1(t) - \varphi_2(t)]\} + m(t)\cos\{2\pi f_{\mathrm{RF}}t + [\varphi_1(t) - \varphi_2(t)]\})^2 .$$

After filtering $I_2(t)$ with a low-pass filter to eliminate the higher-frequency components, only the baseband signal

$$I_3(t) = 4\mathcal{R}^2\{m(t)\cos[2\pi(f_{\mathrm{IF}})t]\}$$

remains. This result demonstrates that there are no laser phase noise terms, implying that the proposed technique is both capable of mitigating laser phase noise effects and robust against chromatic dispersion. However, it should be noted that this technique requires good long-term stability of the optical sources, which may require additional circuitry to give the necessary wavelength stability and accuracy. The requirement is that the wavelength drift present is not large enough to shift the received signal outside the bandwidth of the envelope detector. In the demonstration of this technique described in [32.74], both optical sources were tunable lasers that utilized thermoelectric cooler (TEC) modules to ensure wavelength stability and accuracy throughout their tuning ranges. Using such lasers guarantees that the beat term falls within the 60 GHz band. More recently, such an approach has been demonstrated in the 60 GHz band with 512-QAM-OFDM [32.77]. This produced a data transmission of 21 Gbit/s. A specially developed coherent photonic mixer was used as the optical receiver.

There are a number of systems in the literature that make use of the term *coherent*, usually because they borrow from the baseband coherent techniques that are now ubiquitous in 40 and 100 Gbit/s systems. One of the first demonstrations of a coherent system was by *Kitayama* et al. in 2012 [32.78]. In this technique, a digital baseband data signal is modulated onto one tone of a coherent two-tone optical generator. In this case, the modulating signal is a complex QPSK signal of 20 GBd. After transmission across a fiber network and reception with a fast photodiode, the signal undergoes wireless transmission at 92.5 GHz. On reception, the signal is downconverted to an IF and sampled using a high-speed analog-to-digital converter (ADC). It is here that the system borrows from baseband tech-

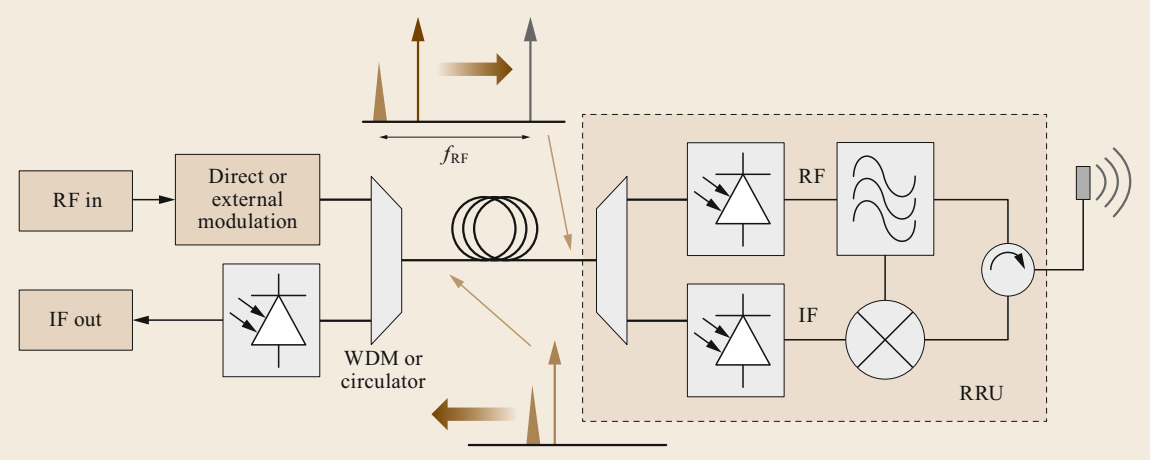


Fig. 32.14 RF downstream–IF upstream configuration

niques, as it utilizes many of the same signal processing methods. This work has since been further extended to demonstrate the QPSK modulation of two wireless polarizations to produce a mm-wave MIMO RoF system at 90 GHz that delivers 74.4 Gbit/s transmission [32.79].

32.3.5 Hybrid Techniques

For some applications, a hybrid RF down–IF up technique can be applied as an alternative. As shown in Fig. 32.14, the downstream (to the RRU) signal is carried using a RF-over-fiber technique. At the RRU, a RF component of this signal is tapped off and used as the LO signal to downconvert the upstream signal to an IF for transmission using lower-speed optics. Utilizing RF transmission in the downlink removes the need for a stable LO source to be placed or derived at the RRU, as would be needed in IF down techniques, while the use of IF upstream offers the all-important cost savings of high-frequency optical components at the RRU. However, it must be noted that careful selection of the up- and downstream frequencies is required, which may not always be possible or may be constrained by the radio standard itself. Because of this, this technique lends itself to use in a system deploying frequency-division duplexing.

32.4 Concluding Remarks

In this chapter, we have reviewed the fronthaul applications and wireless standards that are likely to benefit from analog radio-over-fiber systems, and the architectures that are typically deployed.

With the increasing bandwidths required for 5G systems and the higher-frequency systems that are being deployed, the case for radio-over-fiber systems grows. Although digital techniques currently dominate for fronthaul applications, the increased bandwidth requirements of digital interfaces mean that analog techniques may become favorable for a number of applications as capacity grows. Even though the bandwidth efficiencies of digital techniques will improve as compression and other techniques are introduced, analog techniques have a significant head start in this regard. For baseband transport of processed data from a wireless BS, the total capacity requirement is simply the sum of the total user data available at the cell site once all available channels are considered and including all sectors plus the control and signaling overhead. For a LTE eNodeB, it may be on the order of a few hundred Mbit/s. For CPRI to function, it requires a high-speed and low-latency backhaul connection where digitized signals are transported at one of the specified line rates, which are multiples (1, 2, 4, 5, 8, 10, 16) of 614.4 Mbit/s [32.80]. Detailed knowledge of the radio channels used is needed to accurately calculate the capacity required. However, using 1 Gbit/s per antenna carrier as a rule of thumb [32.81] for a current LTE implementation, we find that the requirement can be significant. Consider a single-sector LTE FDD cell site with two transmit and two receive antennas. If we assume a total bandwidth of 20 MHz, 16 bit sampling, two times oversampling, and an overhead of 20%, the CPRI-defined line bit rate required would be 4.9152 Gbit/s. Introducing sectorization, a capacity of

> 10 Gbit/s may be needed at the uncompressed CPRI interface.

Capacity is not the only issue: there are other tolerances that create additional design constraints on the front/back-haul network. The two most notable are the high bit error rate requirements of 10^{-12} and a low latency compared to FEC-enabled digital links. These are linked in that it is recommended that forward error correction is avoided to reduce the processing delay, which is already on the order of a sub-ms for the round trip (for GSM = 78 km or ≈ 380 μs, 3G = 40 km or ≈ 190 μs, and LTE = 20 km or ≈ 95 μs). In addition, strict requirement on the delay jitter, in the range of tens of nanoseconds, and a frequency jitter of < ±0.002 pm add further issues in digital systems [32.82]. If scaled to 100 MHz channels, which is not unexpected for future systems, bit rates in excess of 24 Gbit/s may well be needed [32.83]. Some are even suggesting that if all of the features of LTE Advanced are enabled, the back-haul traffic requirement may be pushed up to 100 Gbit/s per site [32.84]. This clearly suggests room for analog systems in this space.

References

32.1 OBSAI: Open Base Station Architecture Initiative, http://www.obsai.com/members.htm (2017)
32.2 CPRI: Common Public Radio Interfaces, http://www.cpri.info/ (2017)
32.3 D. Wake, D. Johansson, D.G. Moodie: Passive picocell: a new concept in wireless network infrastructure, Electron. Lett. **33**(5), 404–406 (1997)
32.4 L. Westbrook, D.G. Moodie: Simultaneous bi-directional analogue fibre-optic transmission using an electroabsorption modulator, Electron. Lett. **32**(9), 1806 (1996)
32.5 C.P. Liu, C.H. Chuang, S. Karlsson, O. Kjebon, R. Schatz, Y. Yu, T. Tsegaye, A.B. Krysa, J.S. Roberts, A.J. Seeds: High-speed 1.56-μm multiple quantum well asymmetric Fabry–Perot Modulator/Detector (AFPMD) for radio-over-fibre applications. In: *2005 Int. Top. Meet. Microw. Photonics, Seoul, Korea* (2005) pp. 137–140
32.6 H. Pfrommer, M.A. Piqueras, J. Herrera, V. Polo, A. Martinez, S. Karlsson, O. Kjebon, R. Schatz, Y. Yu, T. Tsegaye, C.P. Liu, C.H. Chuang, A. Enard, F. Van Dijk, A.J. Seeds, J. Marti: Full-duplex DOCSIS/WirelessDOCSIS fiber-radio network employing packaged AFPM-based base-stations, IEEE Photonics Technol. Lett. **18**(2), 406–408 (2006)
32.7 B. Kalantari-Sabet, M. Mjeku, N.J. Gomes, J.E. Mitchell: Performance impairments in single-mode radio-over-fiber systems due to MAC constraints, J. Lightwave Technol. **26**(15), 2540–2548 (2008)
32.8 S.M. Redl, M.K. Weber, M.W. Oliphant: *GSM and Personal Communications Handbook* (Artech House, London 1998)
32.9 The 3rd Generation Partnership Project (3GPP): http://www.3gpp.org/
32.10 IEEE Std 802.11b-1999 (R2003): Higher-speed physical layer extension in the 2.4 GHz band, https://standards.ieee.org/getieee802/download/802.11b-1999.pdf (2003)
32.11 IEEE Std 802.11a-1999: High-speed physical layer in the 5 GHz band, https://standards.ieee.org/getieee802/download/802.11a-1999.pdf (1999)
32.12 IEEE Std 802.11g-2003: Further higher data rate extension in the 2.4 GHz band, https://standards.ieee.org/findstds/standard/802.11g-2003.html (2003)
32.13 IEEE Std 802.11n-2009: Enhancements for Higher Throughput, https://standards.ieee.org/findstds/standard/802.11n-2009.html (2009)
32.14 T. Baykas, C.-S. Sum, Z. Lan, J. Wang, M.A. Rahman, H. Harada, S. Kato: IEEE 802.15.3c: the first IEEE wireless standard for data rates over 1 Gb/s, IEEE Commun. Mag. **49**(7), 114–121 (2011)
32.15 UTRA-UTRAN Long Term Evolution (LTE): www.3gpp.org/technologies/keywords-acronyms/98-lte
32.16 IEEE Std 802.11ad-2012: Enhancements for very high throughput in the 60 GHz band, https://standards.ieee.org/findstds/standard/802.11ad-2012.html (2012)
32.17 OFCOM: Improving consumer access to mobile services at 3.6 to 3.8 GHz (2016)
32.18 I. Kanaras, A. Chorti, M. Rodrigues, I. Darwazeh: Spectrally efficient FDM signals: Bandwidth gain at the expense of receiver complexity. In: *Proc. IEEE ICC* (2009) pp. 1–6
32.19 G. Fettweis, M. Krondorf, S. Bittner: GFDM – Generalized frequency division multiplexing. In: *VTC Spring 2009 – IEEE 69th Veh. Technol. Conf., Barcelona* (2009)
32.20 F. Schaich, T. Wild: Waveform contenders for 5G – OFDM vs. FBMC vs. UFMC. In: *2014 6th Int. Symp. Commun. Control Signal Proc. (ISCCSP), Athens* (2014) pp. 457–460
32.21 C. Reichert: StarHub clocks 35Gbps speeds in 5G trial with Huawei, http://www.zdnet.com/article/starhub-clocks-35gbps-speeds-in-5g-trial-with-huawei/ (2017)
32.22 Commscope: http://www.commscope.com
32.23 Zinwave: http://www.zinwave.com
32.24 LGC Wireless: http://www.lgcwireless.com
32.25 ADC: http://www.adc.com
32.26 Foxcom: http://www.foxcom.com
32.27 Y. Le Guennec, M. Lourdiane, B. Cabon, G. Maury, P. Lombard: Technologies for UWB-over-fiber. In: *IEEE Lasers Electro-Opt. Soc. Ann. Meet* (2006) pp. 518–519
32.28 M.L. Yee, V.H. Pham, Y.X. Guo, L.C. Ong, B. Luo: Performance evaluation of MB-OFDM ultra-wideband

signals over single mode fiber. In: *IEEE Int. Conf. Ultra-Wideband* (2007) pp. 674–677
32.29 M.L. Yee, H.L. Chung, P.K. Tang, L.C. Ong, B. Luo, M.T. Zhou, Z. Shao, M. Fujise: Radio-over-fiber EVM measurements for IEEE 802.11g WLAN and cellular signal distribution. In: *Eur. Microw. Conf* (2006) pp. 882–885
32.30 M.J. Crisp, L. Shen, A. Wonfor, R.V. Penty, I.H. White: Demonstration of a radio over fibre distributed antenna network for combined in-building WLAN and 3G coverage. In: *Proc. Opt. Fiber Commun. Conf. OFC* (2007)
32.31 R.E. Schuh: Hybrid fiber radio for second and third generation wireless systems, Int. Top. Meet. Microw. Photon. **1**, 213–216 (1999)
32.32 R. Yuen, X.N. Fernando: Enhanced wireless Hhtspot downlink supporting IEEE802.11 and WCDMA. In: *IEEE Int. Symp. Personal Indoor Mobile Radio Commun.* (2006)
32.33 L. Roselli, V. Borgioni, F. Zepparelli, F. Ambrosi, M. Comez, P. Faccin, A. Casini: Analog laser predistortion for multiservice radio-over-fiber systems, J. Lightwave Technol. **21**(5), 1211–1223 (2003)
32.34 3GPP: http://www.3gpp.org/specifications/releases (2017)
32.35 5G: 5g.co.uk/guides/what-is-5g/ (2017)
32.36 Federal Communications Commission: Operation within the band 57–64 GHz, Rep. FCC 15.255 (2002)
32.37 Industry Canada: Low-power licence-exempt radiocommunication devices (all frequency bands): Category I Equipment, IC RSS-210 (2007)
32.38 Australian Communications and Media Authority: Radiocommunications (low interference potential devices) class licence 2000, ComLaw F2011C00543 (2011)
32.39 Association of Radio Industries and Businesses: Millimeter-wave data transmission equipment for specified low power radio station (ultra-high-speed wireless LAN system), ARIB, STD-T74 v1.1 (2005)
32.40 G.H. Smith, D. Novak: Broadband millimetre wave fibre radio network incorporating remote up/down conversion. In: *IEEE MTT-S Int. Microw. Symp. Dig*, Vol. 3 (1998) pp. 1509–1512
32.41 FCC-Memorandum Opinion and Order on Reconsideration: https://apps.fcc.gov/edocs_public/attachmatch/FCC-05-45A1.pdf (2017)
32.42 ETSI EN 302 217-3: Fixed Radio Systems: Characteristics and requirements for point-to-point equipment and antennas, Part 3: Equipment operating in frequency bands where both frequency coordinated or uncoordinated deployment might be applied, Harmonized EN covering the essential requirements of article 3.2 of the R&TTE Directive (2017)
32.43 Y. Azar, G.N. Wong, K. Wang, R. Mayzus, J.K. Schulz, H. Zhao, F. Gutierrez, D.D. Hwang, T.S. Rappaport: 28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in New York City. In: *IEEE Int. Conf. Commun. (ICC), Budapest* (2013) pp. 5143–5147
32.44 B. Guo, W. Cao, A. Tao, D. Samardzija: CPRI compression transport for LTE and LTE-A signal in C-RAN. In: *7th International ICST Conf. Commun. Netw. China (CHINACOM)* (2012) pp. 843–849, https://doi.org/10.1109/ChinaCom.2012.6417602
32.45 N.J. Gomes, P. Chanclou, P. Turnbull, A. Magee, V. Jungnickel: Fronthaul evolution: from CPRI to Ethernet, Opt. Fiber Technol. **26**, 50–58 (2015)
32.46 A. Kaszubowska, P. Anandarajah, L.P. Barry: Improved performance of a hybrid radio/fiber system using a directly modulated laser transmitter with external injection, IEEE Photon. Technol. Lett. **14**(2), 233–235 (2002)
32.47 K. Asatani: Nonlinearity and its compensation of semiconductor laser diodes for analog intensity modulation systems, IEEE Trans. Commun. **28**(2), 297–300 (1980)
32.48 H. Lin, Y. Kao: Nonlinear distortion and compensations of DFB laser diode in AM-VSB lightwave CATV applications, J. Lightwave Technol. **14**, 2567–2574 (1996)
32.49 G.E. Betts, F.J. O'Donnell: Optical analog link using a linearized modulator. In: *IEEE Lasers Electro-Opt. Soc. Annu. Meet.* (1994) pp. 278–279
32.50 T. Ismail, C.-P. Liu, J.E. Mitchell, A.J. Seeds: High-dynamic-range wireless-over-fiber link using feedforward linearization, J. Lightwave Technol. **25**(11), 3274–3282 (2007)
32.51 C.H. Cox: *Analog Optical Links: Theory and Practice* (Cambridge Univ. Press, Cambridge 2004)
32.52 G. Betts, L.M. Johnson, C.H. Cox, S.D. Lowney: High-performance optical analog link using external modulator, IEEE Photon. Technol. Lett. **1**(11), 404–406 (1989)
32.53 C.H. Cox, E.I. Ackerman, G.E. Betts, J.L. Prince: Limits on the performance of RF-over-fiber links and their impact on device design, IEEE Trans. Microw. Theory Tech. **54**(2), 906–920 (2006)
32.54 U. Gliese, S. Norskov, T.N. Nielsen: Chromatic dispersion in fibre optic microwave and millimetre wave links, IEEE Trans. Microw. Theory Tech. **44**(10), 1716–1724 (1996)
32.55 G.H. Smith, D. Novak, A. Ahmed: Overcoming chromatic dispersion effects in fibre-wireless systems incorporating external modulators, IEEE Trans. Microw. Theory Tech. **45**(8), 1410–1415 (1997)
32.56 G.H. Smith, D. Novak, Z. Ahmed: Overcoming chromatic-dispersion effects in fiber-wireless systems incorporating external modulators, IEEE Transact. Microw. Theory Tech. **45**(8), 1410–1415 (1997)
32.57 A. Loayssa, R. Hernandez, D. Benito: Optical single-side band modulators and their applications, Fiber Integr. Opt. **23**(23), 171–188 (2004)
32.58 J. Conradi, B. Davies, M. Sieben, D. Dodds, S. Walklin: Optical Single Sideband (OSSB) transmission for dispersion avoidance and electrical dispersion compensation in microwave subcarrier and baseband digital systems. In: *Opt. Fiber Commun. Conf. (OFC '97) Dallas* (1997), post deadline paper
32.59 E. Vergnol, J.F. Cadiou, A. Carenco, C. Kazmierski: New modulation scheme for integrated single side band lightwave source allowing fiber transport up

to 256 QAM over 38 GHz carrier, Opt. Fiber Commun. Conf. **4**, 134–136 (2000)
32.60 J.J. O'Reilly, P.M. Lane, R. Heidemann, R. Hofstetter: Optical generation of very narrow linewidth millimetre wave signals, Electron. Lett. **28**, 2309–2311 (1992)
32.61 J.J. O'Reilly, P.M. Lane: Fibre supported optical generation and delivery of 60 GHz signals, Electron. Lett. **30**(16), 1329–1330 (1994)
32.62 J.E. Mitchell: Simultaneous up-conversion of multiple wavelengths to 18 GHz and 36 GHz using 4-f technique and optical filtering. In: *Int. Top. Meet. Microw. Photon* (2006), Paper W.4.3
32.63 M. Sauer, K. Kojucharow, H. Kaluzni, D. Sommer, W. Nowak: Simultaneous electro-optical upconversion to 60 GHz of uncoded OFDM signals. In: *Int. Top. Meet. Microw. Photon.* (1998) pp. 219–222
32.64 R.A. Griffin, P.M. Lane, J.J. O'Reilly: Radio over fibre distribution using an optical millimetre wave/DWDM overlay. In: *Opt. Fiber Commun. Conf.* (1999), Paper WD6
32.65 J. Park, W.V. Sorin, K.Y.E. Lau: Limitation of the fibre chromatic dispersion penalty on 1550 nm millimetre wave optical transmission, Electron. Lett. **33**(6), 512–513 (1997)
32.66 A. Nirmalathas, C. Lim, D. Novak, R. Waterhouse: Optical interfaces without light sources for base station designs in fiber wireless systems incorporating WDM. In: *Int. Top. Meet. Microw. Photon*, Vol. 1 (1999) pp. 119–122
32.67 R.A. Griffin, H.M. Salgado, P.M. Lane, J.J. O'Reilly: System capacity for millimeter wave radio over fiber distribution employing an optically supported PLL, J. Lightwave Technol. **17**(12), 2480–2487 (1999)
32.68 T. Ismail, C.-P. Liu, J.E. Mitchell, A.J. Seeds, X. Qian, A. Wonfor, R.V. Penty, I.H. White: Transmission of 37.6-GHz QPSK wireless data over 12.8-km fiber with remote millimeter-wave local oscillator delivery using a bi-directional SOA in a full-duplex system with 2.2-km CWDM fiber ring architecture, IEEE Photon. Technol. Lett. **17**(9), 1989–1991 (2005)
32.69 S. Kawanishi, A. Takada, M. Saruwatari: Wideband frequency-response measurement of optical receivers using optical heterodyne detection, J. Lightwave Technol. **7**(1), 92–98 (1989)
32.70 R.T. Ramos, A.J. Seeds: Fast heterodyne optical phase-lock loop using double quantum well laser diodes, Electron. Lett. **28**(1), 82–83 (1992)
32.71 R.A. Griffin, K. Kitayama: Optical millimetre-wave generation with high spectral purity using feed forward optical field modulation, Electron. Lett. **34**(8), 795–796 (1998)
32.72 L. Noel, D. Marcenac, D. Wake: 120Mbit/s QPSK radio fibre transmission over 100Km of standard fibre at 60 GHz using a master slave injection locked DFB laser source, Electron. Lett. **32**(20), 1895–1897 (1996)
32.73 L.A. Johansson, D. Wake, A.J. Seeds: Millimetre-wave over fibre transmission using a BPSK reference-modulated optical injection phase-lock loop. In: *Opt. Fiber Commun. Conf.*, Vol. 3 (2001) p. WV3-1
32.74 O. Omomukuyo, M.P. Thakur, J.E. Mitchell: A simple 60-GHz MB-OFDM ultra-wideband RoF system based on remote heterodyning, IEEE Photon. Techn. Lett. **25**(3), 268–271 (2014)
32.75 I.G. Insua, D. Plettemeier, C.G. Schäffer: Simple remote heterodyne radio-over-fiber system for gigabit per second wireless access, J. Lightwave Technol. **28**(16), 2289–2295 (2010)
32.76 L. Tomba: On the effect of Wiener phase noise in OFDM systems, IEEE Trans. Commun. **46**(5), 580–583 (1998)
32.77 A. Stoehr, B. Shih, S. Abraha, A.G. Steffan, A. Ng'oma: High spectral-efficient 512-QAM-OFDM 60 GHz CRoF system using a coherent photonic mixer (CPX) and an RF envelope detector. In: *Opt. Fiber Commun. Conf. (OFC 2016), Anaheim* (2016), Paper Tu3B.4
32.78 A. Kanno, P. Dat, T. Kuri, I. Hosako, T. Kawanishi, Y. Yoshida, Y. Yasumura, K. Kitayama: Coherent radio-over-fiber and millimeter-wave radio seamless transmission system for resilient access networks, IEEE Photon. J. **4**(6), 2196–2204 (2012)
32.79 K.I. Kitayama, A. Maruta, Y. Yoshida: Digital coherent technology for optical fiber and radio-over-fiber transmission systems, J. Lightwave Technol. **32**(20), 3411–3420 (2014)
32.80 CPRI Specification V4.2: http://www.cpri.info/spec.html
32.81 S. Dahlfort, K. Laraqui: Exploring the antenna lambda connection. In: *Opt. Fiber Commun. Conf. Expo., Los Angeles* (2012), Paper NTh4E.5
32.82 C. L. I: Optical networks enabled C-RAN fronthaul. In: *IEEE ICC Workshop Opt.-Wirel. Integr. Technol. Syst. Netw., Budapest* (2013)
32.83 D. Wake, S. Pato, J. Pedro, E. Lopez, N. Gomes, P. Monteiro: A comparison of remote radio head optical transmission technologies for next generation wireless systems. In: *Proc. IEEE Photon. Soc. Annu. Meet.* (2009) pp. 442–443
32.84 V. Jungnickel, K. Manolakis, S. Jaeckel, M. Lossow, P. Farkas, M. Schlosser, V. Braun: Backhaul requirements for inter-site cooperation in heterogeneous LTE-advanced networks. In: *Proc. IEEE Int. Workshop Opt.-Wirel. Integr. Technol. Syst. Netw.* (2013) pp. 905–910

John E. Mitchell

Dept. of Electronic and Electrical Engineering
University College London
London, UK
j.mitchell@ucl.ac.uk

John Mitchell is Professor of Communications Systems Engineering in the UCL Department of Electronic and Electrical Engineering. His research work focuses on optical access technologies with a specification consideration of optical network support of wireless techniques. He is currently Coordinator of the EU FP7 European Teaching Network FiWiN5G (FIber-Wireless Integrated Networks for 5th Generation delivery).

33. Optical Networking for 5G and Fiber-Wireless Convergence

Gee-Kung Chang, Mu Xu, Feng Lu

Due to the drastically increased capacity and exploitation of higher RF bands by the upcoming 5G New Radio (5G NR) standards, revolutionary changes are set to occur in next-generation mobile data networks. In the 5G NR specification being developed by 3GPP NR, three different application scenarios are defined: enhanced mobile broadband (eMBB), massive machine-type communication (mMTC), and ultra-reliable low-latency communication (URLLC). Furthermore, following recent debate, the major standards bodies have agreed that both low-density parity check (LDPC) and polar codes should be employed as data and control channel coding options, respectively. Thus, it is clear that, in addition to higher throughput, the flexibility to meet different needs in various application scenarios and to utilize the advantages of different technologies will be key to the construction of 5G fiber-wireless converged heterogeneous networks. In this chapter, to show how optics can support 5G heterogeneous wireless communications, technologies used in fiber-wireless integrated radio access networks (RANs) are reviewed. Aspects such as mobile fronthaul evolution, all-spectrum fiber wireless access technologies, and optical signal processing techniques in 5G converged networks are also discussed.

4k/8k streaming, public cloud services, virtual reality (VR), the Internet of Things (IoT), and machine-type communications have driven the exponential growth in both the overall traffic and the number of connected devices in wireless communication systems in recent years [33.1, 2]. There are also emerging bandwidth, latency, and reliability requirements of vehicle-related communications and other mission-critical applications [33.3] that cannot be met by the 4G Long Term Evolution Advanced (LTE-A) standard. According to the roadmap of the Third Generation Partnership Project (3GPP), 5G mobile networks are expected to be standardized and deployed in the early 2020s to serve these applications [33.4].

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_33

33.1 Challenges Associated with the Introduction of 5G

Based on projections by service providers and major equipment suppliers, 5G mobile networks should be able to support more than 1000 times the overall traffic volume and 100 times the peak data rate per user that can be achieved with the current LTE-A standard. 5G should also be able to support other features such as ultrahigh connection density, increased traffic density, reduced latency, and improved mobility (up to 500 km/h), making it highly adaptable to different application scenarios. These requirements represent a challenge to vendors and service providers in all network segments, including the core network, mobile backhaul, mobile fronthaul, and wideband or ultrareliable wireless access. A series of technologies to facilitate 5G are being investigated, as shown in Fig. 33.1 [33.5].

The high-level key performance indicators (KPI) targets for next-generation mobile networks are shown in Table 33.1. These objectives were developed as part of the IMT-2020. The various applications of 5G can be mapped to three different classes of usage scenarios: enhanced mobile broadband (eMBB), massive machine-type communication (mMTC), and ultra-reliable low-latency communication (URLLC) (Fig. 33.2).

33.1.1 Enhanced Mobile Broadband

eMBB is intended for applications that require high bandwidth and involve high traffic density, such as VR, augmented reality (AR), and ultrahigh-resolution, high frame rate 4k or 8k video streaming [33.6, 7]. As it is designed to support a wide range of services, eMBB is expected to underpin the first commercial utilization of 5G technology. eMBB can be applied in several service categories, including high-speed access in traffic-dense areas, broadband everywhere, and high-speed mobility. The targets for the application of eMBB are shown in Table 33.2.

These tremendous traffic densities are supported by the capacity gain enabled by multiple-antenna systems, including massive MIMO and beamforming technologies. Especially for millimeter-wave (MMW) transmission, a large MIMO antenna array that is within the specified physical limit on antenna size can be em-

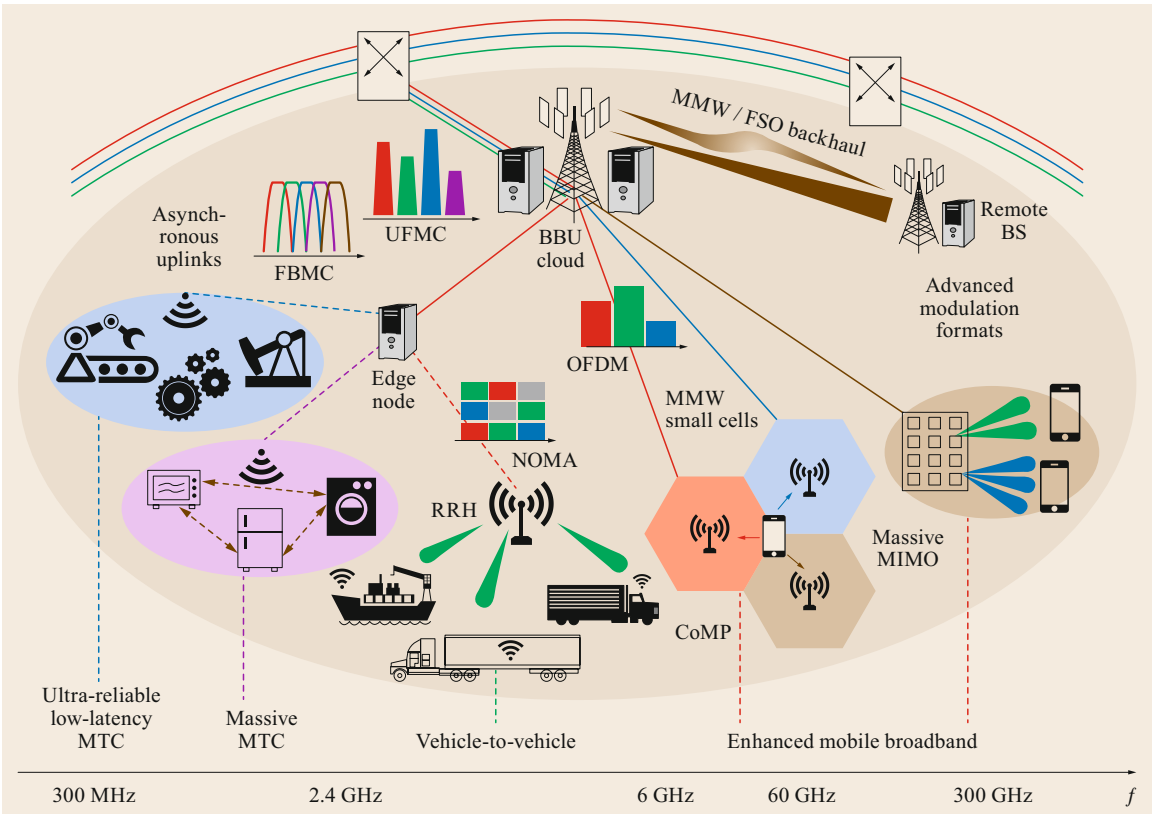


Fig. 33.1 5G technology in both the optical and the wireless domains

Table 33.1 KPIs for next-generation mobile networks

KPI	IMT-2020 performance values
Peak data rate per user	> 10 Gb/s
Minimum data rate per user	> 100 Mb/s
Supported user density	> 1 000 000 connections/km^2
Supported traffic density	> 10 (Tb/s)/km^2
Latency	< 1 ms
Mobility	Up to 500 km/h
Energy consumption	1/10× compared to 2010
Network management operational expenditure (OPEX)	1/5×
Service deployment time	1/1000×

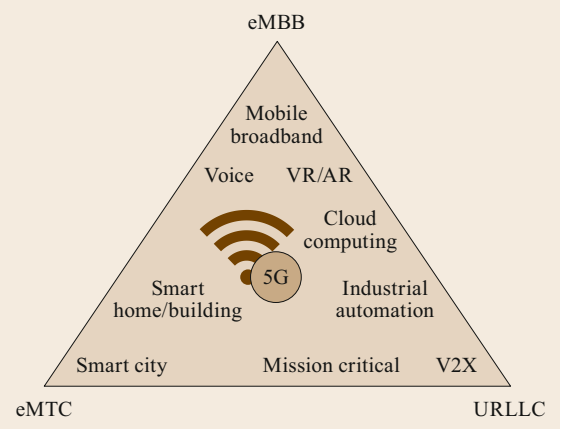


Fig. 33.2 Applications of eMBB, URLLC, and mMTC within the scope of 5G

Table 33.2 eMBB use-case categories and targets (DL: downlink; UL: uplink)

Use-case category	Connection density	Traffic density
Broadband access in traffic-dense areas	200–2500 per km^2	DL: 750 Gb/s per km^2 UL: 125 Gb/s per km^2
Broadband access in a crowd	150 000 per km^2	DL: 3.75 Tb/s per km^2 UL: 7.5 Tb/s per km^2
Mobile broadband in vehicles	2000 per km^2	DL: 25 Gb/s per train 50 Mb/s per car UL: 12.5 Gb/s per train 25 Mb/s per car

ployed due to the reduced wavelength [33.8]. 5G will use MIMO extensively with up to 32 antenna ports at current phase. With MIMO and beamforming, operators will be able to deploy 5G base stations on existing sites to achieve similar coverage, while the signal quality received by the user will be significantly improved, so better performance is expected. Further performance enhancement can be accomplished by using 3D MIMO instead of conventional horizontal or vertical MIMO antenna components with flat-panel antennas [33.9]. In addition, high-order modulation (HOM) of up to 256-QAM or 1024-QAM can be applied due to the increased SNR enabled by high-gain beamforming.

Another approach that can be used to improve capacity is to enlarge the wireless bandwidth. Two methods of achieving this have been proposed and extensively investigated: increasing the number of channels utilized simultaneously by the same user (known as carrier aggregation, CA) [33.10], and expanding the bandwidth of a single channel. 32 CA is defined in LTE-A Pro to facilitate larger data pipes. Intraband and interband operations are applied to utilize the spectral resources available in both licensed and unlicensed ranges, as shown in Fig. 33.3. Greater bandwidths than the currently used 20, 40, and 80 MHz, and so on are expected in the 5G standard. By combining both schemes, wireless bandwidths of more than 1 GHz are achievable. These technologies can provide the capacity density required without introducing a substantial load on the digital signal processing (DSP) side, but they are significantly limited by the channels available in bands below 6 GHz. These bands are crowded and an exclusive license for operation in these bands to

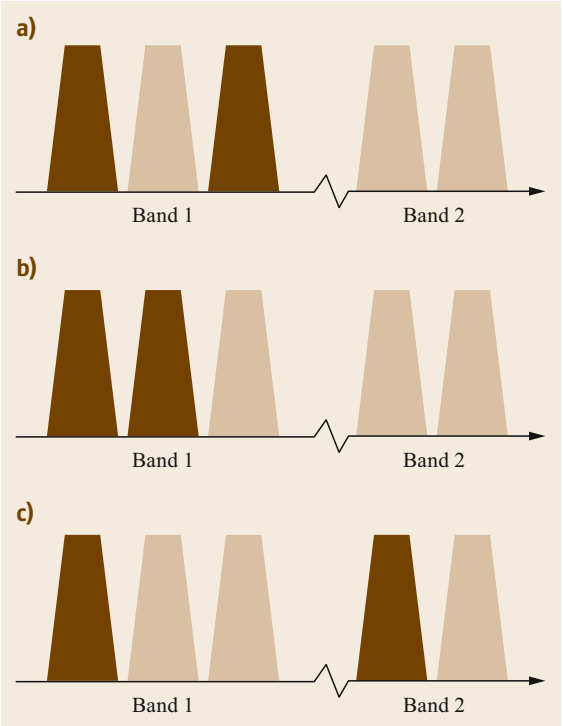


Fig. 33.3a–c Various carrier aggregation modes defined in LTE-A and 5G. **(a)** intraband noncontiguous, **(b)** intraband contiguous, and **(c)** interband

maintain the quality of service required by 5G is expensive. MMW bands, which have vast potential, are being considered for the 5G spectrum, especially in eMBB scenarios [33.6].

To meet the requirements of eMBB under extreme conditions, high-density deployment will be necessary [33.11]. As they are a well-established technology, heterogeneous networks incorporating macro and small cells are known to permit a good balance between maximal capacity density, coverage, and deployment difficulty. With small cells, the reduced cell size increases the spatial reuse ratio such that the data rate can be increased dramatically for each user [33.12]. However, synchronizing and controlling high-density cells is challenging given the severe intercell interference (ICI), and coordination is needed to achieve the expected capacity gain. Coordination can be achieved by using centralized radio access network (C-RAN) architectures, which allows a significant portion of the transmission/reception DSP to be conducted in a centralized infrastructure shared by multiple cell sites [33.13].

33.1.2 Machine-Type Communications

An unprecedented number of services and applications that require communication among machines, such as drone-based packet delivery, autonomous vehicles, smart homes, smart factories, smart cities, and large-scale sensor applications, are expected to be served by wireless networks in the future. The communication schemes associated with these applications are different from traditional human-centric communications based on eMBB [33.14]: they have distinct requirements in various respects, including latency, data rate, reliability, and energy efficiency. The coexistence of human-centric communications and machine-type communications will lead to more diverse wireless networks in the future. The International Telecommunication Union (ITU) divides these machine-related applications into those based on ultra-reliable low-latency communication (URLLC) and those based on massive machine-type communication (mMTC).

mMTC facilitates services such as logging, metering, security monitoring, and asset condition tracking in which a massive number of machine-type devices are connected within a limited area. High connection densities of up to 1 million connections per km^2 and ultraefficient energy consumption allowing a battery life of decades are supported by this 5G scenario. Current mMTC technologies, including NB-IoT, SigFox, and LoRA, aim to improve coverage while maintaining high power efficiency, but a significant increase in the number of users supported is yet to be reported. Some solutions based on nonorthogonal multiplexing and multiple access (NOMA) that use nonorthogonal resources to serve multiple users within limited time and frequency intervals have been proposed [33.15]. NOMA can remove the strict user volume limitation based on availability, allowing far more users to be supported than in the conventional approach [33.16].

URLLC focuses on optimizing the reliability and latency performance of wireless transmission, making it suitable for applications that are mission critical, such as autonomous cars, wireless industrial control, and drone-related applications [33.17]. The design of the waveform for an URLLC service is challenging and problematic, as reliability and low latency lead to conflicting requirements to some degree, and high mobility also needs to be accounted for to ensure better reception performance [33.18]. The utilization of short packets is typically preferable when the aim is to reduce the latency, leading to a small forward error correction (FEC) overhead that limits the coding gain and reduces the link performance. A stringent latency requirement also limits the availability of retransmission, a technique that is widely used in current wireless communications to enhance reliability. There are three methods that can be used to achieve both reliability and latency from a physical layer perspective: reducing the transmission overhead, improving the reception of the URLLC packet, and reducing the delay in transmitting the URLLC packet. More specifically, the system overhead should be optimized and minimized, including the information related to channel estimation, scheduling, resource allocation, and data verification; high channel coding overhead or other spreading techniques should be considered as a means to increase the success decoding probability, since retransmission will not satisfy the strict latency requirement; and finally, when the URLLC packet is generated from higher layers, the packet should be transmitted immediately instead of waiting for the next available slot. When URLLC coexists with eMBB, eMBB frame puncturing by URLLC packets would be expected [33.19].

33.1.3 Coordination and Network MIMO

There are several aspects to 5G coordination: spatial coordination, spectral coordination, and service-wise coexistence and coordination [33.20]. Spatial coordination becomes critical when small high-density cells are deployed and ICI is inevitable. When the capacity penalty due to ICI is of a similar scale to the capacity gain, the spatial densification of cells does not yield a gain in performance. Hence, intercell coordination plays an important role in mitigating interference between cells using joint processing and joint transmission [33.21]. Coordinated multipoint (CoMP) and

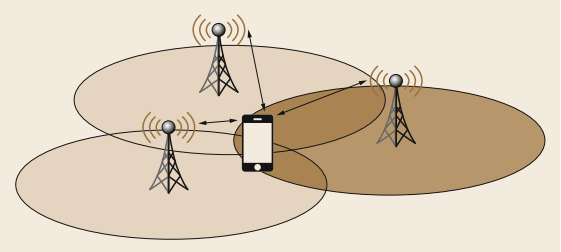


Fig. 33.4 Coordinated multipoint transmission and network MIMO

network MIMO are promising concepts for improving the cell-edge user data rate and spectral efficiency. Both of these network coordination schemes have been proposed as approaches that can be used to mitigate intercell interference and increase spectral efficiency. Figure 33.4 shows the concept of CoMP, where one user is served by all available cells nearby using the same spectral resources. The interference in the cell-edge area is minimized and the transmission power and equivalent gain are boosted through centralized processing or by exchanging information among the cells. In particular, now that MMW is becoming a prominent candidate spectrum for 5G exploration, the line-of-sight requirement for MMW directional beams limits service availability significantly in single-cell operation [33.22]. However, using multiple cells significantly increases the probability of achieving a high-quality radio path, meaning that the user has a better chance of receiving good SNR. The backhauling link is the key challenge in spatial coordination, since an enormous amount of overhead information is generated between traditional cell towers. The highly dynamic channels associated with mobile users require the latency of this traffic to be extremely low. Another issue is the synchronization among these geodistributed cells, since any offset in the frequency or time domain will reduce the performance gain or even turn the constructive signal into a destructive one and severely degrade the transmission.

Spectral coordination comes into play when multiple-band operation is included in a single wireless service. For example, CA utilizes multiple contiguous or noncontiguous bands for joint simultaneous wireless transmission to enhance the system capacity, which is important for eMBB. Interband or intraband CA was standardized in 3GPP release 13 and 14 (also known as LTE-A Pro), and coordination between bands below 6 GHz and MMW bands is possible in 5G, enabling both the low-loss properties of low-frequency channels and the wide bandwidth of MMW to be exploited. In addition, both licensed and unlicensed bands can be coordinated for data pipe expansion with the same or similar standards.

The capacity requirement of 5G necessitates more efficient use of all available licensed and unlicensed bands. Several bands have been opened up for mobile services, including the 2.5 and 5 GHz bands for the proposed use of unlicensed LTE carriers as secondary LTE carriers in CA. However, WiFi is currently running on these bands. These emerging bands will lead to the cochannel deployment of multiple radio access technologies [33.23]. However, under the coexistence of multiple services in the same frequency and space, resource utilization is only improved when multiple services are properly coordinated and controlled.

33.1.4 Mapping 5G Challenges to the Optical Network

5G imposes various requirements on next-generation access networks. As it will be the dominant supporting media for these networks, the optical transport network will also need to be upgraded to accommodate emerging applications.

Optical networks that support 5G C-RAN can be divided into mobile fronthaul and backhaul. Mobile backhaul refers to the communication network infrastructure for transporting packet data from end users or nodes to the central network or infrastructure and vice versa. It carries the raw communication data. So, for 5G, bigger data pipes (up to Tb/s) in the optical network and ways to offload traffic to the network edge are needed. In terms of latency, a fast and resilient forwarding engine that leverages the edge routing and switching infrastructure is needed. From a cost perspective, the use of multiplexing technologies such as time-division multiplexing (TDM) and wavelength-division multiplexing (WDM) can reduce the need for fiber deployment. Also, deployment and upgrading costs can be significantly reduced by ignoring specialized switching hardware and using general-purpose infrastructure and network function virtualization (NFV). Moreover, the system should be flexible and software-defined so that it is virtualized and self-contained and thus complies with the requirement for rapid deployment in 5G.

On the other hand, mobile fronthaul refers to a new network architecture in which centralized baseband controllers are connected to remote standalone radio heads at cell sites. The connections may be standardized digital optical links that use pluggable modules or they may be customized analog optical links. There are, however, various challenges associated with mobile fronthaul, including those relating to capacity, latency, and efficiency. Currently, mobile fronthaul based on the common public radio interface (CPRI) requires massive bandwidth, as shown in Table 33.3. Methods of reducing the required fronthaul capacity using com-

Table 33.3 Data requirements of the CPRI in optical links

Number of antenna ports	Wireless bandwidth		
	20 MHz	160 MHz	1 GHz
2	2 Gb/s	16 Gb/s	100 Gb/s
8	8 Gb/s	64 Gb/s	400 Gb/s
64	64 Gb/s	512 Gb/s	3.2 Tb/s
256	256 Gb/s	2 Tb/s	12.8 Tb/s

pression or multiple functional splits to decrease the bandwidth needed and maintain good signal quality to and from the radio head are being investigated [33.24, 25]. The optical bandwidth can also be reduced by utilizing analog mobile fronthaul, which makes use of linear optical components and adaptively designed optimization DSP [33.26, 27]. Since mobile fronthaul latency can add to the physical (PHY) layer delay, to fit into the URLLC framework, the optical link must utilize fast networking and adaptively designed scheduling schemes when multiple cells are sharing optical transmission resources. Statistical multiplexing can be leveraged in mobile fronthaul to increase the efficiency by shifting the dedicated constant-bit-rate (CBR) CPRI links to a more shareable infrastructure.

Analog mobile fronthaul provides another approach for achieving high spectral efficiency and minimal latency in front ends. The bandwidth occupation is the same as in free space and is less than 10% of that required with CPRI. The latency is also negligible as the front-end processing can be purely analog. That said, there are several challenges with analog technology. During analog signal delivery, it is very hard to verify and correct any transmission errors, and the link suffers from noise and burst interference. The linearity of the optical transceiver must be very high because of the multicarrier wireless waveforms delivered along the fiber. In addition, since multiple carriers must be supported simultaneously for MIMO and CA, multiplexing is another issue in an analog link that uses radio-over-fiber technology [33.28]. There are considerable discussions regarding the utilization of TDM, FDM, and WDM in analog mobile fronthaul.

33.2 Overview of Fiber-Wireless Integrated Fronthaul Systems in 5G

Future 5G mobile data networks empowered by spectral aggregation and cell densification pose a huge challenge for building next-generation mobile fronthaul systems with large capacities, scalability, and high energy efficiency. Traditional solutions based on CPRI or channel aggregation with DSP are not sufficient to support heterogeneous ubiquitous wireless access. In this section, we compare different fiber-wireless integrated mobile fronthaul solutions and envisage future optical networking technologies that could potentially reshape next-generation radio access networks.

33.2.1 C-RAN and Digital Mobile Fronthaul

Due to the constant demand for higher internet speeds and seamless expanded wireless signal coverage, fifth-generation (5G) mobile data networks are currently being constructed. The new 5G standards are under review and the related technologies have been under deployment since 2019. 5G products are predicted to emerge onto the market around 2020 [33.29]. Among all of the promising new features of 5G, millimeter-wave frequency exploration, ultralow latency, and small-cell densification are critical drivers of significant improvements in system capacity. Considerable challenges are brought by carrier aggregation and cell densification. Figure 33.5 depicts the impact of spectral aggregation and cell densification on future 5G network systems with centralized control and management. Higher throughput to subscribers is achieved using an expanded spectral bandwidth and higher spectral efficiency. Meanwhile, 5G also makes it possible to utilize fragmented frequency resources and enables the coexistence of multiple radio access technologies (RATs). On the other hand, an increased small-cell density potentially facilitates wide and seamless wireless signal coverage with a greatly improved customer experience through frequency reuse and millimeter-waveband exploitation.

As mentioned above, spectral aggregation and cell densification bring some challenges. Above all, the smaller cell size makes it more difficult to rebalance the performance in different areas. It is important to keep the quality of service (QoS) stable and smooth at the cell edge or during the handover between adjacent cells. The situation is also complicated by the significantly increased antenna and equipment densities associated with 5G networks [33.30]. Moreover, in the 5G NR system, cell splitting, coordinated multipoint, massive MIMO, and beamforming are important tools for optimizing the spatial distribution of the signal strength and improving performance at cell edges. To achieve this, a powerful central office (CO) with a high processing speed and enhanced computing capability is needed to coordinate and schedule on a large scale. On the other hand, upgrading to next-generation

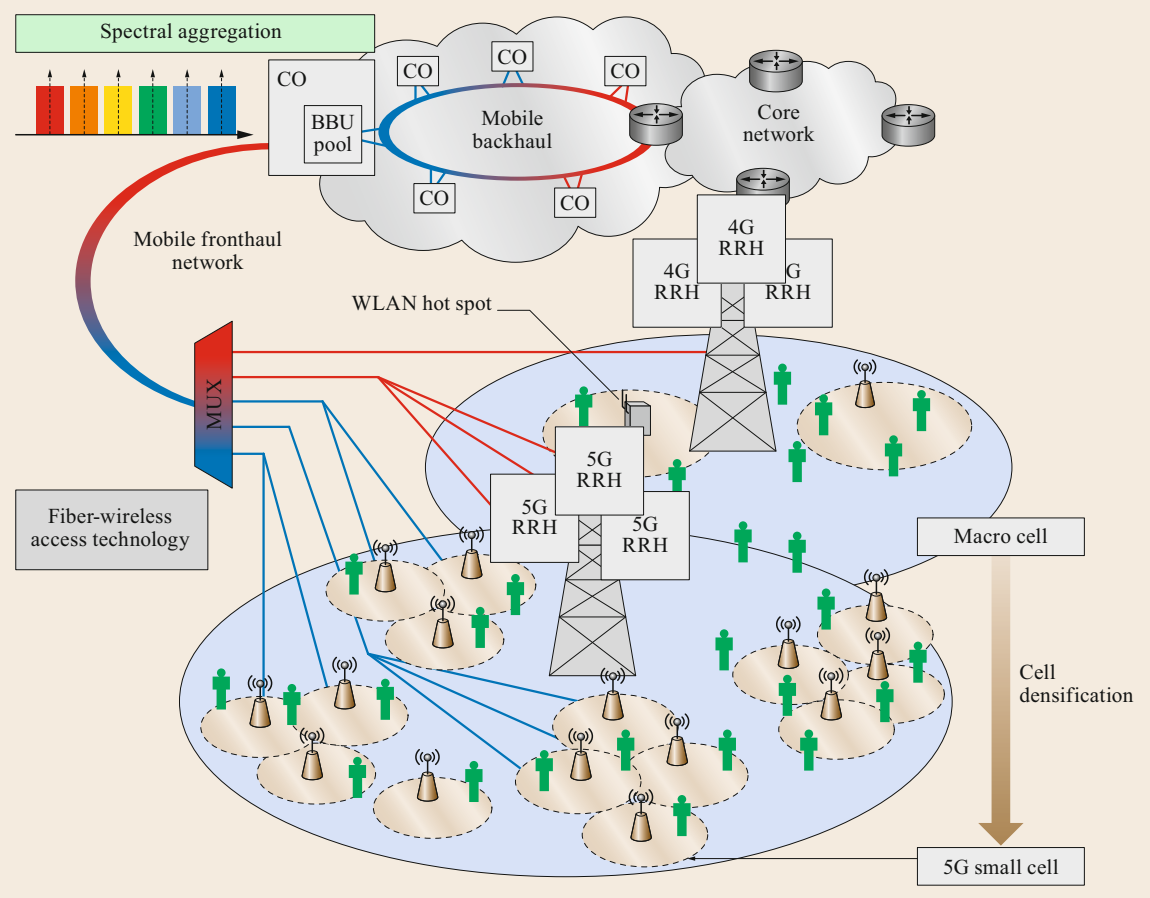


Fig. 33.5 Conceptual diagram of the impact of spectral aggregation and cell densification on a 5G radio access network

mobile technologies also requires increased capital expenditure (CAPEX), operational expenditure (OPEX), power, and time. While unavoidable, those increased inputs should be kept within a reasonable range. Meanwhile, 5G NR must also be sufficiently adaptable to support delay-sensitive and energy-efficient applications such as tactile networks, IoT, and machine-type communications. Thus, compatibility with different usage scenarios is of great importance and should be considered in efforts to design and optimize the frame structure, network functions, and hardware implementation. So far, based on information that has leaked from major standards bodies, the consensus appears to be that the new 5G NR system should not only support higher bandwidths and throughput but also it should be more efficient in terms of CAPEX, OPEX, as well as energy consumption. the end-to-end latency should also be significantly reduced to support URLLC [33.13].

The C-RAN architecture has been proposed by China Mobile [33.31–33] as a solution to address the challenges involved in large-scale cell control and management. In this proposal, baseband units (BBUs) that perform some processing are decoupled from the radio equipment controllers below antenna towers and aggregated into a pool of BBUs that work as a central unit. As shown in Fig. 33.5, this central unit can control and manage hundreds of distributed remote units and radio heads in various small cells. The central unit has a large storage volume and high data-processing capabilities. Some cutting-edge networking technologies such as software-defined networking, network function virtualization, and cloud-edge computing will also be deployed to facilitate real-time management and time scheduling on a large scale. This approach to constructing future fiber-wireless converged networks will allow the throughput of the fixed network to be boosted. Optical fibers will be used for the transmission links in the C-RAN architecture due to their low propagation loss, high data capacity, and low cost in building a full duplexed network connecting the central unit to a large number of remote units.

There are multiple options for digital data transmission in C-RAN, most of which are realized utilizing a digitized radio-over-fiber (D-RoF) technology, CPRI (more information about CPRI can be found in Chap. 31) [33.34]. CPRI defines a set of rules for the digitization, packetizing, and forward error coding processes. In CPRI fronthaul systems, the baseband in-phase (I) and quadrature (Q) analog waveforms are first reduced to discrete samples. Then the discrete I/Q samples are quantized with 15 digits, and one control word digit is also added before each digitized sample; these digits together form an antenna component (AxC) chip. Different AxC chips to different antennas are interleaved at the BBU pool before being transmitted to the remote units. At the remote units, the AxC chips are collected and used to reconstruct the analog baseband signals before they are upconverted to radiofrequencies and sent to the radio equipment (RE). However, one of the main problems with CPRI is its low bandwidth efficiency. CPRI requires more than 16 times more bandwidth than LTE to transmit a particular amount of information. After quantization, a component carrier in LTE that typically occupies 20 MHz is expanded to a 490 Mb/s signal in a non-return-to-zero (NRZ) format. Calculations indicate that the speed of a CPRI link within a standard three-sector 8-by-8 MIMO LTE base station can reach 150 Gb/s, which is intolerable and significantly increases the cost. On the other hand, D-RoF schemes may be indispensable for short-distance low-latency transmission using high-order wireless formats in 5G NR systems. The latest 5G NR specifications feature orthogonal frequency-division multiplexing (OFDM) and high-order modulation (256 and 1024-QAM). However, integrating these two technologies to achieve the desired operational flexibility and spectral efficiency also leads to several new challenges, such as high sensitivity to nonlinear distortions and increased demands on the high-resolution digital-to-analog converters (DAC), limiting the quality and transmission distance of analog radio-over-fiber (A-RoF) optical networks in mobile fronthaul (MFH) [33.35]. D-RoF is the preferred technology in this case. It is compatible with different formats, making it suitable for the 5G NR environment with its diverse services and spectrum. It also has high immunity to nonlinear distortions. Error-free transmission can be realized using FEC coding. Its shortcoming—low bandwidth efficiency—can also be mitigated through data compression [33.36] and the use of advanced modulation formats, making D-RoF a promising candidate for the digital transportation of high-quality wireless signals across increased transmission distances between the BBU pool and radio access units (RAUs) and with improved power budgets.

33.2.2 Functional Splitting and the Next-Generation Fronthaul Interface (NGFI)

To meet the requirements of 5G NR, current 4G architectures must be upgraded to meet the different latency and throughput requirements of different scenarios. The legacy C-RAN architecture is evolving into the next-generation fronthaul interface (NGFI), meaning that functionality must be relocated to the remote radio unit (RRU) to alleviate the processing burden in the mobile fronthaul [33.37]. Functions are then split between two segments: a radio cloud center (RCC) and the RAU. The functionality can be reconfigured and reorganized depending on the splitting point selected. For example, a simple remote radio head (RRH) that can only handle the RF (radio frequency) functionality may be used (this is known as option 8). Alternatively, the radio head could inherit part or all of the functions in the PHY layer (this is known as option 7). There are also many other possible splitting points besides these two options, each of which has unique properties and advantages that may make it the best option for a particular application scenario. Therefore, a flexible functional split is proposed in [33.38], where the different split options can coexist and can be dynamically adjusted over time.

NGFI, with its capacity for reconfigurable and multiple coexisting functional splits, is proposed in the IEEE 1914.1 standard. Figure 33.6 shows a conceptual diagram of a NGFI-based mobile fronthaul network with a two-layer architecture: fronthaul I is a network that connects each distributed unit (DU) to several RRUs, while fronthaul II links a central unit (CU) to multiple DUs. The location of the CU can be tailored to different application environments. For example, the CU could be located in the (transport) aggregation layer, yielding a mMTC scenario with a large number of connections. For eMBB, the CU could be deployed at the access layer to augment the throughput. Forward and backward compatibility must also be considered. Legacy fronthaul infrastructure inherited from 3G and 4G systems is incorporated into the networks in options 1 and 2. Small cell densification is supported by options 7 and 8, since the centralized network architecture can process a large number of remote access points at the same time, helping to simplify the functions assigned to and the cost of the remote units. A reconfigurable functional split (option X) can also be implemented depending on the requirements of the system. Such a two-layer architecture could represent a good trade-off between latency and capacity. For some services with stringent delay requirements, such as unmanned self-driving, security, and positioning, a DU can make a quick decision without sending the dataflow to the

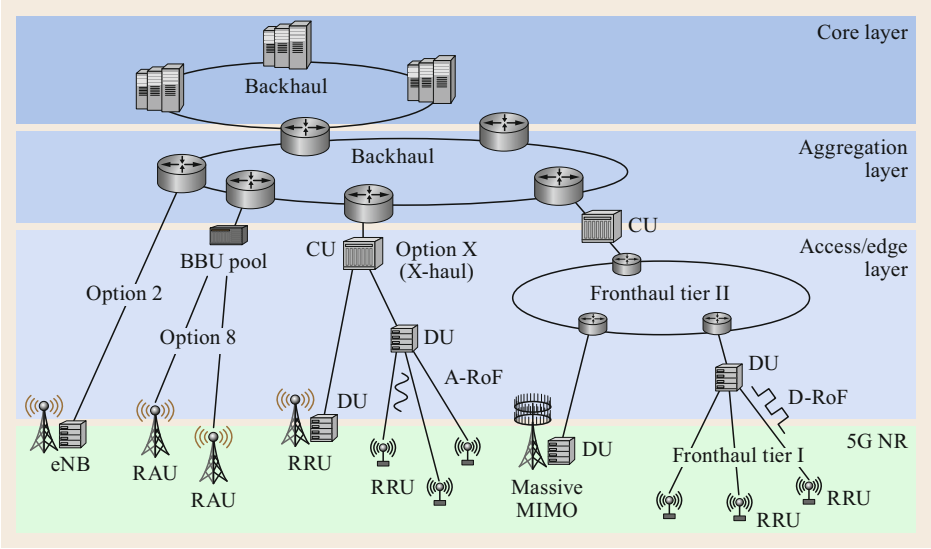


Fig. 33.6 Conceptual diagram of the next-generation fronthaul interface (NGFI)

CU. However, for some tasks that require complex computing and scheduling, the CU can be included in order to utilize its high-performance computational resources and large storage capacity.

The various potential functional splits can be classified into four main categories, as shown in Fig. 33.7. It was mentioned earlier that functions are redistributed between the CU and DUs in NGFI. Eight functional split options are defined and recommended by the 3GPP [33.39]. These options have different requirements in terms of bandwidth, latency, jitter, and synchronization. Options 1 and 2 are fully distributed architectures in which most of the functions are aggregated at the DUs. In these two options, Ethernet packets are transmitted between the CU and DUs. There is almost no wireless overhead, leading to high transmission efficiency, low latency, and low synchronization. However, their distributed architectures make them incompatible with MIMO and coordinated multipoint (CoMP) transmission. It is worth noting that in these options, the resources can be aggregated at the network edge, thus leveraging mobile edge computing for delay-sensitive services. Options 3 to 5 are intermediate between distributed and partly centralized architectures. On the other hand, options 6 and 7 are partly centralized architectures that can provide a good trade-off between transmission efficiency and radio performance, although they have high synchronization and latency requirements because the resource blocks of LTE need to be aligned between the DUs and RRUs. Option 8 is a fully centralized option; some researchers also refer it as the A-RoF option. Since the LTE component carriers are not modified, MFH and air transmission can be seamlessly integrated together. The best radio performance is obtained with centralized coordination and scheduling, making it fully compatible with CoMP. However, implementing mobile fronthaul using this option will necessitate very strict requirements in terms

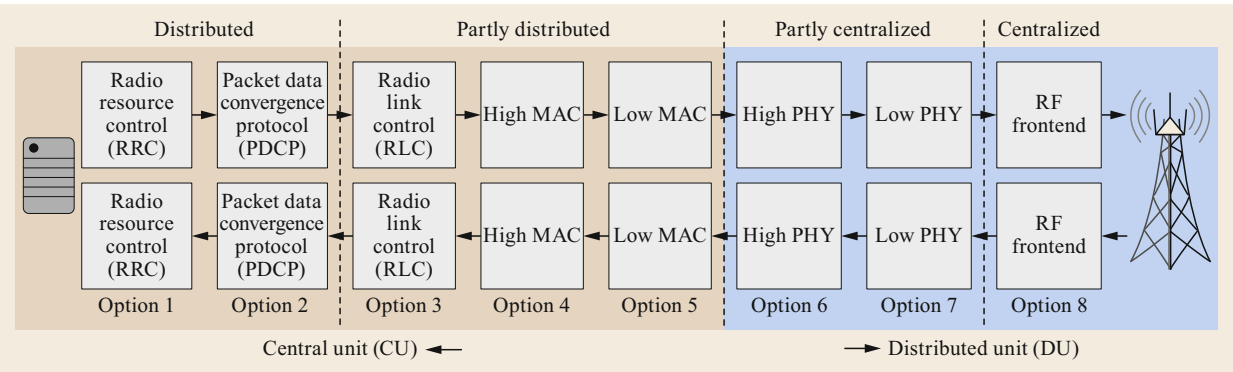


Fig. 33.7 Conceptual diagram of different functional split options in mobile fronthaul

Table 33.4 Comparison of different functional split options in NGFI

Option	Description	Applications
1	RRC-PDCP split: In this split option, RRC is in the CU. PDCP, RLC, MAC, the physical (PHY) layer, and RF are in the DU, so the entire user plane is in the DU.	Edge computing
2	PDCP-RLC split: In this split option, RRC and PDCP are in the CU; RLC, MAC, PHY layer, and RF are in the DU.	Low-latency transmission (already standardized)
3	Intra-RLC split: Low RLC (one function of RLC), MAC, PHY layer, and RF are in the DU; PDCP and high RLC (other function of RLC) are in the CU.	Low-latency transmission Implementing intra-gNB (next-generation node B) RAN-based mobility Fronthaul flow control
4	RLC-MAC split: MAC, PHY layer, and RF are in the DU; PDCP and RLC are in the CU.	Not recommended by 3GPP
5	Intra-MAC split: RF, PHY layer, and some of parts of the MAC layer (e.g., HARQ, hybrid automatic repeat request) are in the DU; upper layer is in the CU.	Collecting UE statistics Measuring/estimating fronthaul activities
6	MAC-PHY split: MAC and upper layers are in the CU; PHY layer and RF are in the DU. The interface between the CU and DUs carries data, configuration, and scheduling-related information (e.g., MCS, layer mapping, beamforming, antenna configuration, and resource block allocation), and measurements.	Centralized scheduling Resource pooling at MAC layer in the central office
7	Intra-PHY split: In the UL, FFT (fast Fourier transform) and CP (cycle prefix) removal reside in the DU. The remaining functions reside in the CU. In the downlink, IFFT and CP addition reside in the DU. The rest of the PHY layer resides in the CU. Multiple realizations are possible.	Centralized scheduling Compatible with joint processing and coordinated multipoint transmission Massive MIMO High-performance radio
8	This option allows the RF and the PHY layer to be separated. It permits the centralization of processes at all protocol layer levels, resulting in very tight coordination of the RAN. This allows efficient support of functions such as CoMP, MIMO, load balancing, and mobility. Realizations include A-RoF, D-RoF, and delta-sigma modulation.	Low-cost RRUs Short-distance PTP transmissions Efficient support of functions such as CoMP, MIMO, load balancing, and mobility

of latency control. Table 33.4 compares the technical features and application scenarios of functional split options 1 to 8.

33.2.3 Analog Radio-over-Fiber Systems

To overcome the drawback of the low spectral efficiency of a D-RoF interface such as CPRI, analog radio-over-fiber (A-RoF)-based channel aggregation/deaggregation (CA/CDA) has been proposed for mobile fronthaul systems [33.40–42], as shown in Fig. 33.8. In such a scheme, multiple LTE signals are transmitted with frequency-division multiplexing (FDM) over one WDM channel, allowing the spectral efficiency to be greatly improved. For example, in [33.41], a RF bandwidth of 1.5 GHz can support 48 20 MHz LTE signals.

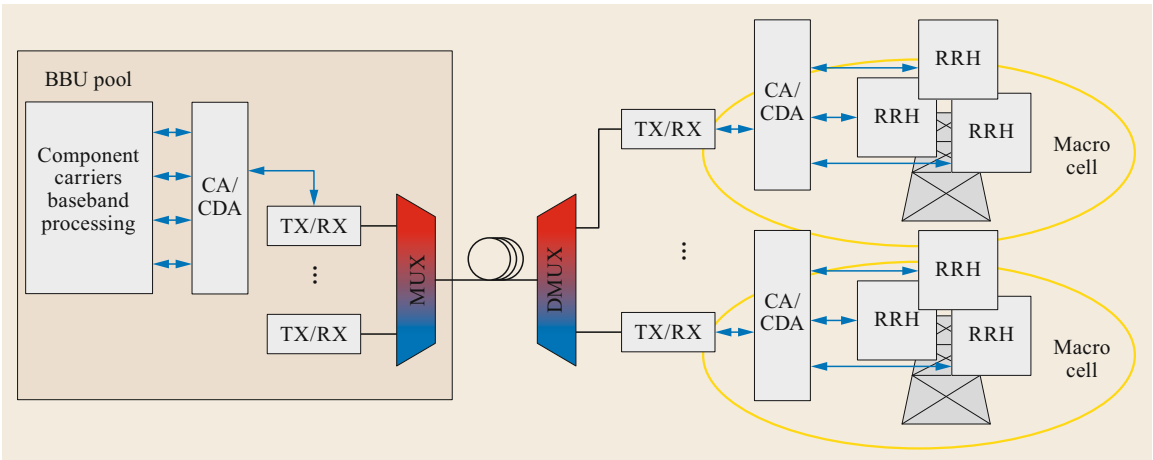


Fig. 33.8 Architecture of channel aggregation/deaggregation (CA/CDA)-based mobile fronthaul

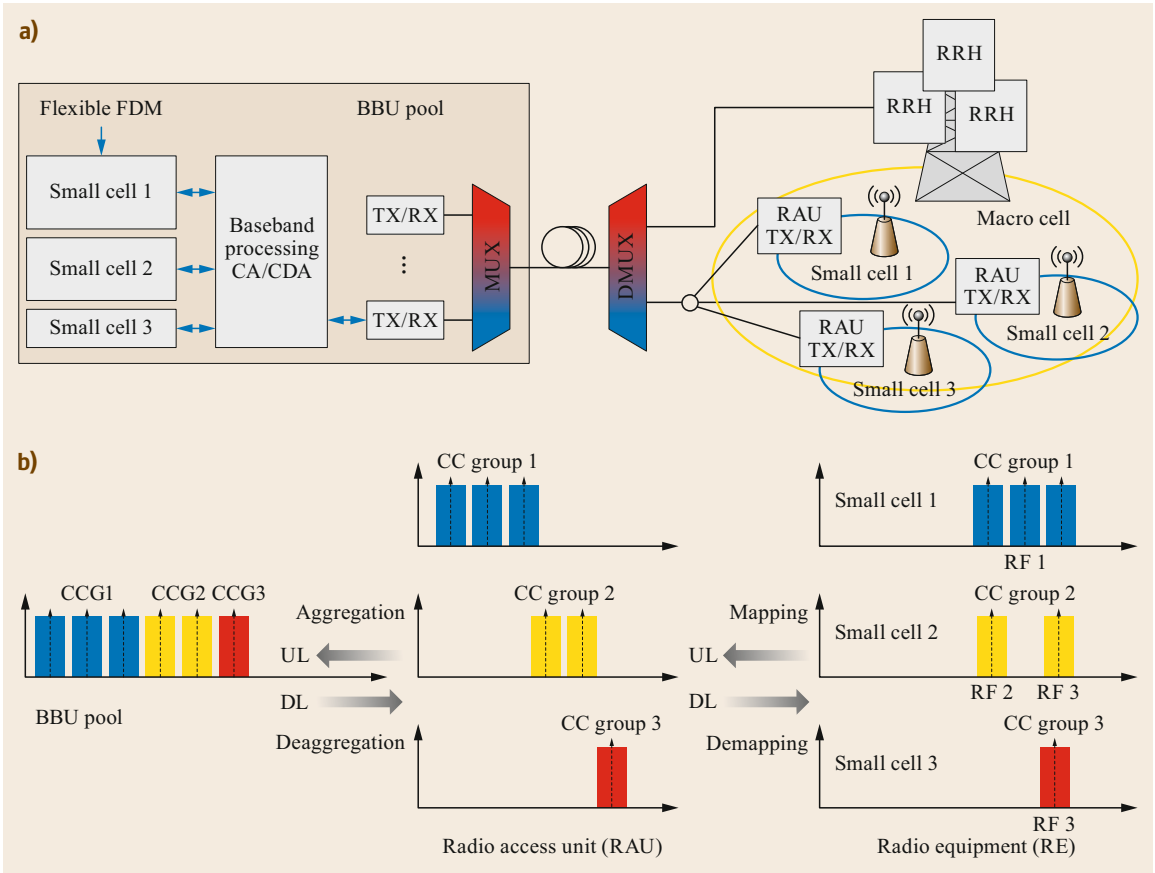


Fig. 33.9a,b A point-to-multipoint mobile fronthaul network: **(a)** architecture, **(b)** working principle

The formats of the wireless signals as well as their analog properties are maintained during the transmission process. A-RoF yields a good radio performance with fully centralized control and coordination. It is worth noting that the FDM can be realized with a pure DSP scheme or a digital–analog hybrid method. Nevertheless, the existing mobile fronthaul system with carrier aggregation/deaggregation is implemented over a WDM passive optical network (PON) with a point-to-point (PTP) multichannel architecture. There are also some issues with this scheme. First of all, it lacks the scalability and flexibility for PTP WDM to facilitate cell splitting and densification in 5G mobile radio access networks. The number of usable wavelengths in a WDM PON is quickly depleted as the number of cells increases because each takes up one wavelength. Moreover, dense WDM (DWDM) multiplexers and wavelength-tunable lasers are expensive.

In addition to point-to-point transmission, analog radio-over-fiber systems are good candidates for bidirectional mobile fronthaul systems to support carrier aggregation/deaggregation in small cells. As indicated by Fig. 33.9a, the proposed network in this case is built on a point-to-multipoint WDM-FDM architecture. Therefore, it can be connected to a great number of small cells or hot spots in a 5G heterogeneous network. On the other hand, the system can support the existing infrastructure with macro cells that provide lower-speed wireless access and complementary control as well as management functions. As illustrated in the inset of Fig. 33.9b, the overall operation of the mobile fronthaul involves two main links. In downlink (DL) transmission, groups consisting of different component carriers allocated to different cells are generated by the baseband units. These groups are received and divided inside the RAUs at the corresponding cell sites after transmission through fiber. After signal reception, the signals in different groups can either be upconverted through analog mixers or reconstructed through a digital signal recovery process that digitally maps the signal onto its corresponding RF bands. This proposed architecture increases the number of connected cells, and the signals allocated to different cell sites can be efficiently and flexibly converted into different intermediate frequencies.

On the other hand, realizing an uplink PTMP (point to multipoint) architecture is more difficult and sophisticated than realizing a downlink scheme if subband or subcarrier multiplexed technologies are utilized in a small-cell mobile fronthaul. Because every cell site is independently operated with distributed clocks, the optical signals from the various cell sites are also unsynchronized and uncorrelated. When they are mixed with each other and beat within a directly detected power receiver, the signal-to-noise ratio of the frequency components is significantly deteriorated by random phase noise from the lasers and optical beating interference (OBI). Although the use of coherent receivers to cancel out the additive interference from OBI is proposed in [33.43–45], the significantly increased complexity of digital signal processing and hardware implementation as well as bias control and polarization tracking inflate the capital expenditure and operational expenditure tremendously. Aside from these issues, due to fabrication defects and the imperfection of the optical hybrid as well as the balanced photodetectors, OBI cannot be eliminated through cancellation inside the coherent receivers. Thus, the residual additive interference will seriously degrade the signal quality and interactions between the OBI and other nonlinear effects will complicate the situation.

Bidirectional transmission techniques based on field modulation and heterodyne detection (FM-HD) have recently been proposed [33.46] for a fronthaul network supporting point-to-multipoint communication from a central unit to distributed units in small cells. Through the utilization of optical beating in heterodyne detection and adjusting to the optical local oscillator wavelength, the desired signals distributed on intermediate-frequency components can be shifted from the region influenced by OBI. Furthermore, inside the heterodyne-detected system, other than beating into the intensity noise, the random optical phase noise can be estimated by tracking the phase variations of the intermediate frequency (IF) or pilot tones, which permits the application of some effective DSP techniques to recover the phase of the received signal. Therefore, the effects of OBIs can be significantly suppressed using the FM-HD scheme. Laser-free RAUs deployed as centralized laser sources that are shared among different cell sites to further reduce the cost and increase the scalability of the bidirectional system have also been proposed. Nonetheless, one problem with this technology is that most Mach–Zehnder modulators are rather sensitive to polarization misalignment, so polarization control and tracking systems are necessary at the distributed RAUs, leading to high transceiver module complexity and cost.

33.2.4 Multiplexing and Coordination in Mobile Fronthaul

In a current D-RoF mobile fronthaul network interface such as CPRI, to reduce the delay and eliminate the need to apply a complex network scheduling mechanism, a fixed-bandwidth allocation algorithm is utilized instead of the dynamic bandwidth allocation (DBA) algorithm used in existing TDM or Ethernet systems (see Chap. 27). Furthermore, time-domain AxC-chip interleaving is used to multiplex the binary codes to distributed radio equipment for analog baseband component reconstruction [33.47]. As demonstrated in Fig. 33.10a, to construct a standardized CPRI frame, digitized samples are arranged in order and periodi-

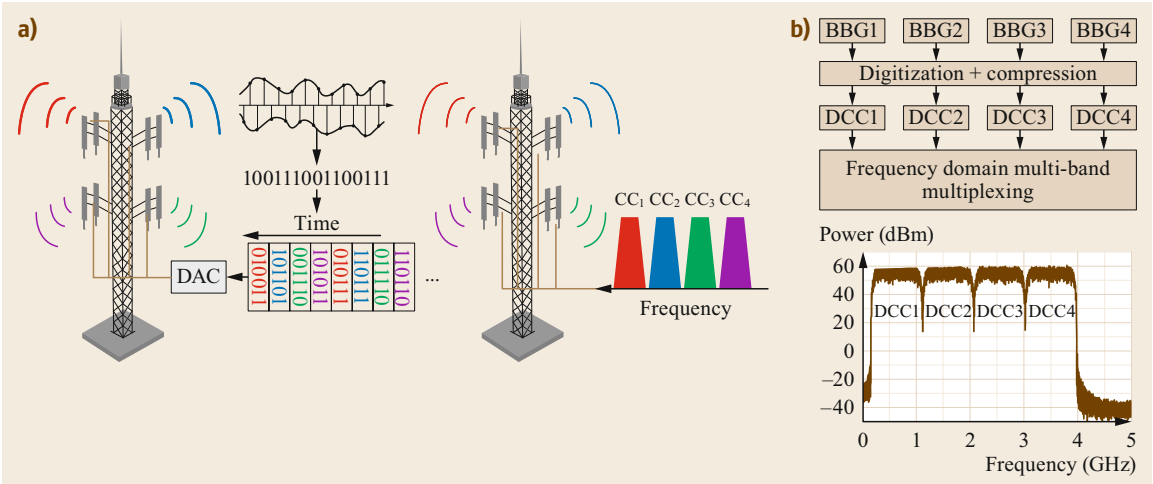


Fig. 33.10 **(a)** Operational principle of data multiplexing in a traditional CPRI-based MFH. **(b)** Flow diagram of the proposed approach to data multiplexing in MFH incorporating frequency-domain multiband modulation (BBG: baseband group; DCC: digitized component carrier)

cally spaced before they are transmitted between a radio equipment controller and radio equipment. This AxC chip interleaving is simple, but there are still some critical problems with this approach that make it unsuitable for 5G mobile fronthaul. Above all, in order to realize the interleaved arrangement, the sampling clock of every component carrier (CC) needs to be the same or to share the same clock granularity. Although this criterion can be met by current LTE systems, to do so reduces the flexibility of the system as well as its compatibility with future developments in multi-RAT coexistence in 5G heterogeneous networks. Another problem is that resampling makes it extremely difficult to keep the 30.72 MHz sampling clock fixed (e.g., for two component carriers of 20 and 15 MHz with resampling rates of 3/4 and 4/5, respectively). In this case, the interleaving algorithm will be rather complicated. Moreover, point-to-multipoint uplink transmission becomes very tricky, as time interleaving requires strict synchronization among all the AxC chips, which is almost impossible to achieve with multiple distributed RAUs.

To solve the problems caused by sample interleaving, it may be feasible to apply multiband multiplexing in the spectral domain over a digital fronthaul system. The proposed concept is displayed diagrammatically in Fig. 33.10b. One or multiple baseband components are interleaved according to the standards of the CPRI or the compressed CPRI. The bundled components are called a baseband group (BBG), which is usually fed to one radio access unit. After being digitally quantized and compressed, the various BBGs are reduced to digitized component carriers (DCCs). Different DCCs that are transmitted to different cell sites are multiplexed in the frequency domain and arranged into multiple intermediate frequencies. There are some modulation formats that can be used to aid this efficient multiband multiplexing; among them, discrete multitone (DMT), discrete Fourier transform spread DMT (DFTS-DMT), and carrierless amplitude and phase modulation (CAP) [33.48–50] are good candidates with comparable spectral efficiencies. It is worth noting that the modulation format and speed can also be adjusted in real time for each band. Adjacent DCCs can be isolated with digital filters, so they do not need to be strictly synchronized. Because it utilizes a computationally efficient FFT algorithm, DMT has the fastest computing speed and the greatest flexibility. However, DMT suffers more from timing and frequency offsets than CAP. CAP may also be a better solution for uplink transmission, since it can be combined with digital filters with less out-of-band leakage.

On the other hand, to cope with different services and optimize the system performance of a next-generation radio access network, it is necessary to jointly implement all three technologies (A-RoF, D-RoF, and a functional split) together with dynamic switching and coordination. Figure 33.11 shows the basic concept of a software-controlled fronthaul architecture in which a flexible functional split is realized over an open air interface (OAI) platform, which is a real implementation of the LTE standard built on low-latency Linux kernels. On top of the OAI is a flexible fronthaul I (FH-I) split controller that can change the splitting option based on traffic conditions and delay requirements. In addition, an A-RoF option is integrated into the platform to support URLLC better with reduced costs and energy consumption at the RRY site. Using this testbed, a series of long-term tests are carried out to collect measurement results such as the throughput, delay, jitter, CPU load, and the power consumption at the RRU. The flexible split allows service providers to optimize their networks and systems to achieve the

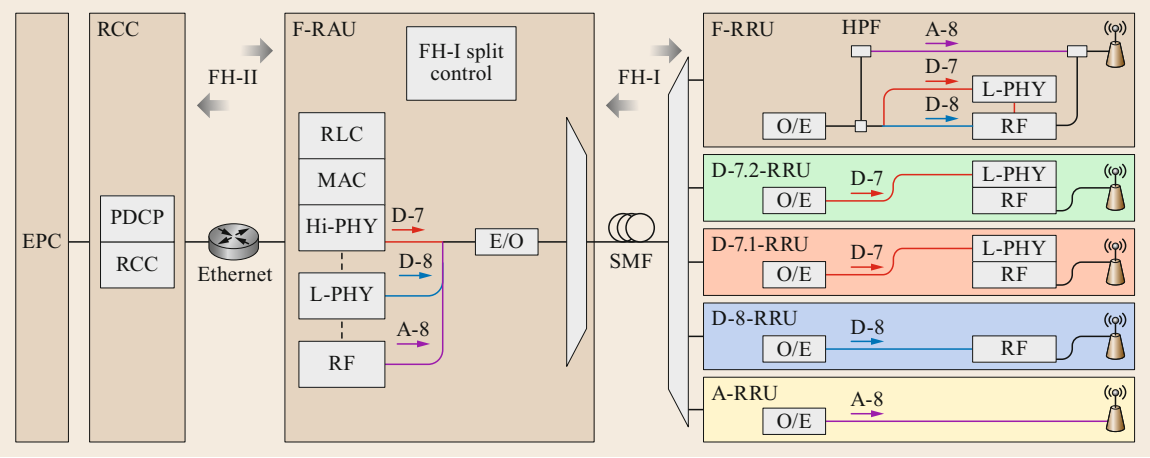


Fig. 33.11 Conceptual diagram of an adaptive functional split in a mobile fronthaul

best performance for different application scenarios, such as a low latency or a high bandwidth, while considering cost and energy efficiency. To enable the different functional split options to work smoothly together and improve reconfigurability, new technologies such as software-defined networking (SDN), functioning virtualization, and mobile edge computing will be considered for flexible mobile fronthaul architectures.

33.2.5 Free-Space Transmission Technologies for Mobile Fronthaul and Backhaul

When deploying 5G, the mobile backhaul and fronthaul are the major cost components, so it is very important to reduce the costs associated with them. When high-density small cells are deployed, large amounts of fiber will need to be deployed both indoors and outdoors. Considering the extremely high cost of fiber installation in urban areas, across long distances, and across geographic barriers, the use of wireless technologies to carry mobile fronthaul or backhaul traffic is crucial to the feasibility of a highly centralized network. Several techniques are discussed in this section: free-space optics (FSO), microwave technologies (including MMW), and LEB-based visible light communication (VLC) technologies.

FSO is an optical technology that utilizes a laser to send very-high-bandwidth data [33.51]. It is natively compatible with the other optical section of the backhaul/fronthaul network and can utilize mature and widely available optical transmitters, receivers, and amplifiers, which is a notable advantage over other wideband wireless connections using MMW or THz bands. Using either coherent or IM-DD-based modulation, a FSO link can easily carry multi-GHz data across long distances to hard-to-reach RRHs or BBUs. As it is a general optical-wireless transmission link, a FSO link is compatible with analog mobile fronthaul and other digital-based functionally split formats, and does not necessitate any modification to the mobile fronthaul front end. However, the optical signal is vulnerable to atmospheric attenuation (absorption, scattering, and turbulence) and other losses such as geometric, pointing, and optical losses in the link.

In MMW mobile fronthaul, a wireless overlay is placed over a densely populated 5G heterogeneous network containing MMW small-cell base stations. The link is used to greatly increase the network capacity at a reasonable cost and make it transparent to users. By increasing the wireless signal frequency to more than 30 GHz, the available channel bandwidth is significantly expanded to multi-GHz, which is capable of delivering the increased traffic generated by 5G networks. Seamless convergence of fiber and MMW mobile fronthaul and backhaul is feasible, with the MMW radio-over-fiber technology playing the important role of a relay. The signal delivered by the fiber is an analog waveform transmitted in the MMW band, meaning that complicated high-speed DSP and digital modulation to the relay node are not necessary. Also, the precise local oscillator used in wireless communication can be eliminated, as upconversion is carried out at the central site, making the fiber-wireless mobile fronthaul link more stable and robust to the environment. However, compared with low-frequency microwave links, MMW and THz links are more sensitive to weather conditions if deployed outdoors.

MMW, THz, and FSO have attracted attention due to their low cost, large bandwidths, high deployment flexibility, and fast deployment, although they have limited delivery distances due to their relatively high propagation loss and sensitivity to weather and atmospheric conditions. Fiber and wireless link convergence has attracted considerable attention because the typical distances for mobile fronthaul and backhaul can exceed 10 km. Most of this distance should still be covered by fiber. When there is an obstacle in the path, one or multiple wireless technologies can be deployed. A promising approach to further improve the wireless transmission quality, especially in the presence of adverse weather conditions, is to combine various techniques. Owing to the drastically different channel responses in FSO and MMW links, these complementary transmission characteristics are exhibited under different conditions. Therefore, a hybrid link with both FSO and MMW transmission should be a good solution for joint deployment. Several studies have reported hybrid MMW/FSO link designs, such as hardware or software switching schemes based on the availability of channels and joint bit-interleaved coding schemes between channels. To avoid the complexity of real-time feedback, the QAM level adaptive diversity combing technique (ADCT) can be used to combine the information from both channels and obtain the symbols with improved SNR [33.52].

The above scheme is mainly intended for outdoor links. For indoor coverage, wireless fronthaul links are also important due to the need for high-density small cells in open public spaces, such as open offices and libraries. We can use low-cost light-emitting diode (LED)-based VLC links in a spatially densified MFH with cell coordination. To provide high cell density in an indoor environment with a massive number of users, two tiers of RRHs are introduced. The first tier of RRHs, referred to as the master RRHs, is connected to

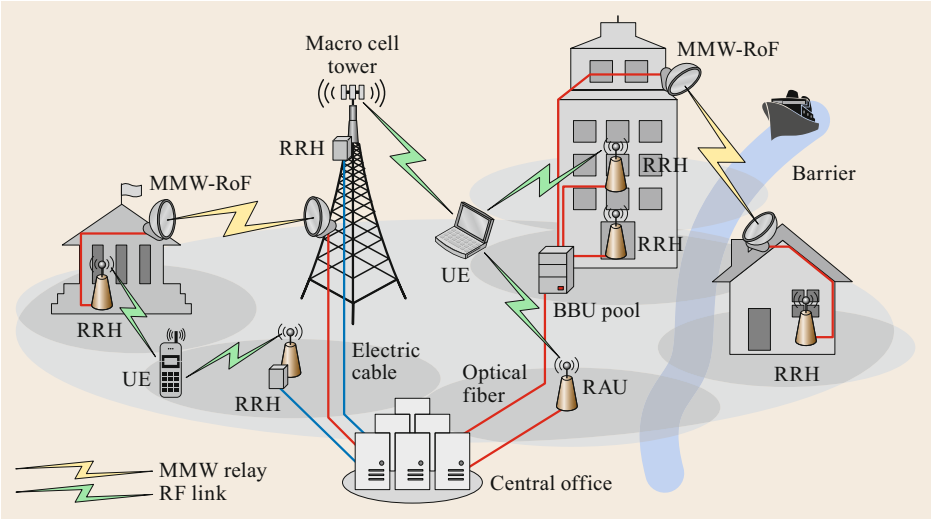


Fig. 33.12 MMW relay links in 5G mobile fronthaul

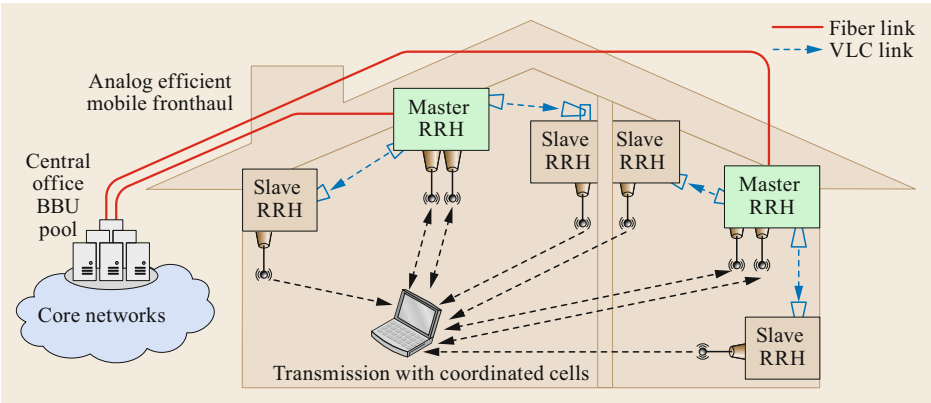


Fig. 33.13 VLC indoor relay links in 5G mobile fronthaul

the fiber MFH network through a direct fiber connection to the BBU pool in the CO. Signals are transmitted between master and slave RRHs by VLC links. Like the waveform in the fiber, the raw wireless streams are used for transmission, which also makes purely analog slave RRH possible without DSP. To compensate for the severe distortion in the LED transmitter, centralized pre-equalization and distributed pre-equalization can be implemented [33.53]. The spatial efficiency of the wireless network can be significantly increased to provide enhanced performance and increased data rates to a massive number of users.

33.2.6 All-Spectrum Fiber-Wireless Access Technologies

5G has specific capacity, latency, and reliability requirements. However, it is very hard to fulfill all of these requirements simultaneously using currently available technology and components. Similarly, multiple technologies and transmission bands can be utilized for user access to serve various user types and scenarios.

The high capacity needed in eMBB requires high-bandwidth channels with enhanced power efficiency. High-frequency channels are attractive in this context as this region of the spectrum is less crowded. However, the penetration and reflection capabilities of these MMW, THz, and FSO channels are poor due to their short wavelengths and the antenna design. As they are directional beams, most transmissions require line-of-sight operation, which limits the service range and coverage of each small cell. It is also hard to provide users (such as those on a high-speed train) with high mobility using such a scheme because physical and electrical beam tracking schemes are problematic to implement and control for short-wavelength beams. In addition, the service quality in such schemes is sensitive to the weather conditions, making it difficult to

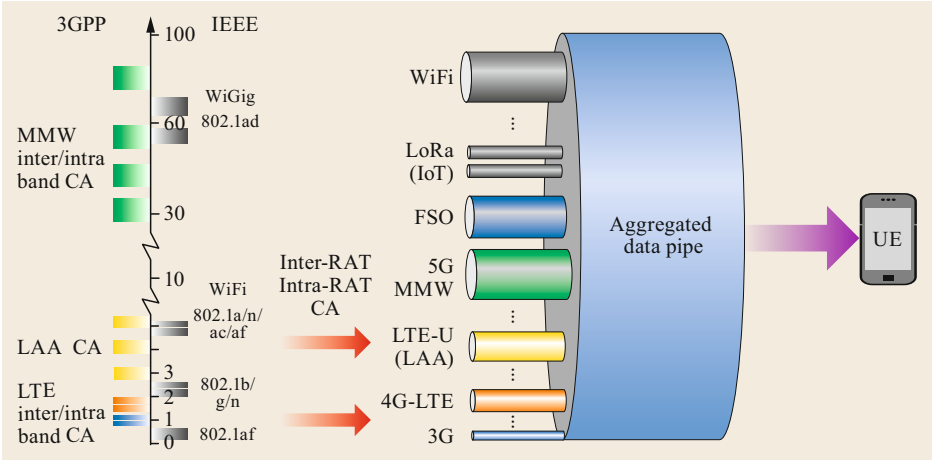


Fig. 33.14 Carrier aggregation and all-spectrum integration for an aggregated data pipe

reliably serve users located outdoors at long distances. As a result, it is more feasible to use high-frequency wireless access technology as the data pipe in eMBB for stationary users located indoors. The availability and stability of transceivers vary among technologies that use high frequencies. FSO transmitters are simple and natively compatible with fiber fronthaul networks. Therefore, it is feasible to use FSO in the downlink with beamforming and tracking technology. However, this requires precise laser temperature control, making it hard to integrate such a system into user equipment, and it also consumes a lot of power. FSO receivers are stable, making it possible for users to incorporate a PD as the FSO receiver. On the other hand, MMW-band transceivers are a mature technology; integrated circuits and a RF front end are available. Hence, it is possible to embed a MMW-band transceiver into user equipment and employ it as the uplink transmitter.

It is unwise to use high-frequency channels to deliver control signals, as these channels suffer from unreliable transmission quality. It is better to use a low-frequency (below 6 GHz) channel due to its extensive coverage and better penetration capabilities. Since the data rate for the control channel is low, each band can support a large number of users, facilitating multiuser synchronization. Signals broadcast by macro cells can be used for multi-small-cell coordination. The broad coverage of low-frequency bands makes it feasible to use them for mMTC.

It is necessary to precisely track the positions of user devices for some applications and for beamforming. If the infrastructure supports high-density cells, the positions of user devices tend to be tracked more precisely. Considering the cost and ease of deployment, it is impractical to deploy high-density MMW or FSO cells. The use of low-frequency cells as anchor nodes is a waste of resources given the resulting severe ICI. Instead, it is better to use LED-based VLC technology as indoor anchors. The transmitter cost is low, with power lines or fiber used for the fronthaul links. With a smart and adaptive sectored architecture, high-density LEDs can be utilized for positioning services in a 5G network. The spectral resources can be reused in the IF domain, so the fronthaul architecture can be reused by multiple LEDs, meaning that it is not necessary to modify the network significantly. The position of a user can be determined approximately by cell ID mapping and more precisely by collecting the information from multiple nearby LEDs with various IFs and then applying a location solver based on machine learning methods [33.54], as shown in Fig. 33.15. Compared with beam sweeping, a LED-based location solution will significantly reduce the delay in the initial connection.

Figure 33.16 shows the complete architecture of the all-spectrum wireless access system obtained when all of the spectral resources and technologies are combined, fulfilling all of the requirements of the 5G standards. All types of devices are supported simultaneously, such as bidirectional data-hungry smartphones and laptops, smart sensors and meters, and mission-critical car-related applications. In addition, signal generation and reception DSP can be executed in the centralized architecture via software-defined radio (SDR) and NFV. Network control and optimization based on machine learning and artificial intelligence can be applied to enable the system to adapt to traffic patterns and the distribution and requirements of users. This centralized architecture allows joint transmission and interspectrum coordination, making it easy for service providers to implement and operate the system.

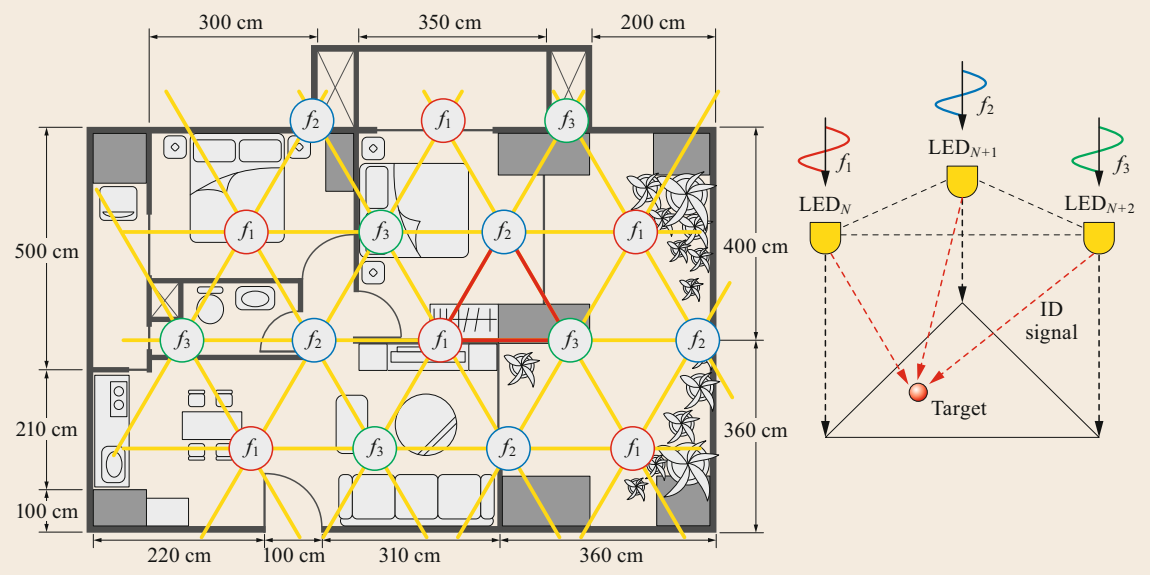


Fig. 33.15 LTE-based small cells and cell grouping for precise positioning

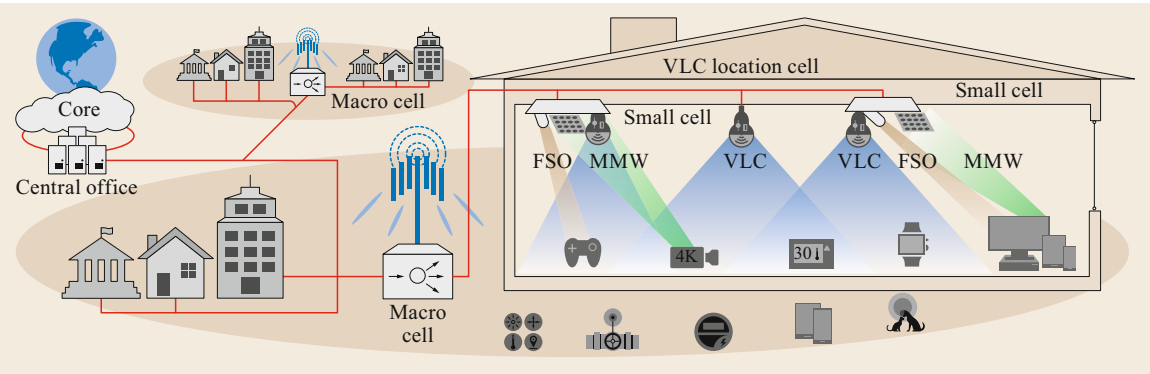


Fig. 33.16 Coordinated all-spectrum wireless access system that serves all devices

33.3 Advanced Digital Signal Processing in 5G Converged Networks

Because of its low bandwidth efficiency, rumors have arisen that digital radio-over-fiber technologies may not be suitable for 5G-compatible high-capacity mobile backhaul or tier-II fronthaul. However, this is not the case for the short-range and latency-stringent tier-I mobile fronthaul network due to its high resistance to nonlinear distortion, its simple and straightforward DSP architecture, and its compatibility with different formats. That said, we cannot ignore the bottleneck induced by its low bandwidth efficiency, which could seriously limit the implementation of enhanced mobile broadband in the future. Thus, it is very important to consider ways to improve the bandwidth efficiency. Fortunately, there are a variety of tools that can be used for this purpose. At the radio interface, functional splitting could help to release a huge amount of bandwidth. However, in many cases—especially for the signal after MIMO precoding—digital quantization is still needed. During DSP, resampling and data compression can help to save more bandwidth. Furthermore, at the hardware and system level, applying advanced modulation formats and coherent optics could further improve the bit information carried by each symbol in the optical signal. On the other hand, nonlinear compensation

methods are crucial to increasing the dynamic range of the system and mitigating nonlinear distortions of analog signals at the fiber-wireless link. Meanwhile, fast statistical estimation and data compression based on the Lloyd algorithm have been shown to reduce the number of quantization digits in a D-RoF-based mobile fronthaul with low complexity and high quality.

33.3.1 Mitigation of Nonlinear Impairment in Fiber–Wireless Access Networks

In view of the recent exponential increases in the number of carriers and wireless bandwidth, the bandwidth required in mobile fronthaul is extremely high. Two solutions can effectively improve this spectral efficiency in the fronthaul fiber link: digital fronthaul schemes with advanced optical modulation (including high-level pulse amplitude modulation, PAM, or DMT) and analog radio-over-fiber mobile fronthaul. Both of these technologies use multicarrier modulation in the fiber link, which generates an increased peak-to-average power ratio. This feature limits the output power as the linear range of optical transceivers is limited.

As a result, nonlinearity mitigation is a crucial aspect of the fiber link [33.55]. Nonlinear precompensation using soft companding or hard clipping techniques can be used to reduce the peak-to-average power ratio (PAPR) of the signal generated. In the receiver, postcompensation is implemented to recover the original signal. The polynomial or Volterra architecture is usually applied with well-trained parameters to achieve this. The use of a neural network to compensate for the nonlinearity of the link using a multinode multilayer structure is a hot topic. Neither precompensation nor postcompensation will modify the embedded information, so it can be applied universally to general multicarrier modulation without the need to modify 5G user devices. In addition, machine-learning-based multilayer neural networks can be leveraged in the receiver DSP for equalization in order to compensate for the nonlinearity in the fiber link and transceivers, and improved performance is expected to support higher modulation.

Another approach is to reduce the inherent PAPR by slightly changing the format or placing of different carriers, or by modifying the symbol carried by subcarriers. Some coding schemes to reduce the PAPR have been proposed. When in-phase signals are added, they produce a peak power that is many times the average power. Therefore, substantial PAPR reduction can be obtained when measures are taken to reduce the probability that the signals will be in phase. This is the fundamental principle of coding schemes, and requires that both the transmitter and the receiver are modified and receive prior information. Partial transmission sequence (PTS) and selective mapping (SLM) are other techniques that can be applied to reduce the PAPR. In a typical PTS approach, the input data block is partitioned into disjoint subblocks using either adjacent partitioning, interleaved partitioning, or random partitioning. However, there are two important issues that must be resolved for PTS: its high computational complexity and the overhead from the optimal phase factors. Similarly, the input data sequences are multiplied by each of the phase sequences to generate alternative input symbol sequences in SLM. Both PTS and SLM are important probabilistic schemes for PAPR reduction. Tone reservation (TR) and tone injection (TI) are another two efficient techniques. Here, the key idea is that both transmitter and receiver reserve a subset of tones for generating PAPR reduction signals. These tones are not used for data transmission, only for PAPR reduction in the transceiver DSP. Note that user devices must be modified to some extent when using the schemes mentioned above.

A better approach that fits with analog mobile fronthaul is to use phase predistortion in the multiband OFDM system. This involves adaptively tuning the phase of each subband in the granularity of time-domain OFDM frames to neutralize the peaks in the time domain, hence reducing the combined PAPR in the fiber link [33.27]. Moreover, the fiber receiver and wireless device do not need to be modified in this approach due to the inherent channel estimation and equalization functions.

33.3.2 Operational Principles of Statistical Data Compression in D-RoF Systems

As discussed earlier, its high compatibility with different data formats, robustness against nonlinear degradation and channel penalties, and the possibility of error-free operation make compressed D-RoF an attractive candidate for future mobile fronthaul networks, especially tier-I fronthaul. Various tools could potentially be applied to improve the bandwidth efficiency of the mobile fronthaul network, as shown in the flow diagram of the whole system depicted in Fig. 33.17. At the radio interface, functional splitting can be utilized to release significant system capacity (depending on the split option selected). In the digital signal processing domain, signal resampling and data compression can help to reduce the bandwidth further. Also, at the hardware and system level, an advanced modulation format and even coherent optics can be implemented to achieve higher spectral efficiency, saving bandwidth for the mobile fronthaul. In the following, we mainly focus on digital compression, which is relatively easy to implement at low cost.

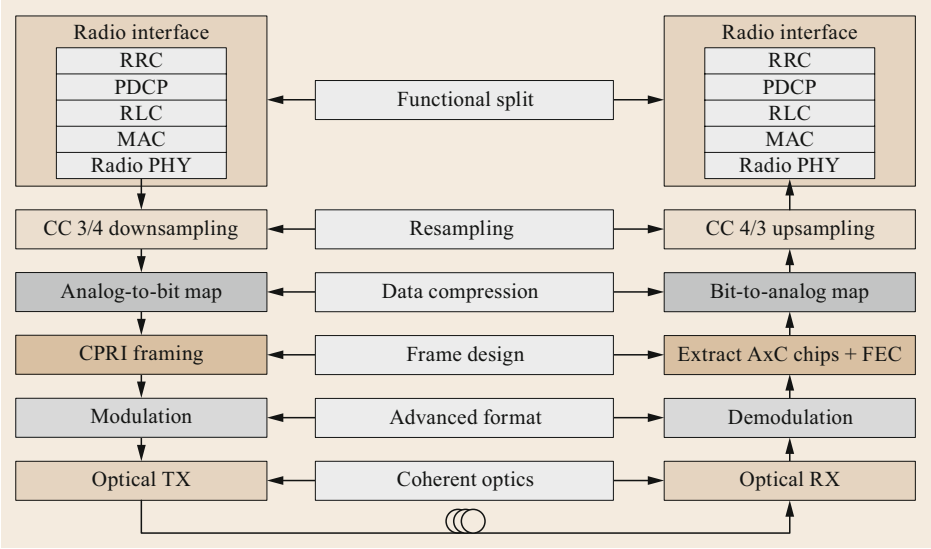


Fig. 33.17 Approaches used to improve bandwidth efficiency in different blocks of a tier-II mobile fronthaul system

Before we discuss the operational principles of data compression in D-RoF systems, Fig. 33.18a, b provides a quick review of some statistical properties of OFDM signals. The amplitude of an OFDM signal follows a Gaussian distribution characterized by the expectation μ and the standard deviation σ, as shown in Fig. 33.18a. On the other hand, the modulus (i.e., the absolute value of the amplitude of the signal) will follow a folded Gaussian distribution, as shown in Fig. 33.18b. It is apparent from the cumulative distribution function (CDF) shown in Fig. 33.18c that the modulus is initially highly nonlinearly distributed and the quantization noise is significantly amplified in some parts of the signal modulus. Upon applying a companding transform to compress and expand parts of the CDF of the signal, the amplitude distribution is linearized, as shown in Fig. 33.18d, and the quantization noise within different quantization sections becomes uniformly distributed. According to the minimum mean square error criterion, the overall quantization noise will be minimized with this uniform quantization noise distribution, which means that fewer quantization levels can be used to achieve the same quantization noise level or that the quantization noise is suppressed with the original number of quantization levels. Multiple methods can be used to achieve companding-based data compression. Fitting-based nonlinear quantization (FBNQ) is among the most mature of those methods [33.56]. In FBNQ, statistical fitting is performed, with the system estimating μ and σ of an OFDM baseband component by calculating the mean $\hat{\mu}$ and the standard variance $\hat{\sigma}$ in a statistical analysis of a large number of samples. The companding functions can then be obtained to compand the CDF of the amplitudes in the real and imaginary parts of the samples. The CDF of the modulus for the real part of an OFDM baseband component is shown before and after companding in Figs. 33.18c and d, respectively. Based on the input and output signal modulus, a look-up table from U digits to V digits ($V < U$) can be generated to accelerate the execution of the digital compression algorithms. The operational process for a digital RoF system with data compression is shown in Fig. 33.19. The analog wireless OFDM waveform is first sampled to get a discrete signal. Each data sample is then quantized to 2^U levels and converted into a binary AxC chip with U digits. In the CPRI standard, $U = 15$. Using a compression process, each AxC chip is mapped from U digits to V digits, thus improving the bandwidth efficiency by $(U - V)/V \times 100\%$.

The procedures involved in using FBNQ in mobile fronthaul are shown in Fig. 33.20a. The I/Q components of a baseband analog LTE signal are quantized by a U-bit analog-to-digital converter. A large number of samples are used for statistical analysis to estimate the probability density function (PDF) of the Gaussian distribution. At the transmitter, after obtaining the fitted Gaussian distribution and look-up table, U-digit antenna component (AxC) chips are efficiently mapped into compressed digital AxC chips with V bits. On the other hand, at the receiver site, a reverse V-to-U-bit look-up table is used to decompress the V-digit samples back to U digits.

However, one problem with using the FBNQ scheme is that the fitting process requires a large number of samples for statistical analysis, which is computationally inefficient and leads to considerable latency in the whole system. The operational procedure for a different method called fast statistical estimation

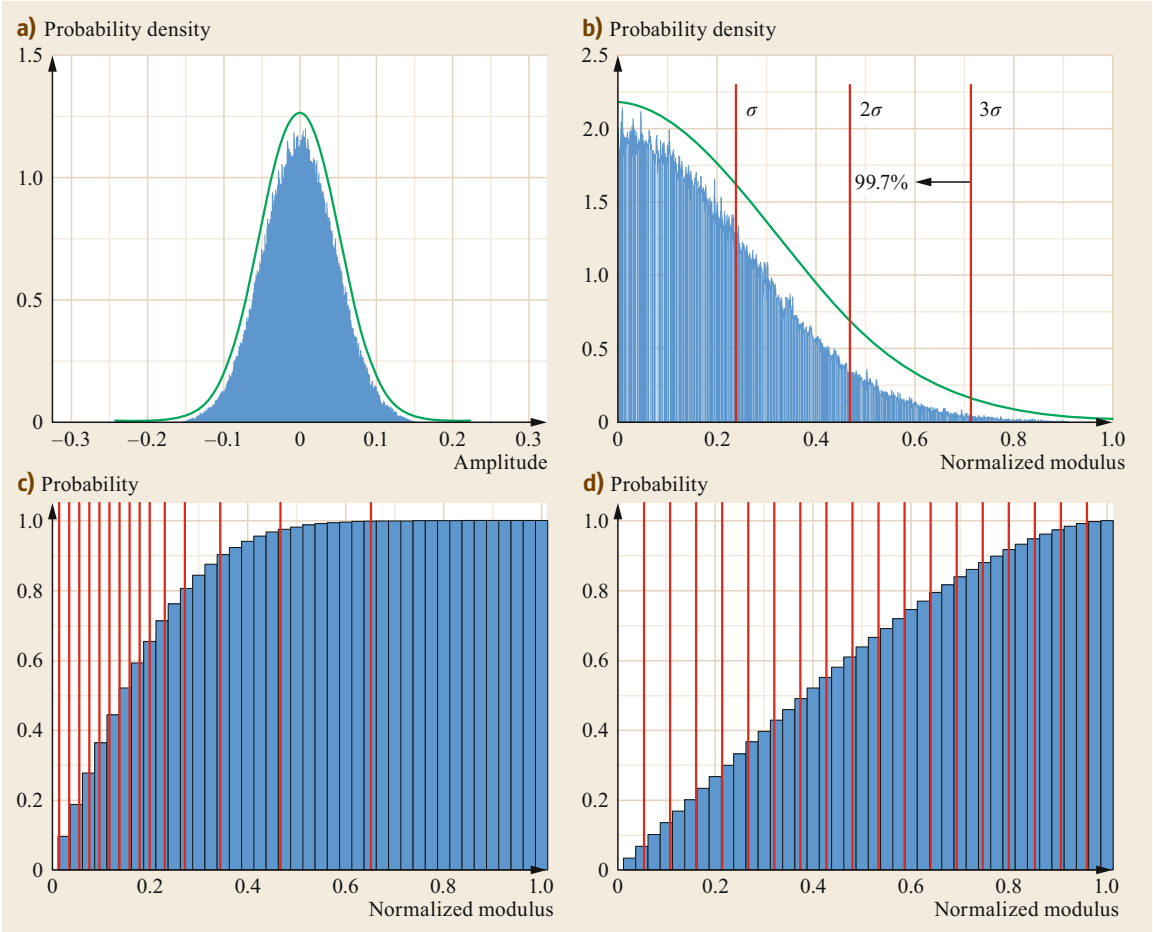


Fig. 33.18a–d Probability distribution functions of **(a)** the amplitude and **(b)** the modulus of a LTE-like OFDM signal; **(c,d)** cumulative distribution functions of the OFDM signal modulus before and after companding, respectively (the *red lines* depict the quantization levels, respectively)

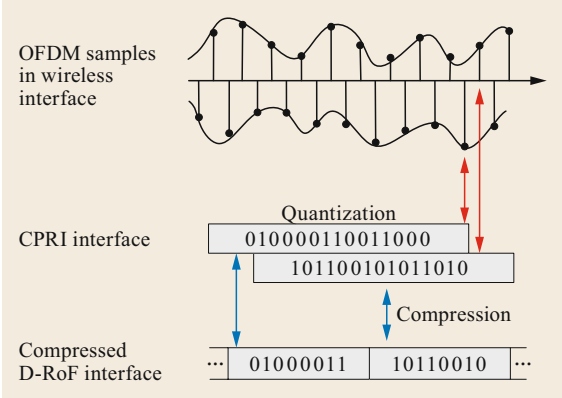


Fig. 33.19 Operational principle of data compression in digital mobile fronthaul

(FSE) [33.57] is shown in Fig. 33.20b. The first step in this method is to let the signal pass through a direct current (DC) block to eliminate the DC component, which enables the modulus of the signal to be readily obtained regardless of its DC variation. The PDF of the modulus for a LTE-like OFDM signal is depicted in Fig. 33.18b. It exhibits a folded Gaussian distribution. Compared with that of PDF of amplitude like Gaussian distribution in Fig. 33.18a, the modulus of the signal is always non-negative and the zero level is very precise and stable. In FSE, the PDFs of the corresponding Gaussian distribution and folded Gaussian distribution are truncated at the edges of the intervals $[-K\sigma, K\sigma]$ and $[0, K\sigma]$ respectively. After truncating and normalizing the signal samples, the statistical value of σ can immediately be expressed as $1/K$. The companding functions can then

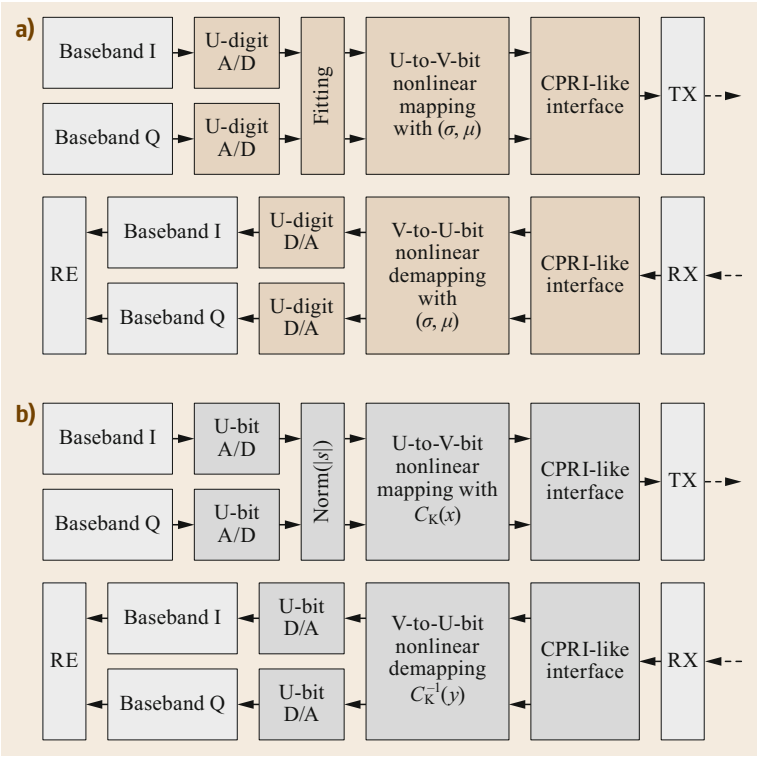


Fig. 33.20a,b DSP flow diagrams of **(a)** FBNQ- and **(b)** FSE-based data-compression methods, respectively

be constructed without the need to incorporate a sophisticated statistical fitting or training process. Recall that an important criterion in a zero-mean Gaussian distribution is the so-called *three-sigma law*, which guarantees that nearly 99.7% of the samples fall within the interval from -3σ to 3σ. Thus, assuming there are enough samples, we may expect that the estimated truncated and folded Gaussian distribution will be very close to the original untruncated and folded Gaussian distribution when K approaches 3 because more than 99% of the possible modulus distribution situations have been considered. Moreover, this precise statistical estimation scheme is very efficient since it does not require the implementation of a computationally complex statistical fitting process.

Figure 33.20b depicts the operation stacks for FSE-based data compression in a D-RoF system. It should be noted that many function blocks are similar to those of FBNQ, as shown in Fig. 33.20a. Nevertheless, there is no need to use a fitting process, and the key part of the DSP stacks is replaced with a fast statistical estimation algorithm. Compression and decompression look-up tables are generated using the companding function $y = C(x)$ and the inverse companding function $x = C^{-1}(y)$, respectively. From the simulation results in [33.57], the best K value with the lowest quantization error was found to be around 2.7. Moreover, it is worth noting that the first digit of each compressed AxC chip is the sign (+ or −) of each amplitude. In fact, only $(V-1)$ bits are used to generate the U-to-V-bit and V-to-U-bit look-up tables. The companding function of the modulus-based fast statistical estimation is

$$C(x) = \frac{K\sigma}{\Delta}\mathrm{erf}\left(\frac{x}{\sqrt{2}\sigma}\right). \tag{33.1}$$

Following the same analytical methods, the companding function of the fast statistical estimation applying to the power of the sample, $x = |s|^2$, $s \in [-K\sigma, K\sigma]$, can be derived, and is shown in (33.2). In this situation, x will present a truncated powered Gaussian (TPG) distribution. Similarly, the companding transform function can be written as

$$C(x) = \sqrt{\left(\frac{K^2\sigma^2}{1-2\Phi(-K)}\right)\mathrm{erf}\left(\frac{\sqrt{x}}{\sqrt{2}\sigma}\right)}. \tag{33.2}$$

Aside from Gaussian-distribution-based statistical estimation methods, there are also some mature companding methods that have already been used to encode and

decode acoustic signals in telephony wired systems, including μ-law and A-law companding. The μ-law companding transform function is

$$C(x) = \mathrm{sgn}(x)\frac{\ln(1+\mu\,|x|)}{\ln(1+\mu)},\quad x\in[-1,+1]\,, \tag{33.3}$$

while the A-law companding function can be expressed in a piecewise manner as

$$C(x) = \begin{cases} \mathrm{sgn}(x)\frac{A|x|}{1+\ln A}, & 0\le|x|\le\frac{1}{A}\,, \\ \mathrm{sgn}(x)\frac{1+\ln(A|x|)}{1+\ln A}, & \frac{1}{A}\le|x|\le 1\,. \end{cases} \tag{33.4}$$

33.3.3 Data Compression Based on the Lloyd Algorithm

A data compression method called the Lloyd algorithm [33.58], which is based on statistical estimation using the minimum mean-square error (MMSE) criterion, has been proposed for use in digital mobile fronthaul. It is format agnostic and can be used to compress non-Gaussian modulation formats, such as single-carrier frequency-division multiplexing (SC-FDM). The Lloyd algorithm first divides the PDF of the signal amplitude into multiple segments with boundaries defined by the thresholds $[t_i, t_{i+1}]$, as shown in Fig. 33.21. After quantization, the amplitudes within each segment are quantized as level l_i. To minimize the MSE between the quantized and original signals, $\{t_i, t_{i+1}\}$ and l_i are related by

$$l_i = \frac{\int_{t_i}^{t_{i+1}} xf(x)\mathrm{d}x}{\int_{t_i}^{t_{i+1}} f(x)\mathrm{d}x} \tag{33.5}$$

and

$$t_i = \frac{(l_i + l_{i+1})}{2}\,, \tag{33.6}$$

where $f(x)$ is the PDF of the signal amplitude. However, if the traditional Lloyd method is applied to calculate the positions of all the quantization levels, the complexity becomes excessive. To achieve a good trade-off between computational complexity and accuracy, a relaxed Lloyd (R-Lloyd) algorithm that only uses the Lloyd method to calculate the major quantization levels has been proposed. In this approach, some minor quantization levels are uniformly inserted between adjacent major quantization levels. A flow diagram of the R-Lloyd algorithm is shown in Fig. 33.20c. First, uniformly separated quantization thresholds $(t_{1,0}, t_{2,0}, \ldots, t_{2P+1,0})$, where P is the number of major digits, are defined. Then the quantization levels $(l_{1,1}, l_{2,1}, \ldots, l_{2P,1})$ are calculated using $(t_{1,0}, t_{2,0}, \ldots, t_{2P+1,0})$ via (33.5). Subsequently, $(t_{1,1}, t_{2,1}, \ldots, t_{2P+1,1})$ are similarly obtained using $(l_{1,1}, l_{2,1}, \ldots l_{2P,1})$ via (33.6). The quantization thresholds and levels are then repeatedly updated based on their counterparts' previous values until the iteration index reaches M (as indicated by the loop in Fig. 33.20c). During this process, both the quantization levels and the thresholds gradually converge to positions that minimize the MSE between the original and quantized signals.

Given a compressed AxC chip with V digits, the R-Lloyd method is applied to reduce the complexity of the traditional Lloyd algorithm. In this case, 2^P major levels from among the 2^V levels in total are first computed using the Lloyd algorithm. $2^{(V-P)}$ minor quantization levels are also uniformly interpolated

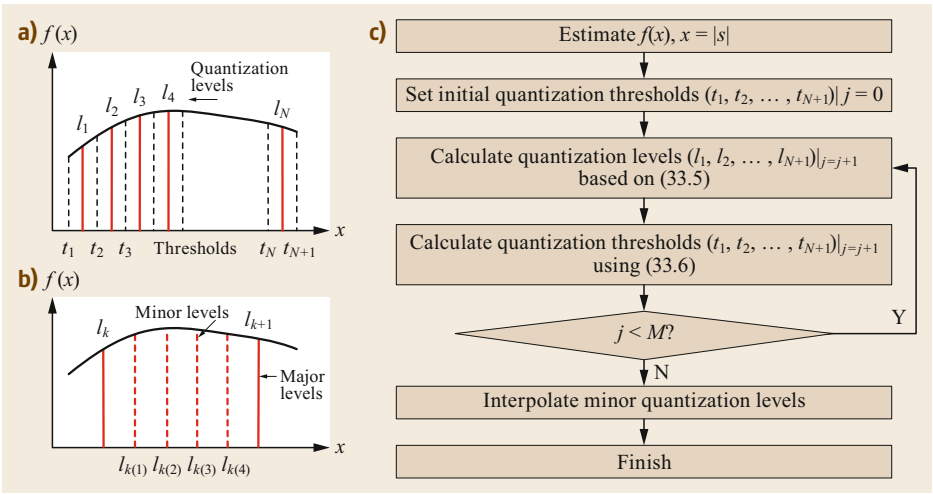


Fig. 33.21 (a,b) Operational principles of data compression based on Lloyd's algorithm; **(c)** flow diagram of the compression process

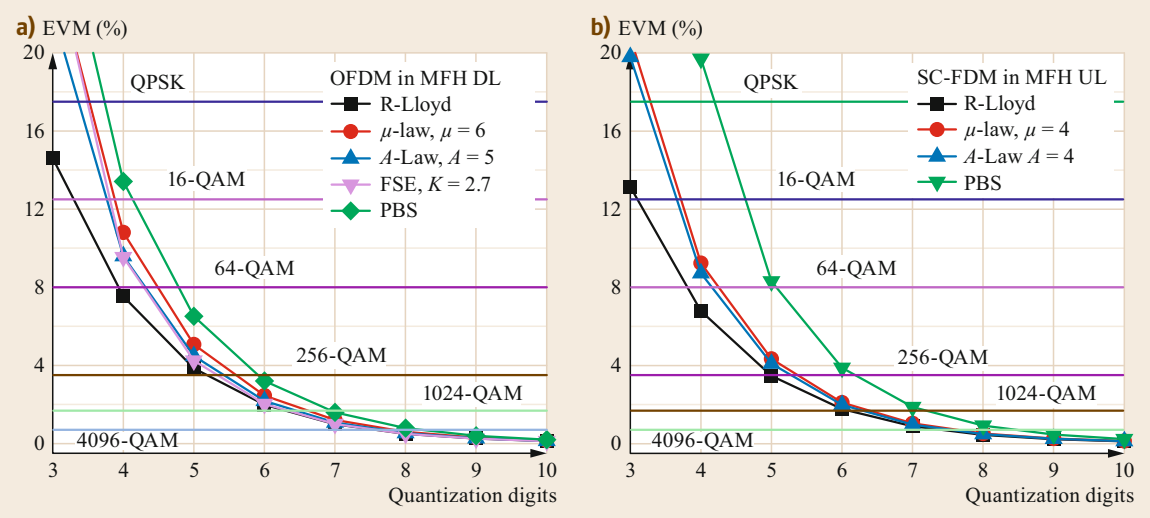


Fig. 33.22 Plots of EVM versus the number of quantization digits when the Lloyd method is used for **(a)** OFDM and **(b)** SC-FDM, respectively

in the interval $[l_i, l_{i+1}]$. The values of P and V are selected according to the trade-off between quantization accuracy and convergence speed. The change in the OFDM error vector magnitude (EVM) convergence speed when the number of major digits used in 15-to-8-digit compression ($U = 15$ and $V = 8$) is varied was studied in [33.58]. It was verified that a faster convergence speed can be achieved with fewer iterations by employing fewer major digits. When applying four major digits ($P = 4$), a converged EVM value was obtained with 100 iterations. To obtain a good balance of complexity and accuracy, it is recommended that four or five major digits should be used. Plots of EVM versus the number of quantization digits for OFDM and SC-FDM radio signals after 15-to-8-digit compression and decompression are shown in Fig. 33.22a, b respectively. It is clear from these plots that, compared with other methods such as μ-law, A-law, or FSE, the R-Lloyd algorithm yields improved EVM performance, especially with a small number of quantization digits. Some methods such as FSE [33.57] are especially designed for OFDM signals with Gaussian-distributed amplitudes. They do not work with non-Gaussian signals such as SC-FDM. However, the performance of the Lloyd algorithm is independent of the signal's statistical properties. As shown in Fig. 33.21b, good EVM is still obtained when using the Lloyd algorithm to compress SC-FDM radio signals. The required EVM thresholds of 64-, 256-, 1024-, and 4096-QAM are set to 8, 3.5, 1.68, and 0.7%, respectively [33.59]. Some key parameters, including the net bits of the original symbols, the minimum number of bits required after compression, and the standardized number of bits in CPRI, are summarized in Table 33.5. It is worth noting that when lower-order modulation formats (e.g., quadrature phase shift keying, QPSK, or 16-QAM) are applied, the minimum number of digits can be significantly reduced with a higher compression ratio. Nevertheless, when using more advanced modulation formats such as 512-QAM or 1024-QAM, the number of digits in every sample after compression approaches the original number of net bits. Thus, the application of an efficient compression scheme can lead to a high D-RoF MFH bandwidth efficiency that is comparable to that of the PHY-I split scheme ([33.60], see also Chap. 29). It is worth noting that in the PHY-I split option, functions such as the inverse fast Fourier transform (IFFT) are shifted to the remote radio access units, meaning that MFH transmits the net bits of each symbol before the IFFT block. Used in combination with cost-effective devices in the distributed radio access units, D-RoF MFH with data compression could become competitive with functional split (FS) in some usage scenarios.

Table 33.5 Minimum number of bits needed to support different modulation formats

Format	Net bits	Minimum bits after compression	Bits in CPRI
QPSK	2	8 (2×4)	30 (2×15)
16-QAM	4	8 (2×4)	30
64-QAM	6	10 (2×5)	30
256-QAM	8	12 (2×6)	30
1024-QAM	10	14 (2×7)	30

33.4 Summary

In this chapter, we have discussed how advanced fiber-optic technologies and fiber-wireless convergence can contribute to the development of next-generation radio access networks that support 5G NR. We briefly introduced 5G, discussing recent progress in 5G NR standardization and how the latest photonic technology could be used to enhance 5G systems. The current status of and progress in the following fields were reviewed:

(a) Integrated RF, mm-wave, and lightwave systems as well as microwave photonics
(b) Fiber-wireless integration and networking in next-generation fronthaul systems
(c) Advanced digital signal processing in 5G converged networks.

By combining and utilizing the advantages of various fiber and wireless technologies, it is possible to improve the system capacity and link throughput of the radio access network by enhancing the spectral and spatial resource utilization rate. Increased reliability and low latency can also be achieved by combining and coordinating the A-RoF and D-RoF options in different application scenarios. We can also expect emerging technologies such as flexible functional splitting, machine-learning-based DSP, and all-spectrum communication to reshape 5G systems and facilitate the further evolution of mobile data networks in the future.

References

33.1 P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, J. Yao: 5G on the horizon: Key challenges for the radio-access network, IEEE Veh. Technol. Mag. **8**(3), 47–53 (2013)

33.2 E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, Y. Selén, J. Sköld: 5G wireless access: Requirements and realization, IEEE Commun. Mag. **52**(12), 42–72 (2014)

33.3 O.N. Yilmaz, Y. Wang, N.A. Johansson, N. Brahmi, S.A. Ashraf, J. Sachs: Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case. In: *IEEE Int. Conf. Commun. Worksh. (ICCW)* (2015) pp. 1190–1195

33.4 J.G. Andrews, S. Buzzi, W. Choi, S.V. Hanly, A. Lozano, A.C. Soong, J. Zhang: What will 5G be?, IEEE J. Sel. Areas Commun. **32**(6), 1065–1082 (2014)

33.5 S. Chen, Z. Jian: The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication, IEEE Commun. Mag. **52**(5), 36–43 (2014)

33.6 T.S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G.N. Wong, J.K. Schulz, M. Samimi, F. Gutierrez: Millimeter wave mobile communications for 5G cellular: It will work!, IEEE Access **1**, 335–349 (2013)

33.7 Qualcomm: Exploring 5G new radio: Use cases capabilities & timeline, White paper, https://www.qualcomm.com/media/documents/files/heavy-reading-whitepaper-exploring-5g-new-radio-use-cases-capabilities-timeline.pdf (2016)

33.8 W. Roh, J.Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, F. Aryanfar: Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results, IEEE Commun. Mag. **52**(2), 106–113 (2014)

33.9 S.M. Razavizadeh, M. Ahn, I. Lee: Three-dimensional beamforming: A new enabling technology for 5G wireless networks, IEEE Signal Process. Mag. **31**(6), 94–101 (2014)

33.10 H. Bogucka, K. Paweł, K. Adrian: Dynamic spectrum aggregation for future 5G communications, IEEE Commun. Mag. **53**(5), 35–43 (2015)

33.11 X. Ge, S. Tu, G. Mao, C.X. Wang, T. Han: 5G ultra-dense cellular networks, IEEE Wirel. Commun. **23**(1), 72–79 (2016)

33.12 V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfrojd, T. Svensson: The role of small cells, coordinated multipoint, and massive MIMO in 5G, IEEE Commun. Mag. **52**(5), 44–51 (2014)

33.13 I. Chih-Lin, C. Rowell, S. Han, Z. Xu, G. Li, Z. Pan: Toward green and soft: A 5G perspective, IEEE Commun. Mag. **52**(2), 66–73 (2014)

33.14 H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, A. Ghosh: Machine-type communications: Current status and future perspectives toward 5G systems, IEEE Commun. Mag. **53**(9), 10–17 (2015)

33.15 L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, Z. Wang: Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends, IEEE Commun. Mag. **53**(9), 74–81 (2015)

33.16 F. Lu, M. Xu, L. Cheng, J. Wang, J. Zhang, G.-K. Chang: Non-orthogonal multiple access with successive interference cancellation in millimeter-wave radio-over-fiber systems, J. Lightwave Technol. **34**(17), 4179–4186 (2016)

33.17 P. Popovski: Ultra-reliable communication in 5G wireless systems. In: *IEEE Int. Conf. 5G Ubiquitous Connect. (5GU)* (2014) pp. 146–151

33.18 F. Lu, L. Cheng, M. Xu, J. Wang, S. Shen, G.-K. Chang: Orthogonal chirp division multiplexing in millimeter-wave fiber-wireless integrated systems for enhanced mobile broadband and ultra-reliable communications. In: *Proc. Opt. Fiber Commun. Conf.* (2017), paper Th4E-5

33.19 C. Li, J. Jiang, W. Chen, T. Ji, J. Smee: 5G ultra-reliable and low-latency systems design. In: *Proc. Eur. Conf. Netw. Commun. (EuCNC)* (2017), https://doi.org/10.1109/EuCNC.2017.7980747

33.20 V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfrojd, T. Svensson: The role of small cells, coordinated multipoint, and massive MIMO in 5G, IEEE Commun. Mag. **52**(5), 44–51 (2014)

33.21 R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.P. Mayer, L. Thiele, V. Jungnickel: Coordinated multipoint: Concepts, performance, and field trial results, IEEE Commun. Mag. **49**(2), 102–111 (2011)

33.22 L. Cheng, M.M. Gul, F. Lu, M. Zhu, J. Wang, M. Xu, X. Ma, G.-K. Chang: Coordinated multipoint transmissions in millimeter-wave radio-over-fiber systems, J. Lightwave Technol. **34**(2), 653–660 (2016)

33.23 N. Rupasinghe, I. Güvenç: Licensed-assisted access for WiFi-LTE coexistence in the unlicensed spectrum. In: *IEEE Globecom Worksh.* (2014) pp. 894–899

33.24 F. Lu, M. Xu, L. Cheng, J. Wang, S. Shen, H.J. Cho, G.-K. Chang: Adaptive digitization and variable channel coding for enhancement of compressed digital mobile fronthaul in PAM-4 optical links, J. Lightwave Technol. **35**(21), 4714–4720 (2017)

33.25 M. Xu, F. Lu, J. Wang, L. Cheng, D. Guidotti, G.-K. Chang: Key technologies for next-generation digital RoF mobile fronthaul with statistical data compression and multiband modulation, J. Lightwave Technol. **35**(17), 3671–3679 (2017)

33.26 F. Lu, M. Xu, L. Cheng, J. Wang, S. Shen, J. Zhang, G.-K. Chang: Sub-band pre-distortion for PAPR reduction in spectral efficient 5G mobile fronthaul, IEEE Photonics Technol. Lett. **29**(1), 122–125 (2017)

33.27 F. Lu, Y.-C. Chi, M. Xu, L. Cheng, J. Wang, C.-T. Tsai, G.-R. Lin, G.-K. Chang: Cost-effective bi-directional mobile fronthaul employing WRC-FPLD for beyond LTE-advanced services. In: *Proc. Opt. Fiber Commun. Conf.* (2016), paper TU2B.5

33.28 F. Lu, M. Xu, L. Cheng, J. Wang, S. Shen, C. Su, G.-K. Chang: Efficient mobile fronthaul serving massive MIMO new radio services using single-IF with sample-wise TDM for reduced RRH complexity and ultra-low latency. In: *Proc. Opt. Fiber Commun. Conf.* (2017), paper Th3A-4

33.29 V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, T. Svensson: The role of small cells, coordinated multipoint, and massive MIMO in 5G, IEEE Commun. Mag. **52**(5), 44–51 (2014)

33.30 Qualcomm Research: *Hyper-Dense Small Cell Deployment Trial in NASCAR Environment* (Qualcomm, San Diego 2014)

33.31 China Mobile: *C-RAN: The Road Towards Green RAN*, White paper (CMCC, Hong Kong 2011)

33.32 K. Miyanabe, K. Suto, Z.M. Fadlullah, H. Nishiyama, N. Kato, H. Ujikawa, K.-I. Suzuki: A cloud radio access network with power over fiber toward 5G networks: QoE-guaranteed design and operation, IEEE Wirel. Commun. **22**(4), 58–64 (2015)

33.33 M. Peng, Y. Li, Z. Zhao, C. Wang: System architecture and key technologies for 5G heterogeneous cloud radio access networks, IEEE Network **29**(2), 6–14 (2015)

33.34 Common Public Radio Interface: The CPRI specification V6.1, http://www.cpri.info/downloads/CPRI_v_6_1_2014-07-01.pdf (2014)

33.35 T. Pfeiffer: Next generation mobile fronthaul and midhaul architectures, J. Opt. Commun. Netw. **7**(11), B38–B45 (2015)

33.36 N. Shibata, T. Tashiro, S. Kuwano, N. Yuki, Y. Fukada, J. Terada, A. Otaka: Performance evaluation of mobile front-haul employing ethernet-based TDM-PON with IQ data compression, J. Opt. Commun. Netw. **7**(11), B16–B22 (2015)

33.37 J. Terada, T. Shimada, T. Shimizu, A. Otaka: Optical network technologies for wireless communication network. In: *Proc. Eur. Conf. Opt. Commun.* (2016) pp. 232–234

33.38 A. Maeder, M. Lalam, A. De Domenico, E. Pateromichelakis, D. Wubben, J. Bartelt, R. Fritzsche, P. Rost: Towards a flexible functional split for cloud-RAN networks. In: *Proc. Eur. Conf. Netw. Commun. (EuCNC)* (2014), paper 6882691

33.39 J.M. Meredith: *Study on New Radio Access Technology: Radio Access Architecture and Interfaces*, GPP Technical Report, Reference 38.801, Release 14 (3GPP, Sophia Antipolis 2017)

33.40 X. Liu, F. Effenberger, N. Chand, L. Zhou, H. Lin: Demonstration of bandwidth-efficient mobile fronthaul enabling seamless aggregation of 36 E-UTRA-like wireless signals in a single 1.1-GHz wavelength channel. In: *Proc. Opt. Fiber Commun. Conf.* (2015), paper M2J.2

33.41 X. Liu, H. Zeng, N. Chand, F. Effenberger: Experimental demonstration of high-throughput low-latency mobile fronthaul supporting 48 20-MHz LTE signals with 59-Gb/s CPRI-equivalent rate and 2-μs processing latency. In: *Proc. Eur. Conf. Opt. Commun.* (2015), paper We.4.4.3

33.42 X. Liu, H. Zeng, N. Chand, F. Effenberger: CPRI-compatible efficient mobile fronthaul transmission via equalized TDMA achieving 256 Gb/s CPRI-equivalent data rate in a single 10-GHz-bandwidth IM-DD channel. In: *Proc. Opt. Fiber Commun. Conf.* (2016), paper W1H.3

33.43 C. Kottke, K. Habel, M.H. Eiselt, H. Griesser, J.P. Elbers: Coherent subcarrier-WDM-PON system with SSB modulation and wavelength reuse. In: *Proc. Opt. Fiber Commun. Conf.* (2013), paper OM2A.3

33.44 J. von Hoyningen-Huene, H. Griesser, M.H. Eiselt, W. Rosenkranz: Experimental demonstration of OFDMA-PON uplink-transmission with four individual ONUs. In: *Proc. Opt. Fiber Commun. Conf.* (2013), paper OTh3A.2

33.45 S. Amiralizadeh, A.T. Nguyen, C.S. Park, L.A. Rusch: Single-fiber lightwave centralized WDM-OFDMA-PON with colorless optical network units, J. Opt. Commun. Netw. **8**(4), 196–205 (2016)

33.46 M. Xu, J.-H. Yan, J. Zhang, F. Lu, J. Wang, L. Cheng, D. Guidotti, G.-K. Chang: Bidirectional fiber-wireless access technology for 5G mobile spectral ag-

gregation and cell densification, J. Opt. Commun. Netw. **8**(12), B104–B110 (2016)
33.47 A. de la Oliva, J.A. Hernandez, D. Larrabeiti, A. Azcorra: An overview of the CPRI specification and its application to C-RAN-based LTE scenarios, IEEE Commun. Mag. **54**(2), 152–159 (2016)
33.48 J. Zhang, J. Yu, F. Li, N. Chi, Z. Dong, X. Li: 11 × 5 × 9.3 Gb/s WDM-CAP-PON based on optical single-side band multi-level multi-band carrier-less amplitude and phase modulation with direct detection, Opt. Express **21**(16), 18842–18848 (2013)
33.49 M. Xu, J. Zhang, F. Lu, J. Wang, L. Cheng, D. Guidott, G.-K. Chang: Orthogonal multiband CAP modulation based on offset-QAM and advanced filter design in spectral efficient MMW RoF systems, J. Lightwave Technol. **35**(4), 997–1005 (2016)
33.50 X. Liu, P.J. Winzer, S. Chandrasekhar, S. Randel, S. Corteselli: Multiband DFT-spread-OFDM equalizer with overlap-and-add dispersion compensation for low-overhead and low-complexity channel equalization. In: *Proc. Opt. Fiber Commun. Conf.* (2013), paper OW3B.2
33.51 S. Bloom, E. Korevaar, J. Schuster, H. Willebrand: Understanding the performance of free-space optics, J. Opt. Netw. **2**(6), 178–200 (2003)
33.52 J. Zhang, J. Wang, Y. Xu, M. Xu, F. Lu, L. Cheng, J. Yu, G.-K. Chang: Fiber–wireless integrated mobile backhaul network based on a hybrid millimeter-wave and free-space-optics architecture with an adaptive diversity combining technique, Opt. Lett. **41**(9), 1909–1912 (2016)
33.53 F. Lu, L. Cheng, J. Shi, M. Xu, J. Wang, S. Shen, G.-K. Chang: Efficient mobile fronthaul incorporating VLC links for coordinated densified cells, IEEE Photonics Technol. Lett. **29**(13), 1059–1062 (2017)
33.54 C.W. Hsu, J.T. Wu, H.Y. Wang, C.W. Chow, C.H. Lee, M.T. Chu, C.H. Yeh: Visible light positioning and lighting based on identity positioning and RF carrier allocation technique using a solar cell receiver, IEEE Photonics J. **8**(4), 1–7 (2016)
33.55 T. Jiang, Y. Wu: An overview: Peak-to-average power ratio reduction techniques for OFDM signals, IEEE Trans. Broadcast. **54**(2), 257–268 (2008)
33.56 N. Shibata, T. Tashiro, S. Kuwano, N. Yuki, Y. Fukada, J. Terada, A. Otaka: Performance evaluation of mobile front-haul employing ethernet-based TDM-PON with IQ data compression, J. Opt. Commun. Netw. **7**(11), B16–B22 (2015)
33.57 M. Xu, F. Lu, J. Wang, L. Cheng, D. Guidotti, G.-K. Chang: Key technologies for next-generation digital RoF mobile fronthaul with statistical data compression and multiband modulation, J. Lightwave Technol. **35**(17), 3671–3679 (2017)
33.58 M. Xu, Z. Jia, J. Wang, L.A. Campos, G.-K. Chang: A novel data-compression technology for digital mobile fronthaul with Lloyd algorithm and differential coding. In: *Proc. Opt. Fiber Commun. Conf.* (2018), paper Tu2K.2
33.59 J. Wang, Z. Jia, L.A. Campos, C. Knittle, G.-K. Chang: Optical coherent transmission of 20 × 192-MHz DOCSIS 3.1 channels with 16384QAM based on delta-sigma digitization. In: *Proc. Opt. Fiber Commun. Conf.* (2017), paper Th1K.1
33.60 K. Miyamoto, S. Kuwano, J. Terada, A. Otaka: Analysis of mobile fronthaul bandwidth and wireless transmission performance in split-PHY processing architecture, Opt. Express **24**(2), 1261–1268 (2016)

Gee-Kung Chang

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, USA
geekung.chang@ece.gatech.edu

Gee-Kung Chang is the Byers Eminent Scholar Chair Professor in Optical Networking and Director of the NSF Center for Fiber-Wireless Integration and Networking at Georgia Institute of Technology. He worked in R&D at Bell Labs, Bellcore, and Telcordia Technologies for 23 years, and has co-authored 55 patents and published 600 journal and conference papers. He is a Fellow of IEEE and OSA.

Mu Xu

Optical Center of Excellence
CableLabs
Louisville, CO, USA
m.xu@cablelabs.com

Mu Xu received his PhD degree in electrical and computer engineering from Georgia Institute Technology at Atlanta, Georgia in 2018. Since graduating, he has been working as a lead architect in CableLabs, Denver, Colorado. His research mainly focuses on system design and signal processing in coherent optical systems and mobile fronthaul networks.

Feng Lu

Wireless and Sensing Products Group
Semtech Corporation
San Jose, USA
flu@semtech.com

Feng Lu received his PhD from Georgia Institute of Technology in Atlanta, Georgia in 2018. He worked at the Fiber-Wireless Integration and Networking Center at Georgia Tech, and has been with Semtech Corporation in California since graduation. He is currently working on 5G and low-power wide-area-network (LP-WAN) transceivers, network architecture, and protocol.

34. Space Optical Links for Communication Networks

Alberto Carrasco-Casado, Ramon Mata-Calvo

Future spacecraft will require a paradigm shift in the way information is transmitted, in light of the continuous increase in the amount of data requiring space links. Current radio-frequency-based communication systems impose a bottleneck on the volume of data that can be transmitted back to Earth due to technological as well as regulatory reasons. Free-space optical communication has finally emerged as a key technology for solving the increasing bandwidth limitations for space communication while reducing the size, weight and power of satellite communication systems, and taking advantage of a license-free spectrum.

This chapter focuses on communication links where one of the terminals is in space, and it is organized as follows: Section 34.1 gives a quick overview of key concepts of FSOC that aid in understanding how this technology is used to design lasercom links. Section 34.2 describes how the laser communication signals are affected as they propagate through the atmosphere. Sections 34.3 and 34.4 explain the two applications with greater potential, each having unique characteristics and advantages. Finally, Sect. 34.5 provides an insight of what optical satellite networks may progress toward in the future.

In recent years, wireless communications have witnessed explosive growth of unprecedented proportions. From cellular networks to satellite links, unguided telecommunications has enabled countless new services, becoming firmly established as a basic part of the current information society. In particular, free-space optical communication (FSOC), despite its only recent emergence, provides a number of advantages that enable the materialization of completely new applications such as quantum communications, as well as the promise to revolutionize traditional applications like satellite communications. FSOC can be applied in a wide variety of scenarios, from cross-links to up- and downlinks between satellites, aircraft, ships, and ground-standing or mobile terminals.

Over the past few years, many missions have demonstrated in orbit the fundamental principles of this technology, proving it ready for operational deployment, and we are now witnessing the emergence of an increasing number of projects oriented towards exploiting space laser communication (lasercom) in scientific and commercial applications. This chapter describes the basic principles and current trends of this new technology.

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_34

34.1 Principles of Free-Space Optical Communication

This section reviews the most fundamental parameters in a space laser communication (lasercom) link. Basically, in any space lasercom system, information will be encoded on a laser beam, collimated and transmitted via a telescope, and, after being propagated through free space, will be collected in another distant telescope and focused on a small spot in the focal plane, where a photodetector will transform the optical signal into an electrical one, which will be decoded to extract the original information. A typical space lasercom link (Fig. 34.1) exhibits three different types of impairments, representative of its free-space nature, i.e., geometrical and pointing losses, degradation of the SNR (signal-to-noise ratio) with the background noise, and losses and perturbation of the received signals due to atmospheric effects. This section reviews the first two types, leaving the atmospheric effects to Sect. 34.2 given their complexity and importance in space lasercom. Other fundamental concepts specific to free-space links such as modulation, coding and sensitivity are introduced in this section, as well as a link-budget calculation as the basic design tool in any lasercom system.

34.1.1 Brief Historical Overview

While optical fiber communications experienced huge growth in the 1970s, the next two decades saw significant slowing of FSOC development due to the effects of the atmosphere on optical signals. In 1992, NASA carried out the first laser transmission to space with the GOPEX project, by emitting 532 nm pulses with power of megawatts from the Earth and detecting them with a camera on board the Galileo probe up to 6 million km [34.1]. In 1994, NICT carried out the first demonstration of a space-to-ground communication downlink using the Japanese geostationary satellite ETS-VI [34.2] (engineering test satellite). After the turn of the century, a number of important milestones were marked in rapid succession, the following being especially noteworthy: the first inter-satellite link between the ESA's ARTEMIS geostationary satellite and the French LEO (low-Earth-orbit) satellite SPOT-4 in 2001 [34.3]; the first link between a high-altitude platform (HAP) and ground by DLR in 2005 [34.4]; the first LEO-to-ground downlink by JAXA and NICT using the OICETS satellite in 2006 [34.5]; the first link between a satellite (ARTEMIS) and an airplane in 2006 [34.6]; the first LEO-to-LEO link using Tesat's LCT terminals aboard a US and a German satellite in 2007 [34.7]; the first aircraft-to-ground link by DLR in 2008 [34.8]; the first deep-space-to-ground link using the LADEE probe in orbit around the Moon in 2013 [34.9]; the first high-speed (over Gb/s) GEO-to-LEO link between the European satellites Alphasat and Sentinel 1A in 2014 [34.10]; the first LEO-to-ground lasercom and quantum key distribution (QKD) experiments using a microsatellite (SOTA) by NICT in 2014 [34.11]; the first link between balloons by Google in 2015 [34.12]; the first ground experiments in GEO-equivalent scenarios, achieving 1.72 Tb/s, by DLR in 2016 [34.13]; and the first quantum entanglement experiment from space by China using the Micius LEO satellite in 2017 [34.14].

34.1.2 Key Parameters

Diffraction Limit

The spot size in the receiver's focal plane and the divergence of the transmitted beam are fundamental parameters in the design of a lasercom system. Ideally, both should be as small as possible: the former gives an idea of how well the receiving optical system can focus the received laser signal, and the latter gives an idea of how narrow a laser beam the transmitting system can transmit. Both can be studied using the concept of diffraction limit, which illustrates one of the best benefits of using optical wavelengths. In optics, the Airy disk has traditionally been used for characterizing the spot size in the receiver's focal plane, because it can be easily measured. Derived from the classical description of a plane wavefront illuminating an aperture homogeneously, the Airy disk is defined by the size of the first ring where the light intensity goes to zero. More generally, the maximum spatial resolution of a telescope is given by the diffraction limit, which is determined by the wave nature of light and the finite character of the aperture of an optical system. If the aperture diameter is D and the wavelength is λ, the angular variation of the

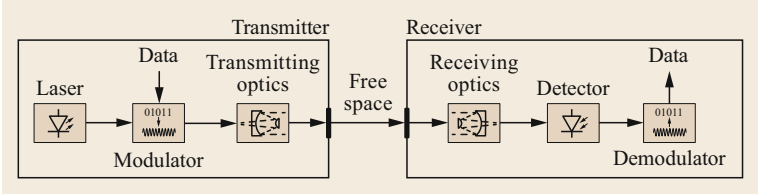


Fig. 34.1 Basic diagram of a generic lasercom system

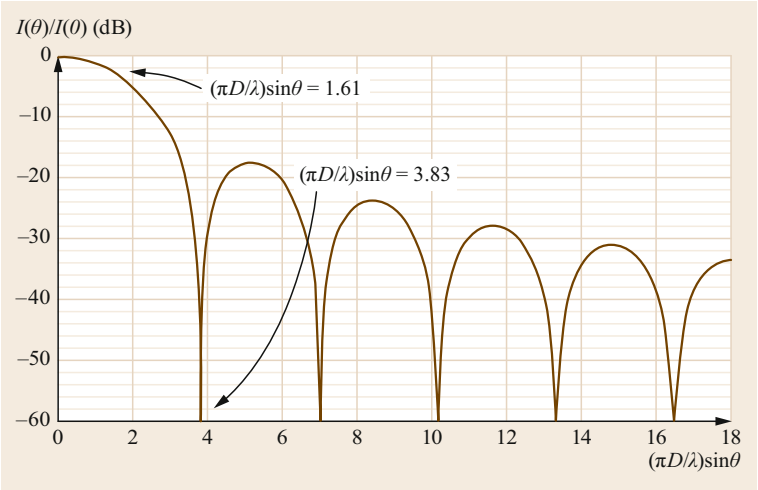


Fig. 34.2 Intensity of radiation as a function of the wavelength λ, the aperture diameter D and the angular width θ

intensity of the radiation $I(\theta)/I(0)$ is given by

$$\frac{I(\theta)}{I(0)} = 2\left(\frac{J_1\left(\frac{\pi D}{\lambda}\sin\theta\right)}{\frac{\pi D}{\lambda}\sin\theta}\right)^2 . \tag{34.1}$$

In (34.1), $J_1(x)$ is the Bessel function of the first kind of x. Its first minimum corresponds to $x = 3.83$, or $x = 1.61$ when $I(\theta)/I(0) = -3\,\mathrm{dB}$ (Fig. 34.2). Using the first-minimum criterion and the approximation $\sin\theta \approx \theta$, the diffraction limit of a telescope can be approximated by (34.2). This limit is sometimes calculated assuming other intensity levels to define the first lobe of the Bessel function, as shown in Fig. 34.2. For example, taking the point where the intensity falls to half, the multiplying factor would be 1.03 in (34.2), instead of 1.22 taking the first minimum. Regardless of the convention, in the graphical representation of (34.1) shown in Fig. 34.2, it can be observed that the width of the main lobe of the Bessel function is proportional to the aperture D and inversely proportional to the wavelength λ. This expression gives an idea of the minimum beam divergence that a perfect—i.e., diffraction-limited—telescope can produce.

$$\theta = 1.22\frac{\lambda}{D} . \tag{34.2}$$

The diffraction limit in (34.2) represents both the radius of the Airy disk in the focal plane of the receiver's telescope and the minimum beam divergence that a given telescope can ideally produce at a certain wavelength, i.e., in a diffraction-limited system. When the equivalent focal length f of the optical system is considered, the diameter d of the point spread function (PSF) can be calculated by (34.3). The PSF is used to quantify the quality of an optical system, and it is defined by the spatial response of the system in the focal plane to a point source in the infinite, which is equivalent to a plane wavefront illuminating the telescope aperture. This expression gives an idea of the minimum spot size that a perfect—i.e., diffraction-limited—telescope can produce in the focal plane

$$d = 2.44\frac{\lambda f}{D} . \tag{34.3}$$

Pointing and Tracking

From the transmitter's point of view, (34.2) illustrates one of the main advantages of FSOC, i.e., the short wavelengths of light can produce very narrow beams, with minimal divergence, where the energy is well confined. In the case of very long distances, as is the case in space links, low divergence becomes a critical factor, allowing a greater density of power per unit of surface area to reach the receiver. This means that much more power can be delivered to the receiver than with radio-frequency (RF), where the wavelength is much longer. Figure 34.3 compares an RF and an optical link from Neptune to Earth transmitting with a 40 cm telescope/antenna at a wavelength of 1 μm (infrared band) in optical and a frequency of 30 GHz (Ka-band) in RF, equivalent to a wavelength of 1 cm. The laser (optical) beam reaching the Earth has a size of around one terrestrial diameter, whereas the RF beam has around times the Earth's diameter. This great directivity demands a high pointing accuracy. After the acquisition, when both terminals establish the line of sight to each other, the procedure to maintain pointing and tracking is several orders of magnitude more complex than with RF. In RF, the pointing accuracy is on the order of milliradians in the Ka-band, whereas a deep-space lasercom link would typically require sub-microradian accuracy.

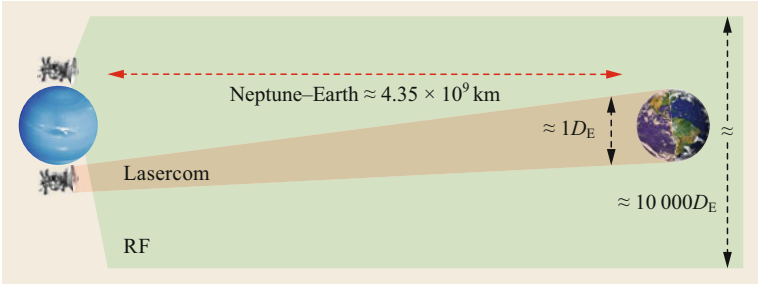


Fig. 34.3 Beam divergence in lasercom and RF from Neptune

To maintain a stable line of sight, it is necessary to use some reference to the other end. This can be achieved by a laser transmitted as a beacon from the ground terminal if the satellite is close to the Earth, or celestial references if it is in deep space. Figure 34.4 shows the main elements in a typical near-to-Earth link. To initiate the acquisition, the beacon is transmitted with a divergence as wide as the uncertainty zone where the satellite is predicted to be according to its orbital elements. Afterwards, the space system searches for the beacon, looking at the predicted direction of the optical ground station (OGS) and transmitting its downlink towards the beacon at a different wavelength or polarization, once it has been found. Lastly, the OGS can transmit a beam much narrower than the beacon by using the downlink reference in a closed loop. Alternatively, scanning algorithms can be implemented, where both terminals angularly scan the counter-partner.

In space lasercom, the so-called point-ahead angle needs to be considered due to the finite speed of light. Since it takes some time for the uplink beam to reach the moving satellite, both downlink and uplink directions are angularly separated by the point-ahead angle, when this is comparable or larger than the beam width. This angle Φ can be calculated with (34.4), where v_t is the tangential velocity of the satellite and c is the speed of light

$$\Phi = \frac{2v_t}{c} . \tag{34.4}$$

For a circular orbit, the worst-case point-ahead angle can be obtained by considering the point where the tangential velocity is the fastest, i.e., the zenith. In this case, the tangential velocity is the same as the orbital velocity v_o, which can be calculated with (34.5), where g is the gravitational constant, M is the mass of the Earth and l is the distance from the center of the Earth to the satellite. As a reference, the point-ahead angle Φ is approximately 51 µrad for 500 km LEO and 18 µrad for GEO

$$v_o = \sqrt{\frac{gM}{l}} . \tag{34.5}$$

Sky Radiance

Atmospheric scattering is an effect of dispersion of light originated by particles suspended in air. Depending on the size of these particles, scattering can be classified into two categories: Rayleigh, when the particles are much smaller than the wavelength, and Mie, when the

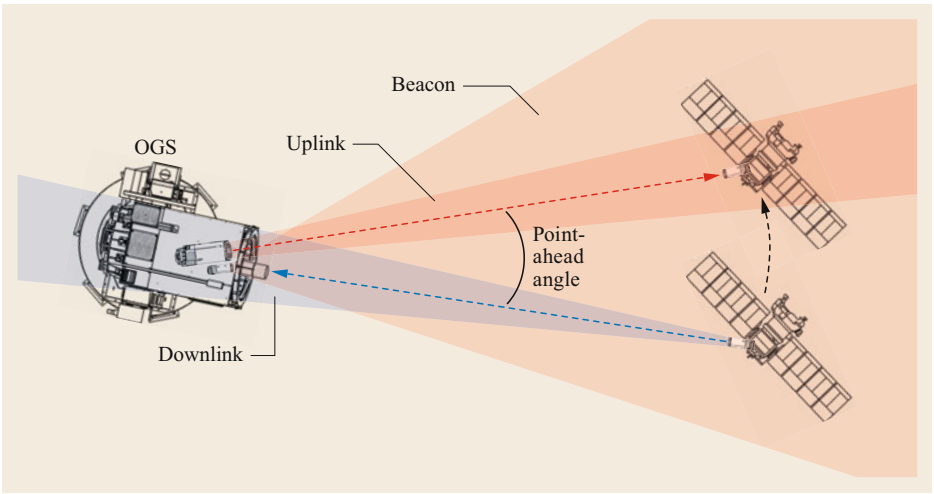


Fig. 34.4 Basic diagram of pointing and tracking

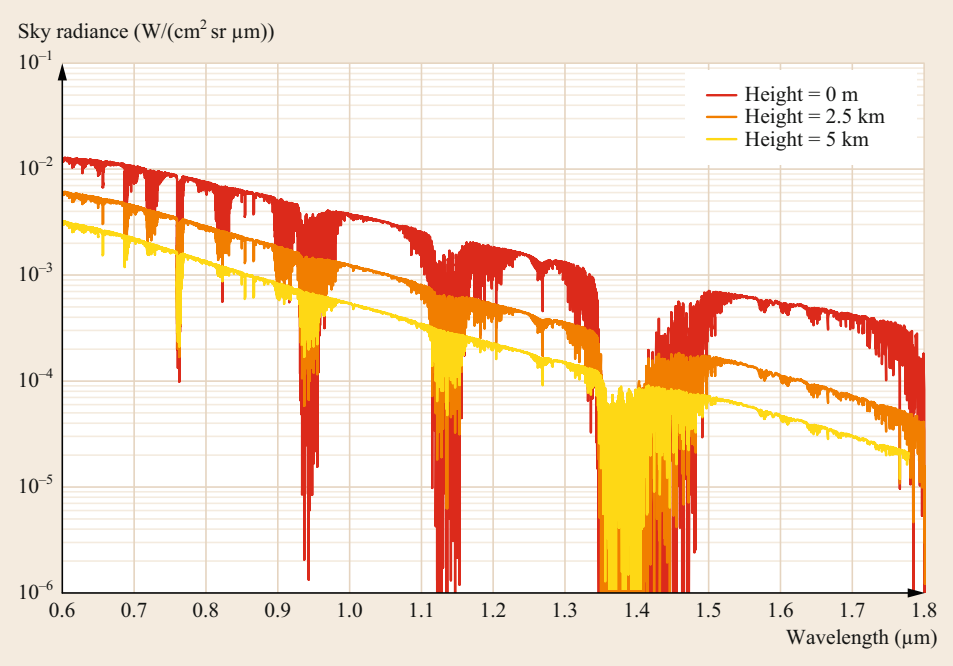


Fig. 34.5 Sky radiance received at three different heights at a zenith angle of 40° with a Sun zenith angle of 60° as a function of the wavelength with good visibility (23 km) and no clouds

particles are comparable to or larger than the wavelength. Although scattering can affect the lasercom link directly as an attenuation of the transmitted signal (see Sect. 34.2.1), a potentially greater impact is due to an indirect effect, i.e., the dispersion of light sources different from the communication laser, such as the Sun or Moon. This diffracted light can enter the field of view of the receiver even if its orientation is angularly far away from these other light sources. Rayleigh scattering is the main contribution when the receiver is oriented with angles far away from the source, and Mie scattering prevails when the receiver points close to the source.

The radiation originated by the scattering of light different from the source is called sky radiance. When it enters the receiver's field of view, it adds to the communication signal as background noise, reducing the dynamic range of the system. This noise source can be modeled as a noise power N_S (expressed in W), which is defined by (34.6), where $L(\lambda, \theta, \varphi)$ represents the sky spectral radiance per area. $L(\lambda, \theta, \varphi)$ depends on the wavelength λ, the receiver's zenith angle θ and the angle φ between the receiver and the Sun (or other noise source), and it is expressed in $\mathrm{W}/(\mathrm{cm}^2\,\mathrm{sr}\,\mu\mathrm{m})$. For a given spectral radiance, the noise power N_S depends on the receiver's aperture area A_R (in cm^2), the field of view Ω_{FOV} (in sr) and the filter's spectral width $\Delta\lambda$ (in μm)

$$N_S = L(\lambda, \theta, \varphi)\, A_R \Omega_{FOV} \Delta\lambda \,. \tag{34.6}$$

Sky radiance varies in a very wide range, depending on many factors such as the presence of aerosols in the atmosphere or the Sun's angular position with respect to the link direction. Figure 34.5 shows an example of its dependence on the wavelength for three different heights of the lasercom terminal location assuming a rural aerosol model with good visibility (23 km) and no clouds. Between 0.8 and 1.5 μm, the most usual wavelengths in free space, there is a difference of more than one order of magnitude, and another order of magnitude between sea level and 5 km at 1.5 μm (half an order for 0.8 μm). Another important factor determining the sky radiance is the angle between the receiver and the Sun. This angle is 10° in Fig. 34.5, although much smaller angles have already been demonstrated (e.g., NASA's LLCD mission went as close as 3° [34.9]). Figure 34.6 shows this dependence, where the sky radiance is represented at 1.5 μm as a function of the zenith angle for several Sun zenith angles, assuming the same conditions of Fig. 34.5 and the middle point height (2.4 km, where the astronomical observatories of Canary Islands, Spain, are located). A difference of more than two orders of magnitude in the sky radiance can be observed between the minimum and maximum angular separation.

Analyzing (34.6), the strategies to reduce the background noise due to the scattering of the Sun's light can be deduced. On one hand, the sky radiance $L(\lambda, \theta, \varphi)$ is usually determined by the mission: the operating wavelength, the satellite's orbit, how close to the Sun the

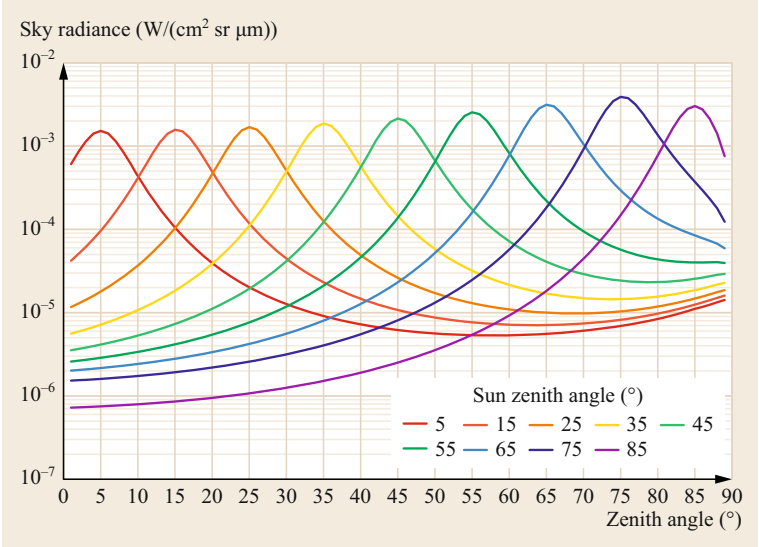


Fig. 34.6 Sky radiance as a function of the zenith angle for several Sun zenith angles at 1.5 μm at a height of 2.4 km for a rural model of aerosols with good visibility (23 km) and no clouds

ground terminal can operate, and the time of the day when the operation is required (this last parameter being almost negligible during the night and a potentially strong source of noise during the day). Reducing the receiver's aperture A_R is not a good strategy for improving the SNR, since it impacts negatively on the signal power. Spectral filtering to reduce $\Delta\lambda$ is an important technique, enabling a decrease in the out-of-band noise, but not in the communication wavelength. Therefore, the field of view Ω_{FOV} would be the parameter to focus on in order to increase the SNR.

Field of View

The field of view (FOV) describes the angular extent that the object plane shows in the image plane of an optical system. FOV depends not only on the characteristics of the optical system, but also on the photodetector that captures the light of that system. Figure 34.7 shows the FOV θ_{FOV} of a generic optical system characterized by its equivalent focal length f and the size d of a photodetector in the image plane.

Figure 34.7 represents the convergence of two collimated beams on the equivalent focal plane of the receiving system, describing the widest angle for a given photodetector size d and a focal length f. The chief ray (going through the center of the optical system) and the marginal ray (going through the edges of the aperture) describe completely the collimated beams going through an optical system. The FOV θ_{FOV} assumes a circular detector can be deduced by (34.7). This equation shows that the FOV is proportional to the detector size and inversely proportional to the focal length, being determined by the chief ray. Therefore, according to (34.6), if the lasercom system has to operate during the day under strong sky radiance, a small detector size and long focal length should be considered (which requires better pointing accuracy) in order to minimize the background noise

$$\theta_{FOV} = 2\arctan\left(\frac{d}{2f}\right). \tag{34.7}$$

Modulation

The simplest way to modulate an optical signal is by turning the transmitter's laser on and off (OOK, on-off keying), as in Fig. 34.8. This is an intensity binary-

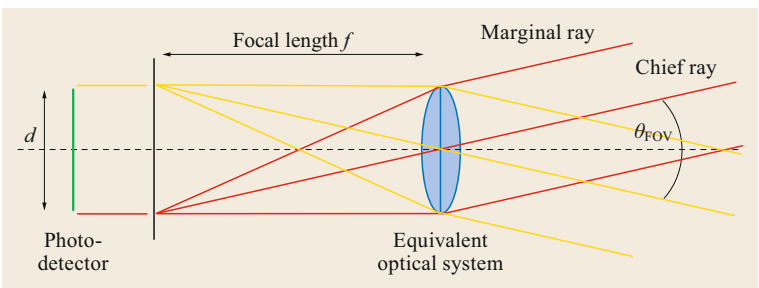


Fig. 34.7 Field of view of a generic optical communication system defined by its focal length f and a photodetector of size d

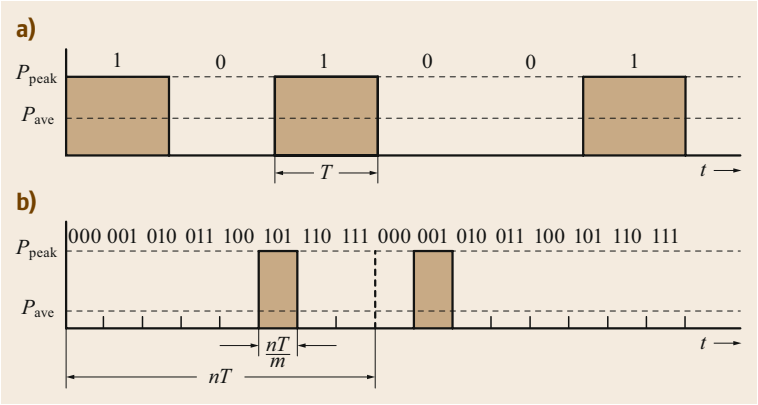


Fig. 34.8a,b Modulation of the sequence 101001 in OOK **(a)** and in 8-PPM **(b)**

level modulation that allows one to use direct detection, which is the most common technique because of its simplicity: IM/DD (intensity modulation/direct detection). These receivers convert the optical signal directly to an electrical current by using detectors following the square law, meaning that the electric output is proportional to the square of the amplitude of the electric field E^2 recovering the original intensity-modulated signal directly. This modulation enables relatively high speed to be achieved while keeping implementation costs low, thus finding good application in scenarios such as LEO-to-ground, especially when small satellites are required. Despite its simplicity, however, this scheme has relatively poor energy and spectral efficiency.

Pulse position modulation (PPM) is a variation of OOK, with lower spectral efficiency but much greater energy efficiency. It finds good application when the spacecraft energy resources are scarce and the losses are high, i.e., low photon flux links, such as deep space. Together with single-photon receivers, PPM is the optimal solution for photon-starved channels with data rates under Gb/s. This modulation allows one to encode more than one bit per pulse by dividing the duration of each sequence of n bits into $m = 2n$ slots, corresponding to m symbols. When each pulse is sent, it is placed in one of the slots, defining the symbol to transmit (Fig. 34.8). In this way, the duty cycle of the laser is reduced, transmitting higher peak power for the same average power (the peak-to-average power is equal to m/n, compared with $1/2$ with OOK) and improving the link SNR at the cost of a $1/m$ times lower spectral efficiency, requiring higher modulation speed to maintain the same binary rate. It is important to note that despite the large bandwidth available in lasercom compared with RF, the spectral efficiency can be a real limitation as well, because it determines the speed of the technology needed to perform the modulation, which can become a difficult requirement in high-PPM orders. PPM can be considered an encoded version of OOK, being based on IM/DD as well, finding its most efficient version when it is applied with photon-counting receivers. However, photon-counting receivers have greater hardware complexity, with the synchronization of the received pulses being the main limitation of high modulation orders in PPM.

Differential phase-shift keying (DPSK) is another scheme that enables the use of IM/DD receivers based on a delay-line interferometer while showing better background rejection than intensity modulations such as OOK or PPM, since the background noise is mostly added to the intensity of the signal, not to the phase. The demodulation is achieved by comparing the phase of two consecutive bits after splitting the incoming PSK signal into two separate paths with a time delay corresponding to one bit between them. Both signals can then be detected by a balanced receiver. DPSK is more robust against the atmospheric effects than intensity modulations, while being much simpler to implement than using a coherent phase receiver, and much more bandwidth-efficient than PPM. It is a good scheme for achieving a high data rate while operating with the atmosphere as the channel.

Coherent demodulation consists in combining the received signal with a local oscillator in the optical domain so that the surface of the photodiode receives a mixture of both signals. For a coherent detector to work properly, it is essential that the local laser matches the received signal in frequency and phase. When that condition is met, this scheme improves background noise rejection because the received signal is amplified after mixing it with the local oscillator, resulting in a higher SNR. Equation (34.8) shows the relation between the SNR of a direct-detection receiver $\mathrm{SNR}_{\mathrm{DD}}$ and a coherent one $\mathrm{SNR}_{\mathrm{CD}}$, both being based on avalanche photodiodes (APD). P_{L} and P_{R} represent the local laser power and the received signal power, respectively, and M, x, R_0, I_{d} and N_{T} refer to the usual APD parameters, i.e., APD multiplication factor,

dependence on the material, responsivity, darkness current, and spectral density of power of the thermal noise, respectively. Equation (34.8) proves that if P_L is large enough, the predominant noise is the shot noise, and SNR_{CD} will always be larger than SNR_{DD}

$$\frac{SNR_{CD}}{SNR_{DD}} = \left(4\frac{P_L}{P_R}\right)\frac{eM^{2+x}R_0P_R + I_d + N_T}{eM^{2+x}R_0P_L + I_d + N_T} . \quad (34.8)$$

Currently, coherent detection based on an analog optical phased-lock loop is applied operationally in intersatellite links. The high sensitivity of this reception technology enables transmission over large distances (LEO to GEO). For satellite-to-ground links, adaptive optics (Sect. 34.4.4) at the receiver is required for either achieving high-heterodyne efficiency when mixing the received signal with the local oscillator, or coupling into a single-mode fiber. The two approaches have similar requirements and depend on the relation between the receiver aperture diameter and the atmospheric coherent length (Sect. 34.2.2), i.e., the Fried parameter [34.15]. Experiments using coherent systems through a turbulent atmosphere have been performed, using both an analog optical phase-locked loop [34.16] and digital signal processing [34.17]. However, for satellite-to-ground communications links up to 10 Gb/s, direct detection is preferred because of the reduced hardware complexity.

Receiver Sensitivity

The receiver sensitivity is defined as the minimum power that a given system needs to reach a given quality measure. It is a fundamental parameter in every lasercom system because it determines all the other design choices. For any signal of a given average power, the photon arrivals in the receiver are not homogeneously distributed in time. Instead, the probability $P(n)$ of detecting n photons during a certain period of time is the average number of photons μ, which follows a Poisson distribution according to (34.9).

$$P(n) = \frac{\mu^n}{n!}e^{-\mu} . \quad (34.9)$$

Considering an ideal detector where there is no noise in the 0s and the quantum efficiency is equal to 1, the decision threshold could be set at 0, because when no photon is transmitted, no photon can be detected. Therefore, the only source of error would be when a 1 is detected as a 0 because of the inhomogeneous Poisson distribution of photon arrivals. If 1s and 0s are equally distributed in the signal, the error probability P_e, or the probability of a 1 being detected as a 0, is $P(n=0) = e^{-\mu}/2$; thus it is possible to derive the average number of photons per 1 bit μ for a given error probability P_e, or bit error rate (BER), which is given by (34.10) and represented in Fig. 34.9.

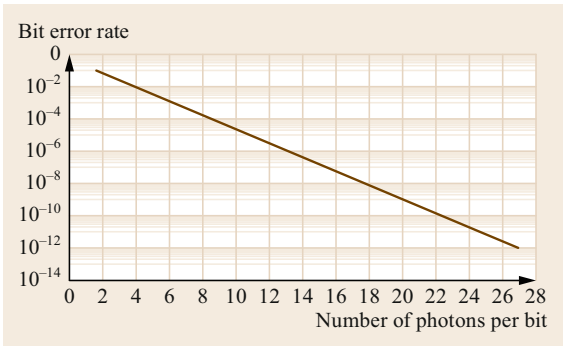


Fig. 34.9 Bit error rate as a function of the number of photons per bit

$$\mu = \ln\left(\frac{1}{2P_e}\right) . \quad (34.10)$$

The average received power P_R can be calculated as the product of the photon arrival rate N and the photon energy $h\nu$, where h is Planck's constant and ν is the frequency of the signal (equal to c/λ), according to (34.11)

$$P_R = Nh\nu = N\left(\frac{hc}{\lambda}\right) . \quad (34.11)$$

Therefore, the received power to achieve a given error probability P_e at a data rate of $1/T_b$, T_b being the bit period, can be calculated as shown in (34.12). Figure 34.10 shows the representation of (34.12), with the relation between the BER and the sensitivity limited by quantum noise for several typical data rates at a 1.55 μm wavelength. In practice, the sensitivity will be higher than this fundamental limit depending on the actual implementation of the receiver, the pre-amplified coherent receivers being the configuration achieving the best sensitivity [34.18] in high-data-rate communications and single-photon-counting PPM with lower data rates limited by the response of single-photon detectors [34.19]

$$P_R = \mu\frac{T_b}{2}\left(\frac{hc}{\lambda}\right) = \ln\left(\frac{1}{2P_e}\right)\frac{T_b}{2}\left(\frac{hc}{\lambda}\right) . \quad (34.12)$$

Coding and Interleaving

The previous section analyzed the quantum-limited performance of a noiseless system. In practice, for non-single-photon PPM, when additive white Gaussian noise (AWGN) is present, the channel capacity C is defined in b/s by the Shannon–Hartley theorem in (34.13), where B is the available bandwidth in Hz and P_R and N_R are signal and noise power in watts, respec-

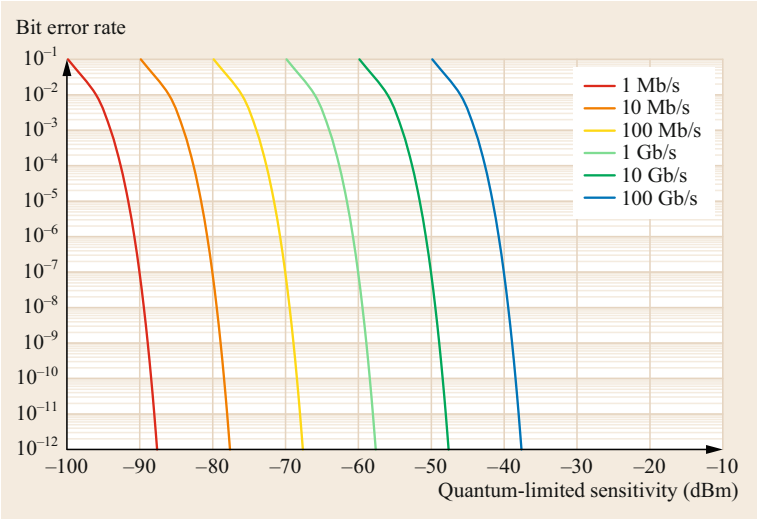


Fig. 34.10 Bit error rate as a function of the quantum-limited sensitivity

tively. It gives another fundamental limit determining the relation between the maximum data rate that can be transmitted with an arbitrarily low BER applying the maximum possible efficiency of error-correcting coding [34.20]

$$C = B\log_2\left(1+\frac{P_\mathrm{R}}{N_\mathrm{R}}\right). \tag{34.13}$$

The link data rate R_b must be lower than the channel capacity C. Since $P_\mathrm{R} = R_\mathrm{b}E_\mathrm{b}$, where E_b is the energy per bit, and $N_\mathrm{R} = N_0 B$, where $N_0/2$ is the noise variance, (34.13) can be expressed as in (34.14), where E_b/N_0 is called power efficiency and the R_b/B is called bandwidth efficiency. Figure 34.11 shows the plot of this equation, where the upper part symbolizes the area where error-free communication is not possible and the area where error-free communication is possible.

$$\frac{E_\mathrm{b}}{N_0} > \frac{2^{\frac{R_\mathrm{b}}{B}}-1}{\frac{R_\mathrm{b}}{B}}. \tag{34.14}$$

The first free-space lasercom links were based more on getting the systems to work than on building efficient systems. How the atmospheric effects affected the communication performance was not well understood, and up to a certain point, it still is not. Therefore, the basic link-budget calculation was assumed, adding a generous link margin to close the link successfully. More recently, the very well-known error-correction coding techniques, used extensively in RF, have been applied to increase lasercom efficiency at a cost of some increase in the electronics complexity, but gaining several decibels in the link margin and moving closer and closer to the theoretical channel capacity limits [34.21].

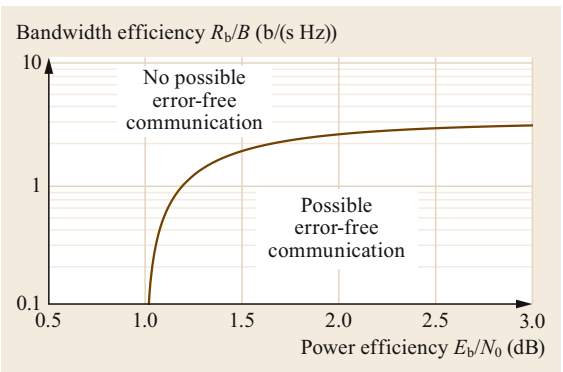


Fig. 34.11 Bandwidth efficiency versus power efficiency

For the successful application of forward error correction (FEC) coding, the atmospheric channel model is a key parameter, especially for understanding how the intensity fluctuates due to the atmospheric turbulence, which is a sizable source of bit errors. There are many different models on intensity fluctuation, although they are gradually being replaced by the gamma–gamma distribution [34.22] because of its applicability to many turbulence regimes, from weak to strong, by multiplying two gamma intensity distributions which consider small- and large-scale cells. When coding techniques are based on realistic modeling of the channel, the SNR gain has a direct impact on the reduction of the power and/or mass/size of the lasercom terminals.

All FEC codes add redundant bits to the output sequence to carry out the error correction in the received signals without retransmission while reducing the effective data rate depending on the expected errors in the channel, thus the amount of redundancy of the code. In the same way as in radio-frequency or optical fiber, the

Table 34.1 Link budget calculation for LEO-to-ground mission

Transmitted power P_T (dBm)	15.40
Transmitting gain G_T (dB)	85.08
Transmitter loss L_T (dB)	1.97
Pointing loss L_P (dB)	5.70
Free-space loss L_S (dB)	259.06
Atmospheric loss L_A (dB)	2.66
Receiving gain G_R (dB)	126.14
Receiver loss L_R (dB)	7.40
Received power P_R (dBm)	−50.18

FEC codes used in free space can be divided into block codes, such as Reed–Solomon (RS) or low-density parity check (LDPC), and convolutional codes, with the turbo codes being the most efficient, performing very close to the Shannon limit at a cost of higher decoding complexity.

Along with FEC codes, interleavers are usually employed in FSOC when the signals go through the atmosphere. Interleavers alter the sequential order of the transmitted bits to reduce the impact of burst errors by taking advantage of the fact that the channel fluctuates several orders of magnitude slower than the bit transmission time. This technique enhances the performance of FEC codes by uncorrelating the signal fading experienced by adjacent bits, distributing the errors in different code blocks in the case of block codes, or separating them in the case of convolutional codes. On the other hand, interleaving introduces latency in the communication and can require large, high-speed memory to store the data during a typical fade, which is an expensive resource in space. As with FEC codes, there are block interleavers and convolutional interleavers, although the convolutional ones provide a reduction of 2× in both latency and memory requirements [34.23].

34.1.3 Link Budget

The link budget is the key method to determine the overall performance of a lasercom system under a set of operating conditions. The basic link budget is given by (34.15) and relates the received power P_R to the transmitted power P_T, the transmission and reception gain G_T and G_R, the losses of the transmitter L_T and the receiver L_R, the atmospheric losses L_A, the pointing losses L_P and the free-space losses L_S

$$P_R = P_T G_T L_T L_P L_S L_A L_R G_R \,. \tag{34.15}$$

The most significant parameters in the link budget can be easily quantified, enabling a quick preliminary analysis of the link to be performed. The transmitted gain G_T and received gain G_R can be calculated with (34.16) and (34.17), where Θ_T is the full transmitting divergence angle in radians, D_R is the telescope aperture diameter, and λ is the wavelength. The pointing loss L_P is defined by (34.18), where Δ_Θ is the pointing accuracy. The free-space loss is given by (34.19), where L is the distance between terminals

$$G_T = \frac{16}{\Theta_T^2} \,, \tag{34.16}$$

$$G_R = \left(\frac{\pi D_R}{\lambda}\right)^2 \,, \tag{34.17}$$

$$L_P = \exp\left[-2\left(\frac{\Delta_\Theta}{\Theta_T}\right)^2\right] \,, \tag{34.18}$$

$$L_S = \left(\frac{\lambda}{4\pi l}\right)^2 \,. \tag{34.19}$$

Table 34.1 and Fig. 34.12 show an example of a basic link-budget calculation for the LEO-to-ground SOTA

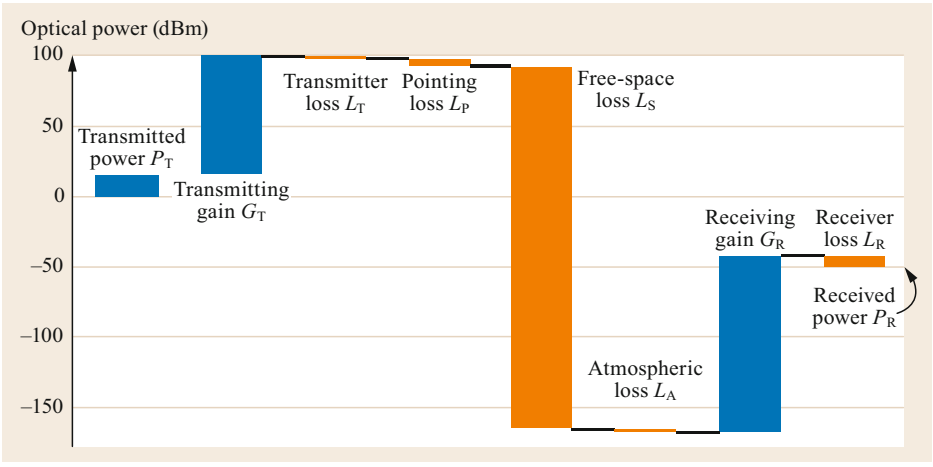


Fig. 34.12 Link budget for the LEO-to-ground NICT's SOTA mission

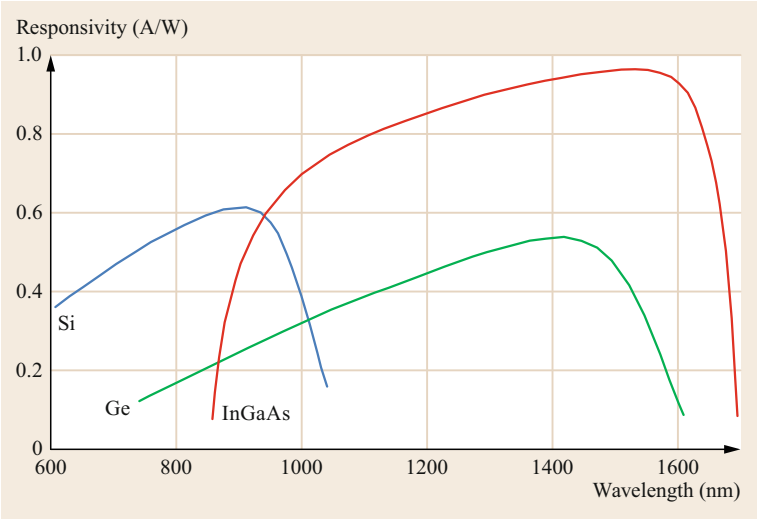


Fig. 34.13 Typical responsivity versus wavelength for silicon, InGaAs and germanium

mission carried out by NICT (Japan) [34.11]. The conditions of this link budget are as follows: the telescope's elevation is 30° for a link distance of 1107 km between the ≈ 600 km SOTA orbit and the NICT's OGS in Koganei (Tokyo, Japan) during the pass on December 9, 2015; the operating wavelength is 1549 nm; the receiver's aperture is 1 m; the optical signal is coupled into multimode optical fiber; and the transmitter, receiver and pointing losses are based on experimental measurements. As a reference, the received power measured in an experiment with the same conditions as in the link-budget calculation was −51.30 dBm. Since there are many factors affecting any link-budget calculation, a common practice is to take a few decibels as the link margin.

Operating Wavelength

Looking at (34.16), (34.17) and (34.19), it is easy to understand how important the wavelength choice is in a lasercom link. According to these equations, the shorter the wavelength, the higher the antenna gain and the lower the free-space losses are. Hence, from the geometrical point of view, shorter wavelengths are preferable (the same consideration is valid for increasing the transmitting and/or receiving aperture in terms of improving the delivery of power due to geometrical factors: in the first case, the beam divergence is reduced, and in the second case, more signal can be gathered). However, the strength of intensity fluctuations due to atmospheric turbulence increases with $\lambda^{-7/6}$, the atmospheric attenuation increases with λ^{-2}, and the scattering attenuation and sky radiance increase with λ^{-4}. Therefore, if the signal must go through the atmosphere, shorter wavelengths provide a larger scintillation (defined in Sect. 34.2) with a stronger impact of the sky radiance.

Currently, there are three important regions where space lasercom operates. The most popular regions are around 1.064 and especially 1.55 μm if the laser beams propagate through the atmosphere because of their better behavior described in the previous paragraph. Furthermore, there is plenty of technology available at 1.55 μm from optical fiber communications, the attenuation is lower and the responsivity of InGaAs photodetectors shows good behavior (Fig. 34.13). From an eye safety point of view, 1.55 μm is preferred, because the eye fluids absorb these wavelengths before being focused on the retina. On the other hand, 1.064 μm has been developed mainly for inter-satellite links, where eye safety is not an issue and there is no atmosphere in the channel. This shorter wavelength takes advantage of lower beam divergence and larger antenna gain, while having available the Nd:YAG technology, especially suitable for coherent communication. When the receiver noise is an issue, e.g., in quantum communications, wavelengths in the band of 800–900 nm may be preferred, taking advantage of the good responsivity of silicon photodetectors (Fig. 34.13), which are less noisy than InGaAs and germanium.

34.2 Characteristics of the Atmospheric Channel

The atmosphere is transparent around 800, 1064 and 1550 nm, which allows for the use of technologies based on silicon, AlGaAs (aluminum gallium arsenide), germanium and InGaAs (indium gallium arsenide). Most commercial off-the-shelf (COTS) components for fiber communications can also be used, especially those at the optical C band, where there are fewer absorption peaks. High-precision lasers at 1064 nm are also available at high power levels, making them suitable for coherent communications.

Apart from absorption or scattering effects, which can be taken into account by increasing the transmitted power by a few decibels (when not transmitting at the absorption line of water or CO_2, for example), atmospheric turbulence is the main challenge when transmitting through the atmosphere. Turbulence causes intensity fluctuations at the receiver which can seriously hamper reliable communication. These fluctuations are due to self-interference processes and pointing jitter, and they can be described statistically by a set of parameters, which are discussed in this section.

34.2.1 Attenuation: Absorption and Scattering

The optical signal in the atmosphere is attenuated mainly by two effects:

- Absorption: Caused by molecular absorption bands, molecular absorption continuums and aerosols.
- Scattering: Caused by molecular effects (Rayleigh scattering) and by larger objects (Mie scattering), e.g., dust particles or fog droplets.

Both absorption and scattering are directly dependent on the particle density and the air density (molecular density). Beer's law in (34.20) gives the power loss L_a through the atmosphere for a given propagation distance z

$$L_a = e^{-\alpha_e z} \quad \text{where} \quad \alpha_e = \alpha_a + \alpha_{sc} \,. \tag{34.20}$$

The extinction coefficient α_e (usually given in km^{-1}) is the sum of an absorption coefficient α_a and a scattering coefficient α_{sc}. For sources with narrow spectral bandwidth such as lasers, the extinction coefficient α_e can be considered constant over the transmitted wavelength portion. The transmission spectrum under clear-sky conditions exhibits windows where atmospheric transmission is conceivable. The absorption coefficient α_a is highly wavelength-dependent: the atmospheric molecules absorb light whose wavelength matches the quantized energy levels of the molecules. The strength of these effects also depends on the location, but it is mainly a function of altitude. There are several databases of scatter and absorption coefficients [34.24, 25]. Generally, volcanic activity is taken into account, since the concentration of aerosols in the atmosphere is highly dependent on it. But they do not include pollution, fog or other local effects, which should be added for each case.

The three common communication wavelengths (810, 1064 and 1550 nm) do not lie in spectrum lines of strong absorption. There is also a decreasing slope as a function of wavelength: this corresponds to the scattering of the air particulates. Moderate volcanic activity has been assumed, which only influences high-altitude curves. The simple approximated model for the clear-sky attenuation in (34.21) can be used for a ground station at 100 m above sea level [34.26]

$$\alpha_{atmos}(\varsigma) = \frac{10}{\left(\frac{\lambda}{1550\,nm}\right)^2 (\varsigma + 1)} \quad [dB] \,. \tag{34.21}$$

In Fig. 34.14, the clear-sky atmospheric attenuation is represented for both wavelengths, 1064 and 1550 nm. At 1064 nm, the atmospheric attenuation can reach 3 dB at low elevation angles, whereas the attenuation at 1550 nm is lower, remaining below 2 dB.

Attenuation caused by rain, snow, fog or clouds, however, is relatively independent of the optical wavelength. Fog and cloud attenuation is strong enough to make many optical links impossible. Near ground, thick fog can reach attenuation values of 300 dB/km ($\alpha_e = 69.8\,km^{-1}$) or more. In the troposphere, cloud attenuation values are on the order of 50 dB/km ($\alpha_e =$

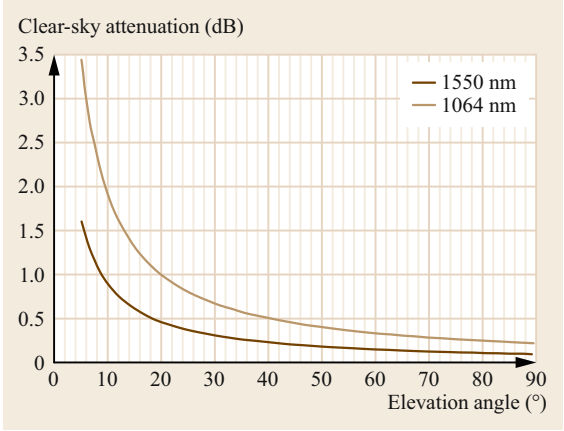


Fig. 34.14 Clear-sky attenuation versus elevation at 1064 and 1550 nm

$11.6\,km^{-1}$). Therefore, in the presence of clouds, optical links are usually not possible. To solve this issue, site diversity is required: a network of ground stations with uncorrelated atmospheric conditions to guarantee that at least one of them is available with a certain probability, increasing the overall network availability [34.27, 28]. The correlation of cloud coverage between two locations usually decreases for distances larger than 80 km. Considering that ground stations in both hemispheres contribute to achieving higher availability with fewer stations, it is important to note that the availability of a single station is not as important as its decorrelation with the other stations in the network. Therefore, stations at locations with high probability of cloud coverage can also contribute positively to network availability.

34.2.2 Atmospheric Turbulence

Atmospheric turbulence modeling is a wide field, which we do not pretend to cover fully in this section. Extensive research on atmospheric turbulence and its effects is available in the literature, covering theory, simulations and measurements [34.29, 30]. The objective of this section is to define the main parameters used to describe the turbulent effects relevant to the communications link.

From the communications perspective, the impact of turbulence can be described with the following parameters:

- Scintillation: normalized variance of the signal fluctuations.
- Intensity correlation time: describes the time characteristics of the signal fading.
- Fading distribution: statistical distribution of the intensity fluctuations.

Turbulence creates phase front distortions on the laser beam propagating through the atmosphere. These distortions lead to destructive and constructive interferences, redistributing the intensity within the beam. The statistical behavior of the atmospheric refractive index is used to model the turbulence and it can be expressed for optical wavelengths in terms of temperature and pressure. Since pressure fluctuations are normally negligible, the index of refraction follows mainly the fluctuations of the air temperature. Therefore, turbulence, i.e., refractive index fluctuations, is created mainly by the mixing of warm and cool air, either by convection or by wind shear. These two phenomena produce eddies (cells) of large scales that break up into smaller eddies, forming an energy transfer cascade. Turbulence cell sizes between a lower limit, called inner scale, and an upper limit, called outer scale, contribute to the optical turbulence. The range between these two sizes is called inertial range. Eddies, below the inner scale, belong to the so-called viscous dissipation range, where the remaining energy in the fluid motion is dissipated as heat. The power spectral density of the refractive index is used to define the contribution of eddies to the atmospheric turbulence. Several models are used to describe the power spectral density considering either one or both inner and outer scales. However, the Kolmogorov model is generally assumed, where the inner scale is set to zero and the outer scale to infinite [34.29]. The main reason is to keep the formulation relatively simple and to avoid estimating inner and outer scales, which usually requires local measurements, not easily practicable. The Kolmogorov model can provide a good estimation; however, it cannot describe all the turbulence effects.

Atmospheric turbulence is a random spatiotemporal field, assumed to be locally homogeneous and defined by variations of the index of refraction around its mean value. This last definition allows one to consider the statistical process stationary in increments, while assuming the mean to change slowly. The structure function is used for the statistical description. This second-order moment is defined as the difference between covariance functions at zero and after a stationary time increment. To relate both spatial and temporal domains, the *frozen turbulence* Taylor hypothesis is assumed, which means that the temporal changes in the atmosphere depend only on the wind velocity orthogonal to the propagation path. The structure function of the refractive index fluctuations is defined as a power law of the distance between two points. This function is directly proportional to a constant, defined as the structure parameter, which describes turbulence strength. The value of the structure parameter depends mainly on two things. The first is the altitude above the ground, as the atmosphere is denser close to the ground, which means that the structure parameter tends to decrease with altitude. The second factor is the time of the day: during sunset and sunrise, the atmosphere becomes quieter because the temperature gradient between the ground and the air decreases, and during midday on a sunny day, the structure parameter tends to be maximal.

For ground-to-space satellite communications, a profile of the refractive index structure parameter is required to describe the turbulence strength along the transmission path. The turbulence strength is described by the structure parameter. The structure parameter is a multiplicative factor of the structure function, i.e., the covariance function assuming that the turbulence is statistically homogeneous and isotropic. Averaged profiles are usually used to qualitatively describe the

atmospheric impact. One of the most widely used is the Hufnagel–Valley profile in (34.22), which basically depends on the altitude above ground h and the turbulence on ground level, defined by the structure parameter at zero height $A = C_n^2(h=0)$ and the mean cross-wind velocity v

$$C_n^2(h) = A\mathrm{e}^{-\frac{h}{100}} + 2.7\times10^{-16}\mathrm{e}^{-\frac{h}{1500}} + 0.00594\left(\frac{v}{27}\right)^2\left(10^{-5}h\right)^{10}\mathrm{e}^{-\frac{h}{1000}}\,. \tag{34.22}$$

For night time, these parameters are usually set to $A = 1.7\times10^{-14}\,\mathrm{m}^{-2/3}$ and $v = 21\,\mathrm{m/s}$, and for daytime, to $A = 1.7\times10^{-13}\,\mathrm{m}^{-2/3}$ and $v = 30\,\mathrm{m/s}$. More complex profiles with more input parameters may properly fit the turbulent path accurately, but as the number of variables increases, the need for an accurate estimation of such parameters may limit their practicability. Using the Hufnagel–Valley profile and the Kolmogorov power spectral density of the refractive index fluctuations, a first-order estimation of the chosen parameters can be performed.

Scintillation

Scintillation is defined as the variance of the signal normalized to its squared mean. This parameter is used to describe the temporal or spatial fluctuations of the received signal. The so-called scintillation index is used to describe the irradiance fluctuations of the optical wave in a single point, i.e., by a point receiver (a receiver can be defined as point receiver when its aperture is smaller than the intensity correlation length, i.e., it will be uniformly illuminated). When the receiver aperture increases and becomes larger than the correlation length of the intensity fluctuations, scintillation decreases due to the so-called aperture averaging. That means the receiver can collect multiple correlation lengths across the transversal plane of the beam, averaging the signal fluctuations.

In satellite communications, the aperture size of the receiver relative to the received intensity correlation length differs between uplink and downlink. For the uplink, the satellite is a point receiver, because the correlation length can be several hundred meters, much larger that the telescope diameter. This happens because the turbulence is closer to the transmitter, and after leaving the atmosphere, only light diffraction occurs. For the downlink, the receiver aperture is usually larger than the intensity correlation length, leading to aperture averaging and a reduction of the scintillation. For a point receiver, scintillation can be estimated with (34.23) and (34.24)

$$\sigma_{I,\mathrm{downlink}}^2 = \exp\left[\frac{0.49\sigma_{\mathrm{Bd}}^2}{\left(1+1.11\sigma_{\mathrm{Bd}}^{12/5}\right)^{7/6}} + \frac{0.51\sigma_{\mathrm{Bd}}^2}{\left(1+0.69\sigma_{\mathrm{Bd}}^{12/5}\right)^{5/6}}\right] - 1\,, \tag{34.23}$$

$$\sigma_{I,\mathrm{uplink}}^2 = \exp\left[\frac{0.49\sigma_{\mathrm{Bu}}^2}{\left(1+0.56\sigma_{\mathrm{Bu}}^{12/5}\right)^{7/6}} + \frac{0.51\sigma_{\mathrm{Bu}}^2}{\left(1+0.69\sigma_{\mathrm{Bd}}^{12/5}\right)^{5/6}}\right] - 1\,, \tag{34.24}$$

where

$$\sigma_{\mathrm{Bd}}^2 = 2.25k^{7/6}\sec(\varsigma)^{11/6}\int_{h_0}^{H} C_n^2(h)(h-h_0)^{5/6}\mathrm{d}h\,, \tag{34.25}$$

$$\sigma_{\mathrm{Bu}}^2 = 2.25k^{7/6}l^{5/6}\int_{h_0}^{H} C_n^2(h) \times\left(1-\frac{h-h_0}{H-h_0}\right)^{5/6}\left(\frac{h-h_0}{H-h_0}\right)^{5/6}\mathrm{d}h\,. \tag{34.26}$$

Both equations are valid under all turbulence regimes and give an estimation of the maximum scintillation that can be expected. The wavenumber is represented by $k = 2\pi/\lambda$, where λ is the wavelength, the link distance is l, and the zenith angle of the link path is ς. The profile is integrated between the OGS height above ground h_0 and the satellite height H.

Scintillation values after aperture averaging can be estimated by (34.27), which however is only valid under weak turbulence conditions

$$\sigma_{I,\mathrm{av}}^2 = 8.70k^{7/6}(H-h_0)^{5/6}\sec(\varsigma)^{11/6} \times\mathrm{Re}\left\{\int_{h_0}^{H} C_n^2(h)\left[\left(\frac{kD^2}{16l}+\mathrm{j}\frac{h-h_0}{H-h_0}\right)^{5/6} - \left(\frac{kD^2}{16l}\right)^{5/6}\right]\mathrm{d}h\right\}. \tag{34.27}$$

As a rule of thumb, when using the Hufnagel–Valley model for the vertical turbulence profile, the weak turbulence approximation holds its validity until 30° for low-strength turbulence. If the turbulence is stronger, the threshold can rise to 60°.

Measured power fluctuations at the satellite are the combination of scintillation and beam wander or pointing jitter. Assuming that the ground station is open-loop pointing towards the satellite, the root mean square (RMS) atmospheric-induced beam wander can be calculated as in (34.28)

$$\sigma_{\mathrm{BW}} = 0.73 \left(\frac{\lambda}{2W_0}\right)\left(\frac{2W_0}{r_0}\right)^{5/6} . \tag{34.28}$$

When the point-ahead angle toward the satellite is small enough, within the coherence cone of the atmosphere, the angle of arrival fluctuations measured in the downlink can be used to pre-correct the beam wander. This point is discussed in a subsequent section.

In Fig. 34.15, the RMS beam wander is plotted versus the transmitted beam diameter for several elevation angles. The RMS beam wander decreases when increasing the transmitted beam diameter, but its impact is larger, as is shown in Fig. 34.17.

Correlation Time and Pointing Jitter

The atmospheric wind affects the temporal bandwidth of the phase perturbations, and it can be modeled by a Gaussian (also called Bufton) model, which is given by (34.29)

$$V(h) = \omega_{\mathrm{s}}[l - l(h)] + 20 + 30\exp\left[-\left(\frac{h-9400}{4800}\right)^2\right] . \tag{34.29}$$

The wind profile was modified, adding the first term, which takes into account the slew rate ω_{s} of the satellite with respect to the ground station; i.e., the angular velocity of the satellite seen by the ground station. This is an important factor especially for LEO satellite links, because of the satellite motion, which virtually contributes as cross wind.

The distance between the ground station and any point in the line of sight to the satellite $l(h)$ can be calculated taking the Earth's curvature R_{e} and the link elevation angle into account as in (34.30), where the link distance can be calculated as $l = l(h = H_{\mathrm{satellite}})$

$$l(h) = -(R_{\mathrm{e}} + H)\cos\varsigma\left[(R_{\mathrm{e}} + H)^2(\cos\varsigma)^2 + h^2 - H^2 + 2(h-H)R_{\mathrm{e}}\right]^{1/2} . \tag{34.30}$$

The correlation time in the signal fluctuations is defined by the Greenwood frequency, i.e., the characteristic frequency of the atmosphere, which can be estimated by (34.31) [34.31]

$$f_{\mathrm{G}} = 2.31\lambda^{-6/5}\sin^{-3/5}(\varsigma)\left[\int_{h_0}^{H} C_n^2(h)V^{\frac{5}{3}}(h)\mathrm{d}h\right]^{3/5} . \tag{34.31}$$

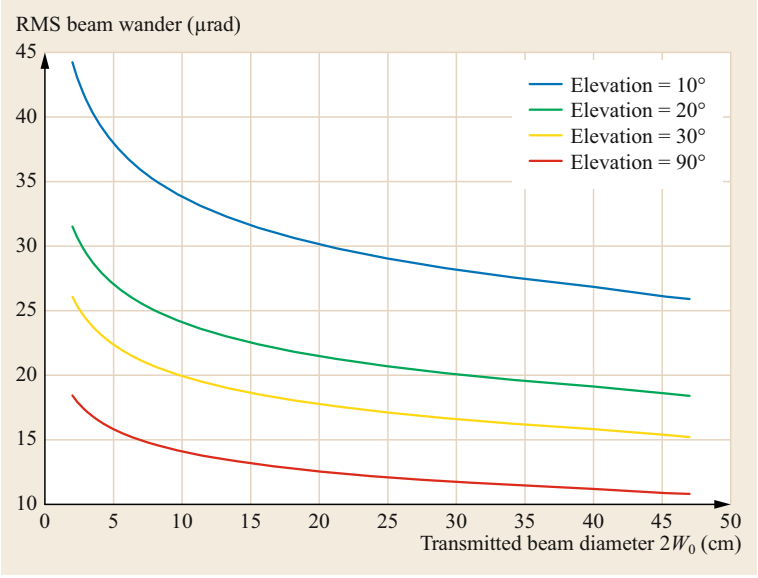


Fig. 34.15 RMS beam wander versus transmitted beam diameter for several elevation angles

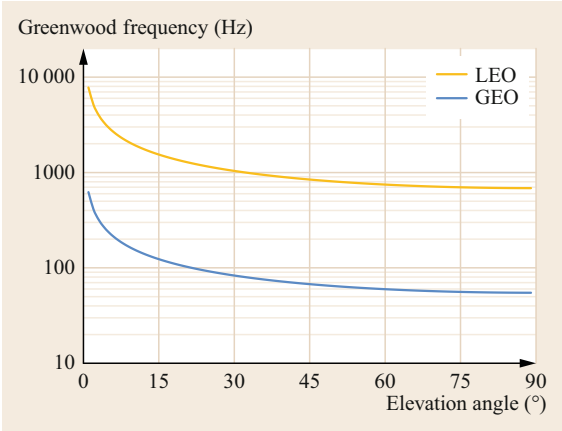


Fig. 34.16 Greenwood frequency f_G versus elevation angle

The correlation time can then be calculated by its inverse $\tau_c = 1/f_G$, and it describes the time behavior of the intensity fluctuations. As scintillation is a self-interference process induced by the phase distortions produced by the atmospheric turbulence, τ_c is introduced to also describe the temporal behavior of the phase distortions. This parameter is especially interesting when designing FEC coding. In Fig. 34.16, the Greenwood frequency is shown versus the link elevation angle. In the calculation, the virtual wind due to the satellite movement has been taken into account by setting the corresponding slew rate ω_s, as defined in (34.29).

In free-space optical communication, the tracking of the optical signal is fundamental to guarantee a stable link. The tilt component of the phase, i.e., the angle of arrival, has another time characteristic as the one considering the whole phase distortions and it can be calculated with (34.32) [34.32]

$$f_{TG} = 0.331 \frac{1}{D^{1/6}\lambda\sqrt{\cos\varsigma}} \left[\int_{h_0}^{H} C_n^2(h)V^2(h)\mathrm{d}h\right]^{1/2} . \tag{34.32}$$

This parameter is particularly interesting for designing the tracking system. Such systems work in closed loop, and the residual error due to the limited closed-loop bandwidth will produce a jitter. The variance of this angular jitter can be calculated by (34.33), where the transfer function $H(f) = 1/(1 + \mathrm{j}f/f_{3\,\mathrm{dB}})$ of the closed loop is assumed

$$\sigma_{TG}^2 = \left(\frac{f_{TG}}{f_{3\,\mathrm{dB}}}\right)^2 \left(\frac{\lambda}{D}\right)^2 . \tag{34.33}$$

Equation (34.33) refers only to one-axis jitter, which is assumed to be normally distributed. Assuming that the two axes are independent, the statistical distribution of the pointing error should be Rayleigh-distributed, with its variance given by (34.34)

$$\sigma_{\mathrm{jitter}}^2 = \left(\frac{4-\pi}{2}\right)\sigma_{TG}^2 . \tag{34.34}$$

For GEO satellites, the uplink pointing uses the information of the downlink angle-of-arrival fluctuations to point towards the satellite and pre-correct the pointing fluctuations of the reciprocal path. Therefore, assuming that transmitter and received beams travel the same path and are driven by the same tracking pointing mirror, the jitter in the tracking system is translated to the pointing system. For LEO satellites, this approach is not used, because the atmospheric path of the uplink and downlink are completely uncorrelated due to the point-ahead angle required to hit the satellite.

The residual pointing jitter impact in the communication can be estimated by normalizing the $\sigma_{\mathrm{pointing}}^2$ by the beam divergence $\vartheta_{\mathrm{beam}}$, here defined as the radius of the transversal Gaussian wave, when the intensity decays to the $1/e^2$. The impact factor $\beta_{\mathrm{pointing}}$ is then defined by (34.35)

$$\beta_{\mathrm{pointing}} = \frac{\sqrt{\sigma_{\mathrm{pointing}}^2}}{\vartheta_{\mathrm{beam}}} . \tag{34.35}$$

Pointing errors can usually be assumed to be negligible when $\beta_{\mathrm{pointing}} < 0.2$. Here, $\beta_{\mathrm{pointing}}$ represents how large the pointing error is with respect to the beam size. When $\beta_{\mathrm{pointing}} = 1$, it means that the shift due to the pointing error is equal to the beam size. However, one may perform an optimization of the beam wander impact (signal fluctuations) with respect to the mean received power. Setting a beam divergence ten times larger than the beam wander implies a huge penalty in terms of received power. Therefore, by designing appropriate coding to protect data transmission, it is possible to substantially reduce the beam divergence. In Fig. 34.17, the $\beta_{\mathrm{pointing}}$ is plotted versus the transmitted beam diameter, which is directly related to the beam divergence as $\vartheta_{\mathrm{beam}} = \lambda/\pi W_0$. For larger beam diameters (narrower beam divergence), the impact on the pointing is larger, although the RMS beam wander is smaller, as already shown in Fig. 34.15. In this calculation, $\lambda = 1550\,\mathrm{nm}$, and the Hufnagel–Valley profile is used for the structure parameter.

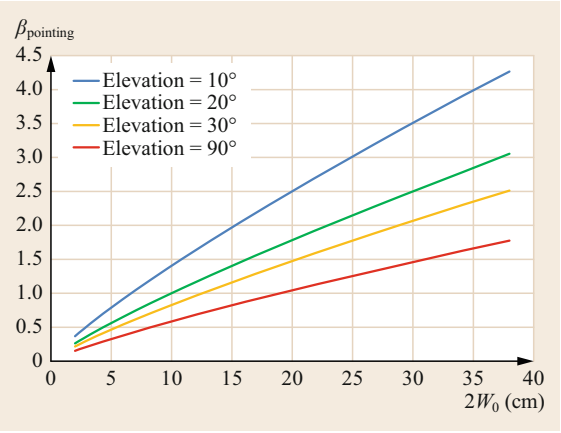


Fig. 34.17 Pointing impact factor $\beta_{pointing}$ versus transmitted beam diameter for several link elevation angles

Fading Distribution

Due to intensity fluctuations, the power might decrease and assume values below the sensitivity threshold of the receiver, leading to a fading or loss of communication. Evaluating the probability of falling under a determined threshold will tell how much the power at the transmitter should be increased in order to ensure the communication availability stays within a desired probability, the so-called fading level.

Fading can be modeled assuming a log-normal probability density function (PDF), especially in weak turbulence conditions. The probability that the received power falls below the threshold P_{min} is defined as in (34.36), where P_{min} is the receiver sensitivity, P_0 is the averaged received power and σ_P^2 is the scintillation of the received signal

$$p_{thr}\left(P_{Rx} < P_{min}\right) = \frac{1}{2}\left(1 + \mathrm{erf}\left\{\frac{\ln\left[\frac{P_{min}}{P_0}\left(\sigma_P^2+1\right)^{1/2}\right]}{2\ln\left(\sigma_P^2+1\right)^{1/2}}\right\}\right). \quad (34.36)$$

The fading level is considered in the link-budget calculation as an extra loss α_{scint}, defined by (34.37)

$$\alpha_{scint} = 10\log_{10}\left(\frac{P_{min}}{P_0}\right). \quad (34.37)$$

Equation (34.38) combines both (34.36) and (34.37)

$$\alpha_{scint} = 4.343\left\{\mathrm{erf}^{-1}\left(2p_{thr}-1\right)\left[2\ln\left(\sigma_P^2+1\right)\right]^{1/2} - \frac{1}{2}\ln\left(\sigma_P^2+1\right)\right\}. \quad (34.38)$$

Under strong turbulence, the intensity distribution does not follow a log-normal distribution. Gamma–gamma and Weibull [34.33] distribution can be used instead, for all turbulence conditions. However, a log-normal assumption can provide a conservative assumption in most cases [34.29].

The log-normal distribution can be modified to include the effects of the pointing jitter as proposed in [34.34]. In this case, it is assumed that the pointing jitter and the scintillation are two independent random processes, and therefore PDFs can be multiplied. The log-normal distribution is used for the scintillation, and the beta-distribution is used for the pointing jitter. The pointing jitter is assumed to be Rayleigh-distributed (normal in each single axis), and it is applied to Gaussian beam form to calculate the intensity fluctuations, leading to a beta distribution. In this case, the combination of the two PDFs is written (34.39), where $I/\bar{I}$ is the normalized intensity, σ_I^2 is the scintillation, and $\beta = \theta_{div}^2/(8\ln 2)\sigma_{jitter}^2$. θ_{div}^2 and σ_{jitter}^2 are respectively the equivalent beam divergence and pointing jitter in the far field [34.35]

$$p(I) = \frac{1}{2\bar{I}}\,\mathrm{erfc}\left[\frac{\ln\left(I/\bar{I}\right)+\sigma_I^2\left(\beta+1/2\right)}{\sqrt{2}\sigma_I}\right] \times \beta\exp\left[\frac{\sigma_I^2}{2}\beta\left(\beta+1\right)\right]\left(\frac{I}{\bar{I}}\right)^{\beta-1}. \quad (34.39)$$

34.3 Low-Earth-Orbit Satellite Communications

Low-Earth orbit (LEO) offers several important advantages to spacecraft including a benign radiation environment, close distance to Earth, low communication latency, and frequent and inexpensive launches. For these reasons, LEO is the most common scenario for small satellites in general [34.36] and the one with the fastest expected growth. The lower cost relative to other orbits makes LEO suitable for missions consisting of single experiments as well as for deploying many satellites in constellations, where they can provide unique coverage of the entire Earth. Launching a satellite to an orbit beyond LEO makes it more difficult to comply with the 25 year post-mission lifetime guideline set by the IADC (Inter-Agency Space Debris Coordination Committee) in order to avoid becoming space debris. For example, CubeSats in orbits above 750 km take cen-

turies to decay, thus requiring some drag strategy. As will be described below, the most important characteristics that define lasercom from LEO are, on one hand, the need to perform accurate pointing and tracking to maintain the link between the fast-moving satellite and the ground station, and on the other hand, the propagation of the laser signals through the turbulent atmosphere before reaching the receiver on the ground. It will be shown that the basic strategies for overcoming the challenges of this type of lasercom link have been already defined, with minor differences, which makes it possible to describe the fundamental principles of this application.

34.3.1 Heritage

Table 34.2 shows an overview of the main characteristics of all in-orbit LEO-to-ground lasercom terminals. The first successful bidirectional LEO–ground lasercom experiment was carried out by JAXA and NICT using LUCE (laser-utilizing communications equipment) aboard OICETS (optical inter-orbit communications engineering test satellite) in 2006 [34.37], which was a 570 kg satellite inserted into a 610 km LEO orbit. LUCE was a 100 kg lasercom terminal based on a two-axis gimballed 26 cm Cassegrain telescope transmitting a 100 mW 847 nm laser at 50 Mb/s, and an accurate fine-pointing system to control a beam with a footprint as small as 5 m reaching the ground station, where a 20 cm telescope was used to gather the received signal, coupling it into an APD.

In 2010, the United States Department of Defense launched the NFIRE (near-field infrared experiment) LEO satellite with Tesat's LCT (laser communications terminal) on board. Although the goal of this terminal was to carry out inter-satellite links, it was used for LEO-to-ground links as a demonstration. The LCT was made up of a two-axis gimballed mirror assembly before a fixed 125 mm telescope, transmitting a 1 W 1064 nm laser at 5.6 Gb/s using homodyne BPSK (binary phase-shift keying) with no beacon. Because of the use of coherent detection on the ground, this system could only close the communication link by using a small receiver aperture (6 cm) located in astronomical observatories in order to watch a Fried parameter larger than the receiver's aperture [34.38].

China in 2011 launched its first lasercom payload LCE (laser communication equipment) to a 971 km LEO orbit on board the Haiyang-2A (or HY-2A) $\approx$ 1500 kg satellite. The lasercom system was based on a 15 cm gimballed telescope and could transmit its 1 W laser beam with tracking accuracy on the order of 1 µrad achieving a maximum data rate of 504 Mb/s [34.39].

In 2014, JPL installed the OPALS (optical payload for lasercom science) terminal in the International Space Station (ISS), at a 408 km LEO orbit [34.40]. OPALS used a two-axis gimbal to move a 5 cm telescope and transmit a 2.5 W laser at 50 Mb/s. The generous transmitted power enabled relaxation of the pointing accuracy down to 300 µrad, enough to point the $\approx$ 1 mrad divergence beam to the ground station.

The first lasercom system on board a small satellite was NICT's SOTA (small optical transponder) [34.11] aboard SOCRATES (Fig. 34.18), which was launched in May 2014 into a 628 km LEO orbit. SOTA was a two-axis gimballed terminal with capabilities to perform a variety of lasercom experiments in a less-than-6 kg compact package. The core experiment was the 10 Mb/s links at 1549 nm using coarse and fine pointing to accurately transmit the 35 mW laser through a 5 cm Cassegrain telescope. SOTA had other additional capabilities, i.e., B92-like QKD protocol at 800 nm band to perform the first quantum-limited demonstration from

Table 34.2 Specifications of in-orbit LEO-to-ground lasercom terminals

	LUCE	LCT	LCE	OPALS	SOTA	OSIRISv2	OSIRISv1	MCLCD
Satellite	OICETS (570 kg)	NFIRE (494 kg)	Haiyang-2A (1500 kg)	ISS (420 t)	SOCRATES (48 kg)	BIROS (130 kg)	Flying laptop (120 kg)	Micius (631 kg)
Operator	JAXA, Japan	DoD-MDA, USA	SOA, China	NASA-JPL, USA	NICT, Japan	DLR, Germany	DLR, Germany	CAS, China
Launch date	Aug 23, 2005	April 24, 2007	Aug 16, 2011	April 18, 2014	May 24, 2014	June 22, 2016	July 14, 2017	Aug 15, 2016
LEO altitude	610 km	495 km	971 km	408 km	628 km	$\approx$ 500 km	$\approx$ 600 km	$\approx$ 500 km
Mass	100 kg	35 kg	67.8 kg	< 180 kg	5.9 kg	1.65 kg	1.34 kg	–
Beacon	808 nm CW	No beacon	–	976 nm CW	1 µm CW	1.56 µm modulated	No beacon	532 nm CW
Downlink	847 nm	1064 nm	–	1.55 µm	800, 980, 1549 nm	1.545, 1.55 µm	–	1549.731 µm
Modulation	On-off keying	Homodyne BPSK	–	On-off keying	On-off keying	On-off keying	On-off keying	DPSK
Max. bitrate	50 Mb/s	5.6 Gb/s	504 Mb/s	50 Mb/s	10 Mb/s	1 Gb/s	200 Mb/s	5.12 Gb/s

Fig. 34.18 Flight model of NICT's SOTA aboard SOCRATES

space [34.41], and 10 Mb/s downlinks at 980 nm using a small lens, both based on coarse pointing only. As a collaboration with the Tohoku University, NICT developed a simplified version of SOTA called VSOTA (very small optical transponder) with a weight of less than 0.7 kg based on body pointing only, thus transmitting a 1550 nm laser beam with a wide divergence (1.3 mrad), low power (80 mW) at a low data rate (up to 1 Mb/s) [34.42]. VSOTA aboard Hodoyoshi-2 (RISESat) lost its launch opportunity planned for 2013 and was successfully launched in January 2019. Based on the SOTA and VSOTA heritage, NICT is currently working towards the next generation of miniaturized high-speed lasercom transmitters compatible with CubeSat platforms for LEO-GEO inter-satellite links as well as LEO–ground.

Since 2008, DLR has been developing optical terminals for small satellites (from CubeSats to $\approx$ 100 kg class) in 1550 nm wavelength based on COTS components to provide solutions to small satellites with limited mass and power. In 2018, the first two generations of terminals were launched into space that base the pointing on the satellite attitude control, reducing the terminal mass to the 1 kg class. The same approach is used in a CubeSat terminal in collaboration with Tesat Spacecom with a $10 \times 10 \times 3\,\mathrm{cm}^3$ form factor, 300 g mass and 100 Mb/s data rate with fine pointing. The 5 kg class OSIRISv3 will include coarse pointing to avoid changing the satellite attitude, providing 10 Gb/s [34.43]. The first generation, OSIRISv1, was launched in July 2017 aboard the Flying Laptop satellite of the University of Stuttgart. DLR's OSIRIS aboard BIROS, known as OSIRISv2 [34.44], was launched in June 2016 into a 500 km LEO orbit, including an InGaAs 4-quadrant-tracking sensor to track the 1560 nm modulated beacon and close the loop with satellite attitude control. The terminal is designed for downlinks up to 1 Gb/s using an OOK-modulated 1 W 1545 nm laser through a 1.5 cm lens with a 200 µrad divergence. OSIRISv2 also includes another downlink capability up to 150 Mb/s using a separate 1.5 cm lens with a divergence of 1200 µrad and transmitted power of 150 mW at 1550 nm. Both terminals were under commissioning at the time of writing this chapter.

In August 2016, the Chinese Academy of Sciences launched the Micius LEO satellite to a 500 km orbit, whose primary mission was quantum communication experiments, but a lasercom experiment called MCLCD (Micius coherent laser communication demonstration) was planned as well [34.45]. The MCLCD space terminal shared the main optics with the quantum experiment used to transmit a 2.2 W 40 µrad 1549.731 nm laser beam with DPSK achieving a data rate of 5.12 Gb/s in the 1.2 m Cassegrain telescope on the ground.

34.3.2 Applications

The most important application of lasercom from LEO is direct-to-ground downlinks, since the main point is being able to download to Earth the increasing amount

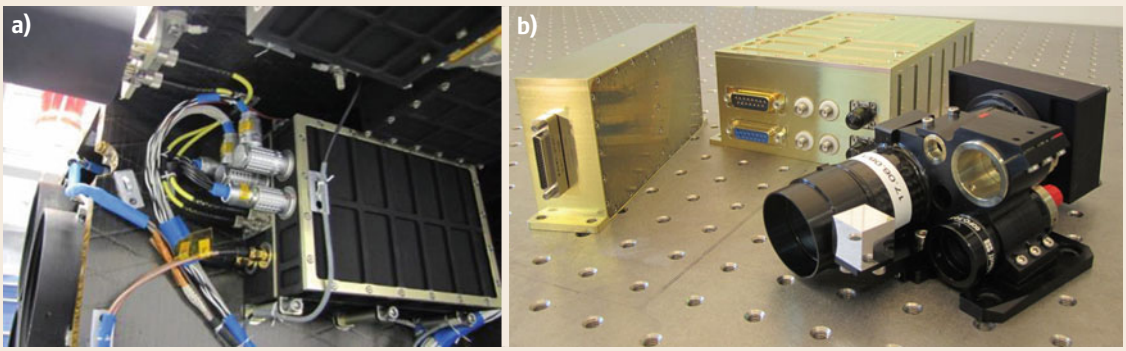


Fig. 34.19a,b OSIRISv1 aboard the Flying Laptop satellite of the University of Stuttgart **(a)**. OSIRISv2 aboard DLR's Biros satellite **(b)**

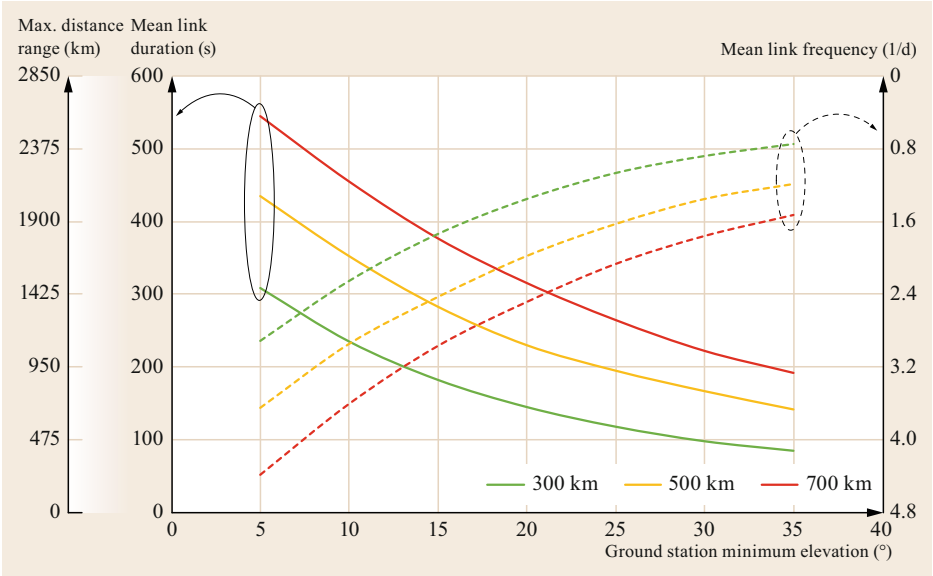


Fig. 34.20 Dependence of the mean link duration and frequency and the maximum distance range with the ground station elevation for three LEO altitudes at NICT's OGS at Koganei (Tokyo)

of data that remote-sensing missions require. The resolution of these sensors is continuously improving, thus demanding more and more bandwidth. Moreover, as launches to LEO become more widely available, spacecraft become miniaturized, as is the case of CubeSats and other small satellites, and with several plans for constellations, the amount of data that will be required to be transferred from LEO to Earth is expected to grow dramatically in the coming years. The already crowded RF spectrum will certainly not be able to support the growing demand for these communications requirements. Therefore, low-complexity and low-SWAP (size, weight and power) high-speed lasercom systems will be a key application to enable the use of such a large number of satellites in LEO.

A typical LEO lasercom link would be as follows: before the scheduled pass, an RF link is used to communicate with the satellite from the ground, updating the orbital data and other relevant pass information, and when the satellite is within line of sight, it makes the tracking sensor face towards the ground station while a powerful beacon is transmitted from the OGS with a divergence wide enough to cover the uncertainty cone of the satellite position. This beacon is used by the satellite as a reference to close the tracking loop with the attitude control of the satellite (body-pointing), with the coarse-pointing system (typically a gimbal) and/or fine-pointing system (in case there is one, usually based on fast-steering mirrors) until the beacon is lost or the communication is degraded.

The LEO scenario generally implies frequent but short passes over a given ground station. Frequency and duration strongly depend on the maximum link range for a given orbit altitude, which is related to the minimum ground station elevation. Figure 34.20 shows the dependence of the average link duration and frequency, and the maximum link distance with the ground station elevation for three typical LEO altitudes. Although these parameters vary with the ground-station latitude, in this case, the NICT's OGS at Koganei (Tokyo) was used, which is in an average latitude (35°41′58″) in the Northern Hemisphere. Taking the 500 km orbit as a typical example of LEO (Fig. 34.20, yellow lines), the maximum contact time would be ≈ 7 min for a 5° minimum elevation, which implies a ≈ 2000 km maximum distance.

Another important limitation to lasercom through the atmosphere is the presence of clouds. Most types of clouds make optical links impossible in practice. Since LEO links are very short, the overall link availability is seriously limited when there are clouds at the time of the experiment. One solution to increase the availability of LEO missions is to relay the data through a GEO satellite. This important application of GEO lasercom can be found in Sect. 34.4. When access to a GEO relay is not available or data must be transferred directly to ground (due to a required short delay or because the onboard technology cannot close a link with GEO), site diversity is usually suggested as the key solution to maximize the probability of achieving cloud-free lines of sight. However, for site diversity to work effectively, two conditions must be met: on one hand, there must be a network of decorrelated ground stations throughout the globe, and on the other hand, they must all use a common set of specifications regarding the wavelength, modulation, codification, etc. In prac-

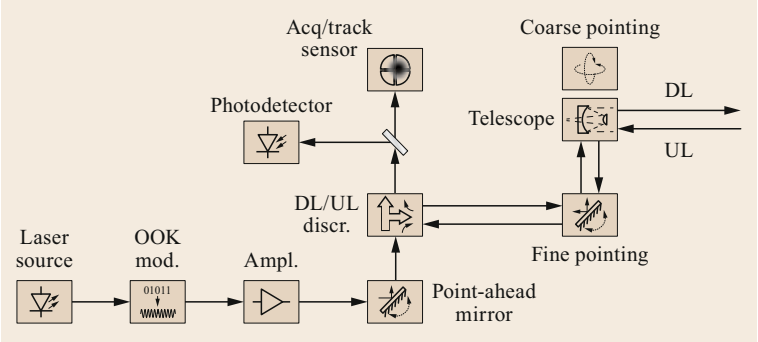


Fig. 34.21 Basic diagram of the optical head assembly in a typical LEO lasercom terminal

tice, currently, neither of these conditions are generally true. Although there is an increasing number of OGSs in many different locations, they are typically built to support specific missions, usually demonstrations, and cannot easily interoperate with each other. The Optical Communications Working Group of the Consultative Committee for Space Data Systems (CCSDS) is currently making efforts to produce standards to enable cross-support between different agencies. A large and growing number of LEO missions could benefit from being able to interoperate with many OGSs spread throughout the world, regardless of the nature of their owners, be they space agencies, universities or private companies. Furthermore, such cross-support would enable many more low-cost LEO missions which could not afford a dedicated OGS.

34.3.3 Space Segment

A typical LEO lasercom terminal basically consists of an optical head assembly and an electronics/processing assembly. As shown in Fig. 34.21, it typically includes a telescope to transmit a collimated beam towards the ground and receive the uplink beam, some device to separate downlink and uplink (typically, based on wavelength, polarization or both), a laser source (with or without amplification), an OOK external modulator (if data rates over Gb/s are required; otherwise, the laser diode can be directly modulated by the input current in a simpler configuration), a tracking detector for the uplink beacon (based on a four-quadrant detector or on a focal-plane array), a fast-steering mirror for the fine-pointing and another for the point-ahead angle if necessary, a fine-pointing assembly (a two-axis gimbal, or alternatively, the satellite attitude control), and optionally a fast photodetector for the communications uplink, if needed. The processing unit typically adds some codification against errors and interleaving against fading, and controls communication, telemetry, command and pointing systems.

Lasercom space terminals usually implement one of the following three kinds of telescopes: In very small terminals, e.g., mounted on CubeSats, based on downlink only, a simple lens (Fig. 34.22a) can produce a collimated beam with a divergence narrow enough for the pointing system, especially when no fine pointing is implemented. For example, a near-diffraction-limited lens with a diameter on the order of 1–2 cm could suffice for a body-pointing-based small satellite. When larger apertures are required, whether it is to produce a very narrow beam supported by a fine-pointing system, or to enable a communications uplink, reflective systems are generally selected. The classic Cassegrain configuration (Fig. 34.22b) is a very common solution for building small and light assemblies with a narrow field of view and good off-axis light rejection. When the receiving aperture must be maximized for the communications uplink, another good solution is an off-axis configuration (Fig. 34.22c), which allows one to make effective use of the whole aperture, avoiding the central obscuration of on-axis reflective telescopes, and provides excellent off-axis light rejection.

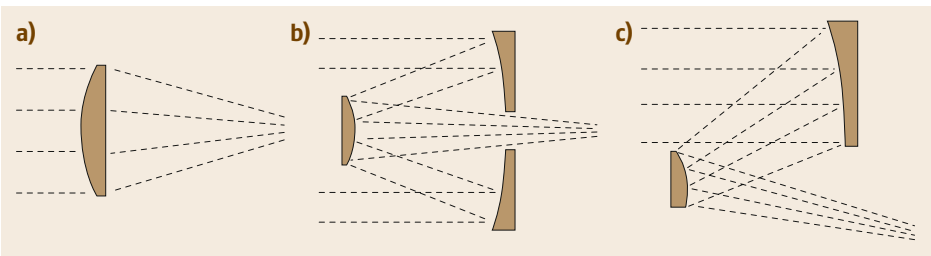


Fig. 34.22a–c Typical configurations for the telescope of lasercom space terminals: **(a)** simple lens, **(b)** classic Cassegrain, **(c)** off-axis Cassegrain

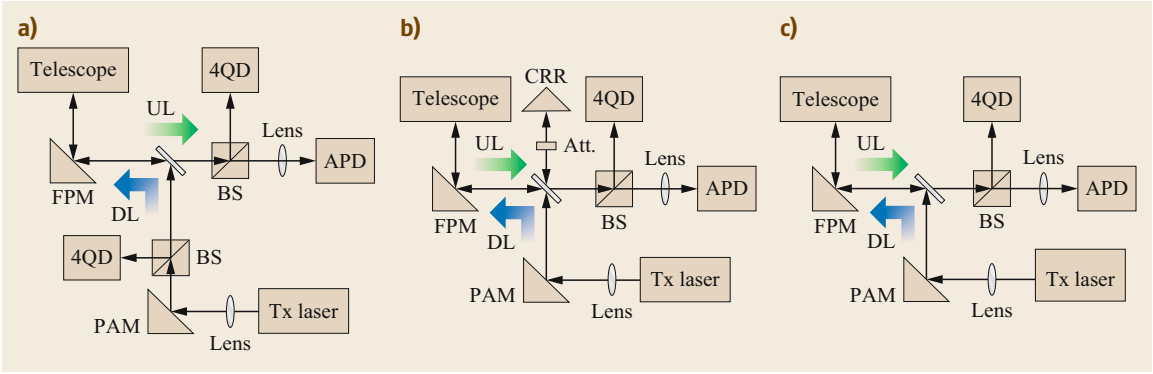


Fig. 34.23a–c Three possible configurations for the point-ahead mechanism in the space terminal

The main challenge for LEO lasercom, with short passes and fast satellite motion, lies in the pointing and tracking system. In terms of beam divergence, (34.2) shows that a relatively small transmitter's aperture can produce a very collimated beam that translates into a small footprint reaching the receiver on the ground. For example, a 5 cm aperture would produce a footprint on the order of 20 m from a distance of 500 km, or on the order of 100 m with an aperture as small as 1 cm. Considering a typical fine-pointing accuracy of 10 μrad and assuming a diffraction-limited transmitting telescope, according to (34.18), the 5 cm aperture would introduce a pointing loss of less than 1 dB, or almost negligible in the case of the 1 cm aperture. With state-of-the-art technology, the pointing accuracy can go down to 1 μrad, enabling a very low pointing loss even with very narrow beam divergence.

A spectral filter of several nanometers before the fine-tracking detector can enable daylight operation of the lasercom terminal, while adding a low-rate modulation to the beacon can improve the background rejection and suppress the downlink scatter.

Acquisition and tracking detectors are typically based on four-quadrant detectors (4QD) or on focal-plane arrays (FPA). The former detectors are based on four active photodiodes used to estimate the spot's center of gravity as the difference in the amplitude of the four elements, and the latter are camera-like sensors based on an array of pixels to directly image the beam. As a rule of thumb, FPAs provide a high spatial resolution, while 4QDs provide a high bandwidth. Additionally, 4QDs allow the use of modulated beacons, and the impact of radiation is lower. Because of their simplicity, 4QDs have been widely used as tracking detectors for many years. However, the latest improvements in FPA technology have been significant, and it may become a good alternative to 4QDs, with the state of the art achieving extremely high sensitivities. Although special care must be taken to protect FPAs against radiation because of their higher vulnerability than 4QDs, their applicability would be more favorable in LEO, where the radiation environment is not as severe as in other scenarios farther from the Earth's magnetic protection.

As explained in Sect. 34.1.2, depending on the beam divergence, it may be necessary to transmit the downlink beam with an offset angle with respect to the detected uplink beam due to the relative motion between the satellite and the Earth, and the finite speed of light. As shown in (34.4), this point-ahead angle is variable depending on the tangential velocity of the satellite as seen from the ground station. For LEO-to-ground links, the maximum velocity, thus the largest point-ahead angle, will happen when the satellite is at the zenith. Taking a worst-case example to illustrate this, a 500 km LEO satellite at the zenith will travel at a velocity of 7.62 km/s, which requires a point-ahead angle of 51 μrad.

Figure 34.23 shows three possible configurations for the point-ahead mechanism. In Fig. 34.23a, the uplink position is detected by a 4QD after separating the uplink and downlink, and the downlink pointing is monitored by another 4QD after the point-ahead mirror (PAM), which makes it possible to compute the required point-ahead angle in real time. In Fig. 34.23b, a small fraction of the transmitted downlink is sent back to the 4QD using a corner retroreflector (CRR) when there is no downlink signal in order to calibrate the necessary point-ahead angle. Real-time point-ahead angle calculation could be implemented if an FPA is used instead of the 4QD, and the power is carefully controlled to produce similar received intensities in the FPA. Figure 34.23c is based on a pre-calibration without monitoring the transmitted downlink, using the beacon from the ground, which despite its simplicity may be a difficult process with fast-moving LEO satel-

lites. When real-time point-ahead angle calculation is not possible, a reference frame transformation is required to obtain the precise direction the point-ahead angle must be applied towards.

In the LEO-to-ground scenario, the preferred wavelengths are around 1550 nm, in the C-band, because of the abundance of devices to generate the laser source, amplify and modulate it. Besides the commercial availability, the reduced atmospheric attenuation, sky background level as well as turbulence effects make this wavelength the most common in this scenario. The wider beam divergence for the same transmitting aperture is not as important as in other scenarios because of the shorter distance. At this wavelength, the basic diagram to generate the downlink signal is shown in the lower-left side of Fig. 34.21. For a high-speed configuration, i.e., multi-Gb/s, the most common scheme would consist of a low-power laser source, followed by an external modulator and an amplifier. The oscillator would be typically based on a semiconductor laser diode. This technology has matured in a significant way to produce high power and high efficiency. However, there is a trade-off between transmitted power, beam quality and modulation rate. Only if power and speed requirements are not demanding, a single laser diode can be used, directly modulated and with no amplification. Laser diodes producing several hundreds of mW are difficult to modulate at high speed (e.g., 1 Gb/s or faster), and the output beam quality is insufficient for the requirements of external modulators.

If data rates on the order of 100 Mb/s and transmitted powers on the order of several hundred mW are sufficient, directly modulated laser diodes can be a simple solution. When both high-speed multi-gigabits per second and high power above 1 W are required, the configuration shown in Fig. 34.21 is the most common one. Despite the additional insertion losses, external modulators are packed in compact devices and enable larger extinction ratios and speeds over tens of gigabits per second. Output powers on the order of several W can be achieved when using fiber amplifiers, usually based on erbium-doped fibers at 1550 nm, i.e., EDFA (erbium-doped fiber amplifier). When high output power is required, several EDFA amplification stages can be used, separated by optical isolators. Since fiber amplifiers typically use single-mode fibers, the output beam quality is excellent, based on a single spatial mode, and the alignment is stable and easily controlled with the fiber connector. The wall plug efficiency (ratio of the total optical output power to the input electrical power) on the order of 10% is the main weak point of EDFA technology, being one of the most power-consuming components in the lasercom system. When available power in the satellite is a strong constraint, using a wavelength around 1 μm could be a solution, since the wall-plug efficiency of YDFA (based on ytterbium) is more than twice that of EDFAs.

34.3.4 Ground Segment

The optical ground station has two important roles in a LEO lasercom system: on one hand, it must provide an uplink beacon so that the satellite can track the ground telescope to transmit its downlink signal, and on the other hand, the OGS must be capable of receiving the communication downlink beam from the satellite and demodulate/decode the received signal.

The first role is simpler to implement than the second. Generally, the beam divergence of the beacon laser should be wide enough to cover the uncertainty cone of the satellite position at its maximum distance, determined by the minimum OGS elevation. For example, a typical uncertainty error on the order of 1 km would require a full-angle divergence over 1 mrad for a LEO satellite at a distance of 1000 km. The optics the beacon is transmitted through are usually mounted parallel to the receiver's telescope and aligned with its tracking detector. Because of its wide divergence, the aperture is usually small (on the order of a few centimeters), and the transmitted power is usually rather high (typically between several and tens of W). Because of the high transmitted power, an eye-safe wavelength is preferred, e.g., in the lower part of the 1.5 μm C-band or in the L-band, leaving the upper part of the C-band for communications, where EDFAs are more efficient and produce lower noise figures, which is more important for communications than for beacons.

A common strategy for reducing the received-power scintillation in the satellite is to transmit several parallel beams [34.46]. If they are separated by a distance greater than the atmospheric coherence length, the effect of the atmospheric turbulence is averaged in the combined signal, and the variance in the power fluctuation reaching the satellite is reduced. This technique is more common for the beacon than for communication uplinks, since the beacon is transmitted with a wider beam divergence to cope with the satellite's position uncertainty, and the communication beam is narrower, enabling more effective power coupling into the receiver. In the case of a beacon, the intensity of the transmitted optical power can also be modulated to improve background noise rejection in the satellite's tracking detector. By using a bandpass filter around the carrier frequency (typically around several or tens of kHz), all the low-frequency and direct current (DC) components present in the time-varying background noise can be eliminated.

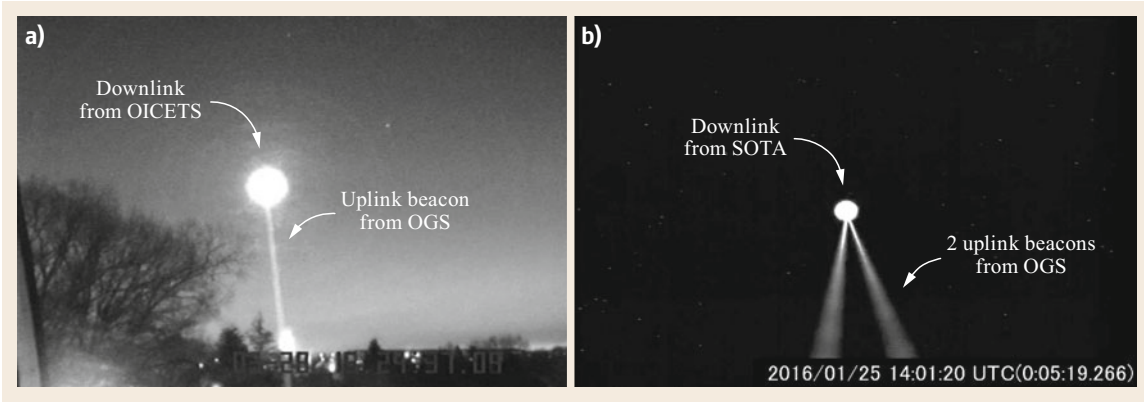


Fig. 34.24a,b Examples of uplink-beacon transmission from two NICT LEO-to-ground missions: 808 nm beacon towards OICETS **(a)** and 1064 nm beacon towards SOTA **(b)**

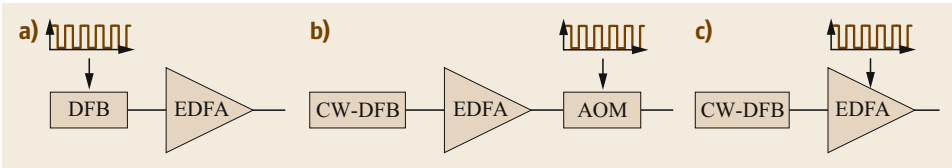


Fig. 34.25a–c Typical modulated beacon laser configurations: **(a)** directly modulated laser source; **(b)** modulated amplified signal; **(c)** modulated pump laser

Figure 34.24 shows images taken with infrared cameras of the uplink-beacon transmission from two NICT's LEO-to-ground missions. In Fig. 34.24a, the beacon towards OICETS can be seen as a straight line in an experiment performed on 28 March 2006, as well as the downlink from the satellite as a bright dot. In Fig. 34.24b, the beacon laser towards SOTA is shown, as well as the downlink signal from the satellite in an experiment performed on 25 January 2016.

Regarding the beacon technological implementation, since the transmitted power is high, a wavelength of 1550 nm is generally preferred because of its eye safeness. At this wavelength, the typical beacon laser configuration would be based on one of the designs shown in Fig. 34.25, depending on how the modulation is introduced in the beacon signal. The design in Fig. 34.25a is the simplest one, based on modulating the seed laser directly by controlling its input current. It is applicable with modulation frequencies of several tens of kHz. For lower frequencies, the excited erbium ions start to lase with no seed signal, causing parasitic oscillation, which may lead to permanent damage. However, there are partial solutions to this even at those lower frequencies, namely, if a dummy continuous wave (CW) signal is introduced at another wavelength (although the total gain would be reduced) or if the modulation depth is lowered to 100% (which makes the demodulation more difficult in the receiver). When a lower modulation frequency is required, the design in Fig. 34.25b is a good solution, based on modulating the amplified signal with an acousto-optic modulator (AOM), which allows a wide range of modulation frequencies, at a cost of a lower output optical power, since the AOM sets a maximum input power on the order of 3 W. The design in Fig. 34.25c consists in modulating the pump laser of the EDFA, and offers high output power and low modulation frequencies, typically below 5 kHz.

The second function of the ground segment in a LEO mission is receiving the communication downlink from the satellite. In LEO-to-ground links, the aperture size of the receiving telescope is an important design parameter. On one hand, the link budget benefits from the greater gain that large apertures provide, but on the other hand, the cost of building monolithic mirrors scales as aperture to the 2.5 power [34.47], which makes them expensive. In addition, large telescopes incur other costs such as operational costs and a dedicated facility. Furthermore, a pointing system that allows a large telescope to track LEO satellites is complex and expensive. For all these reasons, considering the latest improvements in the sensitivity of the receivers, together with the fact that the link budget in LEO-to-ground links is not as demanding as in other scenarios, the tendency has gone towards reducing the aperture size of the ground telescopes. Figure 34.26 shows an example of a 1 m compact OGS built by NICT

Fig. 34.26 Compact and remotely controlled 1 m OGS in Okinawa (Japan) built by NICT

in Okinawa (Japan), which can be controlled remotely from NICT headquarters in Tokyo.

In typical LEO-to-ground lasercom links, where the link budget is not as constrained as in other scenarios, OOK modulation is often used. Even though DPSK offers a 3 dB advantage over OOK, the latter is usually preferred because of the simplicity of the receiver. Whereas a simple photodetector is enough to demodulate the OOK signal, DPSK requires the use of a balanced detector with an interferometer adapted to the specific data rate (as it was explained in Sect. 34.1.2). Furthermore, in LEO-to-ground links, the fast motion of the satellite during the pass produces a Doppler shift on the order of several GHz that can degrade the demodulation process if the path length difference in the interferometer is not precisely controlled by tracking the received wavelength, adding complexity to the system. In addition, the simplicity of OOK receivers allows for the use of a multimode fiber to couple the received signal into, which is relatively insensitive to atmospheric turbulence, as opposed to single-mode fibers due to their small core diameter, i.e., smaller than 10 μm in single-mode fibers and larger than 50 μm in multimode fibers. Moreover, if the area to couple the signal into is larger, a larger PSF can be tolerated, which enables the use of low-cost telescopes with reduced optical quality.

Regardless of the coupling technique the receiver uses, some scintillation effect will usually remain, since this depends mainly on the conditions of the atmospheric turbulence, as explained in Sect. 34.2.2. Together with fading due to pointing errors from the space terminal, the communication link might require the use of FEC coding, depending on the received power and the sensitivity of the receiver. The quantum-limited sensitivity of OOK is 33.9 photons/b [34.48]. However, thermal noise causes worse sensitivity to be seen in practical implementation. As shown in Table 34.3, typical minimum required sensitivities for uncooled APD-based receivers at 1550 nm are on the order of 500 photons/b at 1 Gb/s to achieve a BER better than 10^{-3} (which can be further improved with FEC), or 1000 photons/b to achieve a BER better than 10^{-9} (with no FEC). This kind of receiver enables multimode fiber coupling for the received signal, which is a simple technique, usually only requiring tip/tilt correction with a fast-steering mirror to compensate the beam wander caused by atmospheric turbulence (low Zernike modes) and improve pointing stability (as explained in Sect. 34.2.2).

Table 34.3 Typical sensitivities to achieve a certain hard-decision BER for different receiver schemes at 1 Gb/s and 1550 nm

Receiver scheme	BER (w/o FEC)	Minimum required power
Uncooled APD	10^{-9}	< 1000 photons/b (−38.92 dBm)
Uncooled APD	10^{-3}	< 500 photons/b (−41.93 dBm)
Optically pre-amplified	10^{-9}	< 100 photons/b (−48.9 dBm)

More complex receiver schemes based on optical pre-amplification, i.e., low-noise EDFA before the photo detection, can take the sensitivity from 1000 photons/b in uncooled APDs down to about 100. However, single-mode fiber coupling is required in order to use EDFAs, which makes the receiver more complex, generally requiring adaptive optics to correct the aberrated received waveform. For this reason, this scheme is not usually considered for LEO-to-ground links, where the link budget is not as constrained as in GEO-to-ground

where it finds its natural application, as explained in Sect. 34.4.4. The APD scheme is applicable up to several gigabits per second, but to achieve higher data rates, the optically pre-amplified scheme should be considered, or alternatively and probably more simply, wavelength-division multiplexing (WDM).

34.4 Geostationary Satellite Communications

Satellites at geostationary equatorial orbit (GEO) appear immobile for observers on ground, because the satellite rotates synchronously with the Earth. This property makes GEO satellites especially suitable for communications, streaming or weather monitoring. GEO satellites are interesting because of their large coverage: three satellites can provide worldwide coverage [34.49]. The classical GEO satellite communications market is focused on video broadcasting and Internet access. In recent times, this has been enhanced by backhaul service for edge servers buffering most popular videos, and direct user access to broadband satcom services via VSAT terminals and high-throughput satellites (HTS). HTS can offer communications throughput of more than 100 Gb/s. Current technologies are based on Ka-band and Ku-band technology. Optical communications in GEO have been mainly developed for data relay from LEO, and its major example is the ESA EDRS system. For very-high-throughput satellites (VHTS), a large amount of data needs to be transmitted to the satellite. New communication satellites may require data throughput on the order of Tb/s and more for the feeder uplink; i.e., the link between the ground station and the satellite. Optical feeder links may become the next revolution in space, boosting the available data throughput with potential global coverage using few satellites.

34.4.1 Heritage

ETS-VI: the First GEO-to-Ground Lasercom Demonstration

The Japanese Engineering Test Satellite VI (ETS-VI) (shown in Fig. 34.27a) included the first lasercom terminal in GEO orbit called LCE (laser communication equipment). It was developed by the NICT's Communications Research Laboratory, currently NICT's Space Communications Laboratory and launched on 28 August 1994 by the National Space Development Agency of Japan (NASDA, currently JAXA). The objectives of this mission were to evaluate for the first time the key technologies for satellite optical communications. This downlink signal could transmit an onboard pseudo-random noise sequence to the ground station, relay the uplink signal back to the ground station, or transmit the telemetry data from different lasercom components on LCE at 128 kb/s at an 8× redundancy to achieve the 1.024 Mb/s [34.2, 50]. The LCE lasercom terminal aboard ETS-VI is shown in Fig. 34.27b.

In 1995, the ETS-VI satellite was used to carry out a joint experiment with NASA-JPL, which used two separated OGSs in Table Mountain (Wrightwood) to communicate with the satellite during the project GOLD (ground/orbiter lasercomm demonstration) [34.51].

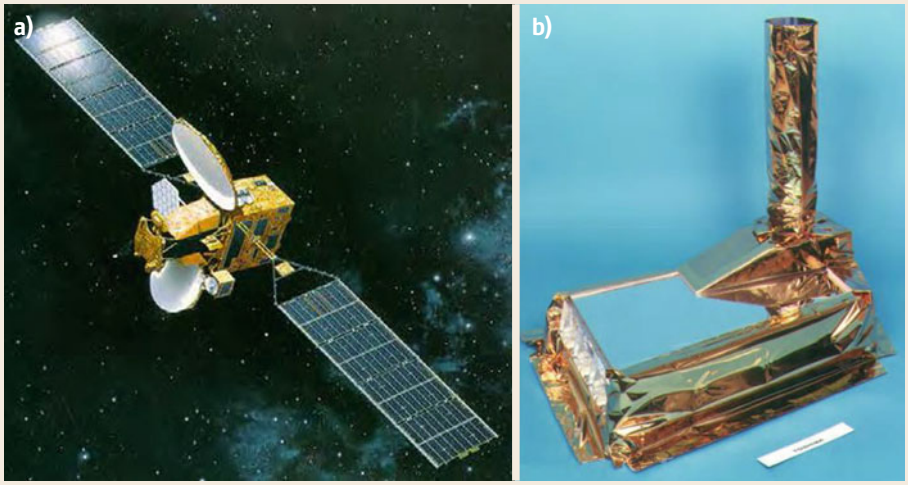


Fig. 34.27a,b ETS-VI satellite **(a)**. LCE lasercom terminal aboard ETS-VI **(b)**

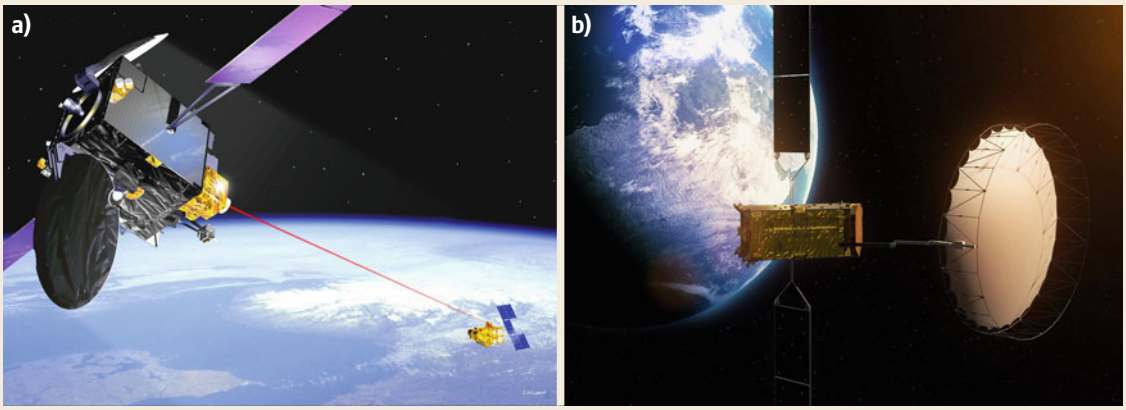


Fig. 34.28a,b Representation of ARTEMIS and SPOT-4 satellites (**a**) and representation of Alphasat (**b**)

After the first demonstration with ETS-VI, two missions with an optical terminal on board were launched to GEO: the Artemis satellite in 1998 by ESA, which will be described in the next section, and the GeoLITE satellite in 2001 by the National Reconnaissance Office (NRO) [34.52], although there is no published information about this mission after its launch, meaning that it probably did not succeed.

Artemis Satellite and the First Inter-Satellite Experiments

The semiconductor laser inter-satellite link experiment (SILEX) was the first civilian optical communications program for space (within the framework of ESA DRTM) [34.3, 53] and the first step towards realizing the European Data Relay System (EDRS). Inter-satellite links were the basis of the SILEX project, with the main objective of relaying video data from a LEO satellite to a ground station, demonstrating the feasibility and performance of optical inter-satellite links in space [34.3]. The experiments involved two satellites which hosted the optical terminals: the ARTEMIS GEO satellite and the SPOT-4 LEO satellite. SPOT-4 was an Earth observation satellite, developed by Matra Marconi Space for CNES. It was successfully launched in 1998, and ARTEMIS, developed by Alenia for ESA, in 2001. The first image transmission was carried out in 2001 between the two optical terminals (Fig. 34.28a), from PASTEL on SPOT-4 to OPALE on Artemis and then to the SPOTIMAGE ground station in Toulouse via a Ka-band feeder link [34.54].

The laser terminals were developed based on OOK modulation and direct detection of laser beams in the 800 nm range (GaAlAs laser diodes and an APD-based receiver). The main parameters of both terminals involved in SILEX experiments are summarized in Table 34.4. The terminals allowed 50 Mb/s data-rate transmission. The terminals on both satellites (OPALE and PASTEL) had a similar structure: a fixed-part electronics and a satellite interface structure with a mobile part. The electronics comprised the onboard processor, the coarse-pointing drive electronics and the communications electronics (interfacing with the signals coming/going from/to the LEO/GEO terminal). The satellite interface structure carried the coarse-pointing mechanism that moved the mobile part, formed by the telescope, the focal plane (with sensitive elements, as sensors and sources) and the required electronics. Because of the relative motion between the satellites, a point-ahead angle assembly with a piezoelectric mirror for the fine tracking was included in the optical system. A high-power laser beacon was used during the acquisition phase for the partner detection on the GEO terminal. The beacon scanned around 750 μrad in the direction of PASTEL. When it was illuminated by the beacon, PASTEL corrected its angle, pointing the communication beam to OPALE, which used the incoming wave to close the loop [34.3, 53–55].

Table 34.4 Main parameters of the optical terminals on ARTEMIS and SPOT-4 satellites

	ARTEMIS	SPOT-4
Antenna diameter (mm)	250	250
Beam diameter Tx ($1/e^2$) (mm)	125	250
Transmit power (mW)	5	40
Transmit data rate (Mb/s)	2	50
Transmit wavelength (nm)	819	847
Transmit modulation scheme	2-PPM	NRZ
Receive data rate (Mb/s)	50	–
Receive wavelength (nm)	847	819
Receive modulation scheme	NRZ	–
Beacon wavelength (nm)	801	–
Optical terminal weight (kg)	160	150

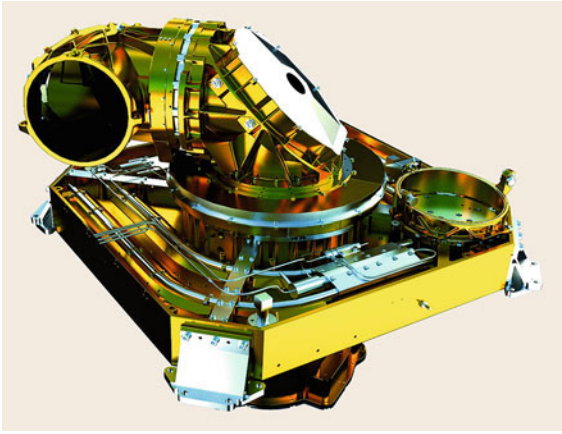


Fig. 34.29 GEO optical communications terminal from Tesat Spacecom

Moreover, since November 2001, bidirectional links have been established between ARTEMIS and the OGS on the Canary Islands. During these experiments, BER and atmospheric parameters were measured [34.56].

Other inter-satellite links were performed between ARTEMIS and the Japanese satellite OICETS. OICETS was a relatively small satellite with a mass of approximately 570 kg. The lasercom terminal on the satellite was designed to have visibility of the geostationary satellite ARTEMIS. The optical antenna had a primary mirror with an effective diameter of 26 cm in a center-feed Cassegrain configuration, and the power consumption of the terminal was around 320 W [34.57, 58].

Inter-satellite laser communications between OICETS and ARTEMIS have been performed several times by JAXA and ESA since 2005, when the first bidirectional inter-satellite link took place [34.58]. Inter-satellite links were also performed through the atmosphere, with the link maintained until the Earth surface blocked the communication. This way, the atmospheric influence on the transmitted beam was measured, and the beam pointing and tracking errors were analyzed as well [34.59].

Alphasat, the Operational Inter-Satellite Links and the News on Relay Systems

In July 2013, the Alphasat satellite was launched and located in the GEO orbit 8° E. This satellite carries several demonstration payloads for satellite communications, among them an optical terminal developed by TESAT Spacecom for LEO-GEO inter-satellite links (Fig. 34.28b). The lasercom terminal (Fig. 34.29) is a made of a central rectangular base structure with a coarse pointer (gimbal) and the optics unit. The telescope aperture is 135 mm [34.60], and the tracking and the communications receiver is based on homodyne reception at 1064 nm. The main parameters of the optical terminal are summarized in Table 34.5.

Table 34.5 Alphasat optical payload main parameters

Dimensions (cm)	$60 \times 60 \times 60$
Weight (kg)	50
Power consumption (W)	160
User data rate (Gb/s)	1.8
Operation range (km)	38 600
Transmit power (W)	5
Wavelength (nm)	1064
Transmit beam diameter (mm)	135
Transmit beam divergence (μrad)	10
Telescope diameter (mm)	135
Nominal receive power density ($\mu W/m^2$)	280 @ BER 10^{-8} (1.8 Gb/s)

The EDRS design originally comprised four satellites: two over Europe (EDRS-A and EDRS-C), one over America (EDRS-B) and one over Asia-Pacific (EDRS-D). In 2018, EDRS-A and EDRS-C were fully deployed, providing operational links since November 2016 for the Sentinel Earth-Observation LEO satellites under the *Space Data Highway* service operated by Airbus [34.61]. The EDRS-B satellite over America most likely will not be developed. Instead, the American LCRD (Laser Communications Relay Demonstration) system will be deployed. At this time, the EDRS-D over the Pacific is planned, and it will provide interoperability with a 1550 nm wavelength, 3.6 Gb/s data transmission, GEO–GEO cross-link communication with the other EDRS satellites and will allow bidirectional links with aircraft. The goal is to create a network that interconnects GEO satellites to LEO and aircraft (or pseudo-satellites) optically and transfers the information from GEO with radio-frequency to the ground. In addition, there are studies to extend the use of data repatriation from the ground [34.62]. This last scenario is especially interesting for isolated areas, where limited or no ground infrastructure is available. An OGS can send the data back to Europe (in this case) over the GEO satellite, avoiding long time delays such as the case of Antarctica [34.63].

The EDRS is being proposed and it is currently being discussed for a possible standardization in the CCSDS Optical Communications Working Group. In this organization, the major international players in space including the space agencies and the industry work mainly in three scenarios: high date rate, low photon flux (deep-space communications) and low complexity (direct LEO downlinks) [34.64].

The Forthcoming Systems

Currently, NASA is developing its own relay system called LCRD (laser communications relay demonstra-

tion), which should start operating in 2019 [34.65, 66]. In Japan, JAXA is developing another relay system called JDRS (Japan data relay system) to start operating in 2019 [34.67, 68], and NICT is developing a GEO feeder link terminal called HICALI (high-speed communication with advanced laser instrument), which aims at demonstrating bidirectional lasercom from GEO up to 10 Gb/s in 2022 [34.69, 70], as well as demonstrating the first LEO-GEO relay using a CubeSat [34.71]. These communications systems are also being standardized in the Optical Working Group of CCSDS under the high-data-rate scenario.

LCRD is the first step towards an American relay system for supporting human exploration and advanced instruments aboard science missions. It is a joint project between NASA's Goddard Space Flight Center, the Jet Propulsion Laboratory, the California Institute of Technology and the Massachusetts Institute of Technology Lincoln Laboratory. The space segment comprises two optical terminals aboard the spacecraft and one high-capacity radio-frequency terminal that will relay data from and to other satellites, spacecraft, airplanes and ground stations. The ground segment includes three OGSs, two optical in Hawaii and California and one in the Ka-band in New Mexico. Two mission operations centers will be connected to the ground stations by terrestrial links. The ground stations will be used to simulate spacecraft users with specific daily data volume requirements. The optical links will provide a bidirectional 1.244 Gb/s data rate. The Ka-band will support one or two users at 32 Mb/s in the forward link and one user at 622 Mb/s or two users at 311 Mb/s in the return link. Each optical terminal consists of a telescope, the electronics for pointing and acquisition, and a modem that supports PPM and DPSK in both link directions. The experiment will also demonstrate specially developed encryption technology for information assurance. Modems supporting both modulation formats were developed and demonstrated previously under the LADEE mission, which in 2014 demonstrated optical PPM data communication from the Moon, achieving a user data rate of up to 622 Mb/s.

In 2019, JAXA plans to launch the first satellite of the Japanese Data Relay System (JDRS). JDRS will include a feeder link in the Ka-band, and in optical it will provide data rates of 2.5 Gb/s for the return link and 60 Mb/s for the forward link including FEC codification, leading to user data rates of 1.8 Gb/s and 50 Mb/s, respectively. The inter-satellite communication system is designed in the optical C-band, 1540 nm for the forward link and 1560 nm for the return link. The optical terminal in GEO will have an antenna diameter of 15 cm, whereas in LEO the aperture will be 10 cm. The acquisition is based on a beaconless sequence to be completed within 60 s. The optical communication system is based on direct detection, with DPSK for the return link and intensity modulation for the forward link. The first JDRS user will be the advanced optical satellite, a Japanese optical observation LEO satellite from JAXA.

Fig. 34.30 Artistic illustration of NICT's HICALI aboard ETS-IX

NICT in Japan is developing a high-throughput GEO satellite called ETS-IX based on hybrid use of radio and optical frequencies (Fig. 34.30). ETS-IX will be launched in 2022 during the second flight of the new Japanese launch vehicle H3. The Ka-band system will include feeder links and user links with 100 Mb/s per user, with flexible allocation of frequencies and steerable beams to handle traffic fluctuations. The lasercom terminal is called HICALI (high-speed communication with advanced laser instrument) and it will demonstrate a bidirectional 10 Gb/s feeder link between GEO and ground. The HICALI terminal will transmit a 2.5 W 1540 nm laser through a 15 cm aperture to be received on the ground by a 40 cm receiver aperture. For the uplink, a beacon system consisting of four apertures of 5 cm will transmit total power below 20 W at 1530 nm, and the feeder uplink will transmit 2.5 W at 1560 nm through an aperture smaller than 40 cm. The HICALI mission will also include feeder link experiments using the NICT OGS network to demonstrate the high availability provided by site diversity, enabling fast handovers between different ground stations depending on the cloudiness over each site.

34.4.2 Applications

For optical communications, satellites at this orbit have been proposed mainly for data relay from LEO. Transmission to GEO satellites extends the ground coverage and decreases the delay between users which are not visible from the same satellite. Moreover, the intermittence of the LEO satellite's visibility is overcome by transmitting to a GEO satellite, reducing the onboard data

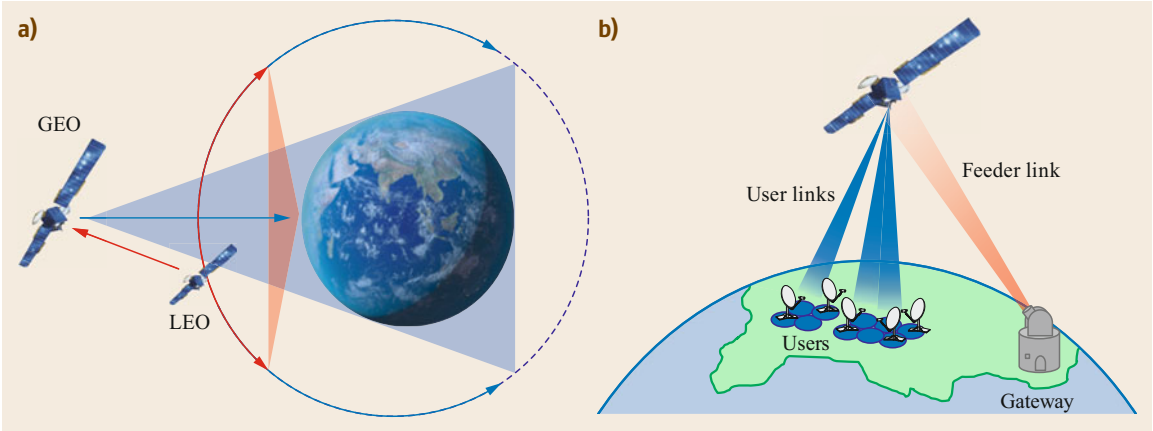


Fig. 34.31a,b Applications for GEO optical satellite communications. The GEO data relay provides larger coverage to LEO for data repatriation **(a)**. Satellite communications for Internet access and streaming based on optical feeder links **(b)**

recorders and increasing the capacity. In Fig. 34.31a, there is a representation of a satellite relay system. The coverage of the lower orbits is larger at the GEO than from a given station on the ground. For this application, the ESA developed the EDRS, which became operational in November 2016, giving service to the European Commission's Earth observation program.

Classical satellite communication is focused mainly on television broadcasting, but in the coming years, the provision of Internet access may gain more weight. The optical fiber infrastructure is costly, especially in less densely populated areas. Several initiatives in Europe target full coverage of Internet access at 50 Mb/s and more, which could not be reached with terrestrial infrastructure alone. For example, 30% of the rural areas in Germany are currently not covered. In this sense, satellite communications can be a good complement to reach the full coverage for broadband Internet access. Another aspect is 5G, the Industry 4.0 and the Internet of Things, which also require global connectivity. In this case, new business models based on cloud services and real-time monitoring of the production and transport of products will require Internet access everywhere and at any time. Geostationary communications time delay of at least $\approx$ 250 ms due to signal propagation may be limiting for applications such as telepresence or augmented reality, where direct human interaction is expected, but for other kinds of applications it may have no impact.

In Fig. 34.31b, a representation of the GEO satellite-based communications is shown. A bidirectional link, called feeder link, between the satellite and a ground station, called gateway, connects the satellite to the network. The satellite gives connectivity to the surface with several spot beams, called user links. The feeder link therefore carries much traffic, with throughputs that currently reach several hundred gigabits per second, but in the future it can potentially reach several Tb/s [34.72]. Current radio-frequency-based technology is reaching the limit on providing the required throughput, due to the limited available bandwidth. Hence, several OGSs operating in the same spectrum feed the satellite to reach the required traffic. The number of these stations increases linearly with the throughput, reaching hundreds when approaching Tb/s. Therefore, in the future, communication satellites will integrate optical communications for the feeder link in order to increase the data throughput. This would have two advantages: first, the feeder link capacity would no longer be limiting, due to the $\approx$ 10 THz of unregulated spectrum around 1064 and 1550 nm, and second, the spectrum currently used for the feeder link could be allocated for the user links, increasing the overall throughput by these two means.

In order to make effective use of the available spectrum in the optical domain, wavelength-division multiplexing (WDM) is required. Optical channels centered at different optical frequencies are multiplexed together, being able to reach an aggregate throughput of multiple Tb/s. In fiber communications, such technology is the state of the art for the optical C- and L-bands. The center frequency of each channel is defined in an International Telecommunication Union (ITU) standard for the dense WDM, with channel bandwidths of 100, 50, 25 and 12.5 GHz [34.73]. The use of such technology in the atmospheric turbulent channel was demonstrated in 2016, transmitting 1.72 Tb/s in 40 channels with 100 GHz bandwidth using OOK modulation [34.74], and in 2017 transmitting 13.16 Tb/s in 51 channels with 50 GHz channels using 16 QAM (quadrature amplitude modulation) and QPSK modulation [34.75].

Current projects in optical communications in space target single-purpose satellites, meaning that only one

application is addressed. In the future, if the demand on connectivity and bandwidth does not drop off, the ground network may be extended to space. In this case, GEO satellites acting as nodes may connect directly to the ground network, as backhaul links. GEO is advantageous due to the relatively small movement of the satellite, compared with LEO satellites, which require handovers quite often. The GEO satellite nodes could then distribute the connectivity to other satellites in GEO, MEO or LEO for relay and telecommunications applications, being then fixed infrastructure in space as an extension of the terrestrial network. By properly designing the ground segment, continuous optical connection to the satellite could be guaranteed, meaning that most of the current bottleneck due to the limited radio-frequency spectrum would be solved. This approach would be compatible with a LEO constellation, for example. Connecting this constellation with the GEO satellites through a relay link would extend the high data rates at all orbits and to any kind of satellite, even small LEO satellites like CubeSats [34.76], which could make use of a LEO constellation to relay the data through the GEO satellite. Traffic demands on small delays could then be routed anyway with a direct link to Earth.

34.4.3 Space Segment

Communications systems on satellites are limited by the platform mass and power budgets, and thus their design is confined to the available onboard resources. This makes the design of very-high-data-rate systems for space very challenging. Currently, there are no satellite buses specifically designed for very-high-throughput communications based on lasercom, especially targeting multiple Tb/s. Therefore, there is still work to be done within the space segment to design appropriate platforms to accommodate such kinds of terminals.

One important point in the design of future systems based on optical feeder links will be to retain the compatibility towards users, which requires including digital satellite television broadcasting standard DVB-S2X for the user links. This standard defines the physical layer, including modulation and coding schemes. The user-link modulation is defined as APSK (amplitude–phase-shift keying) with different modulation orders, each targeting different carrier-to-noise ratio requirements, the lower orders being for poor SNR conditions. For the current system technology in RF, it seems reasonable to use modulation orders up to 16 or 32 APSK, but the standard defines modulation orders up to 256 APSK.

From the communication system point of view, there are several approaches still under study for future-generation systems. The different options can be classified between analog and digital payloads and between transparent and regenerative, and they are summarized in Fig. 34.32. The choice is driven by a trade-off between complexity and efficiency.

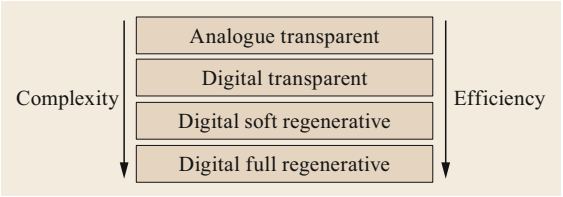


Fig. 34.32 Communications design approaches

Currently, the satellite communication community prefers analog transparent systems, acting as transponders, as are state-of-the-art satellite communication systems based on radio-frequency. This enables the development and upgrading of the communication system without compatibility issues of previous satellites. From the point of view of the optical feeder link, this means that the data need to be analog-modulated on the laser light [34.77], similarly to radio over fiber [34.78], to avoid any onboard data processing. APSK signals can be mapped in amplitude analog modulation by setting the signal to an intermediate frequency, which first would help to convert the signal on the satellite to the Ka-band and enable the multiplexing of more DVB carriers in a DWDM channel. At the satellite terminal, the signal is converted to the electrical domain and frequency up-converted to the Ka-band for the user link. With this approach, no signal processing on the satellite is required, and therefore the ground segment must take care that the signals are correctly shaped for the user link.

Another approach that maintains the transparency of the system is to digitally sample the signal and transmit the samples on the feeder link. This approach is similar to the analog one, but it allows FEC to be added on the bits containing the sampling information for the feeder link in order to mitigate the turbulence effects in the uplink. This means that onboard processing is required for decoding the FEC. In this case, the system is transparent from the point of view of the DVB signal, since the wavefront is digitally sampled. The main drawback of this approach is the band expansion due to the sampling, but in general it is more efficient than the analog approach. Error correction algorithms help achieve better sensitivity. An overview of this approach can be found in [34.79, 80].

Soft-regenerative techniques are halfway between transparent systems and fully regenerative systems. In this case, the system is transparent to the data, but it requires generation of the DVB wavefront for the

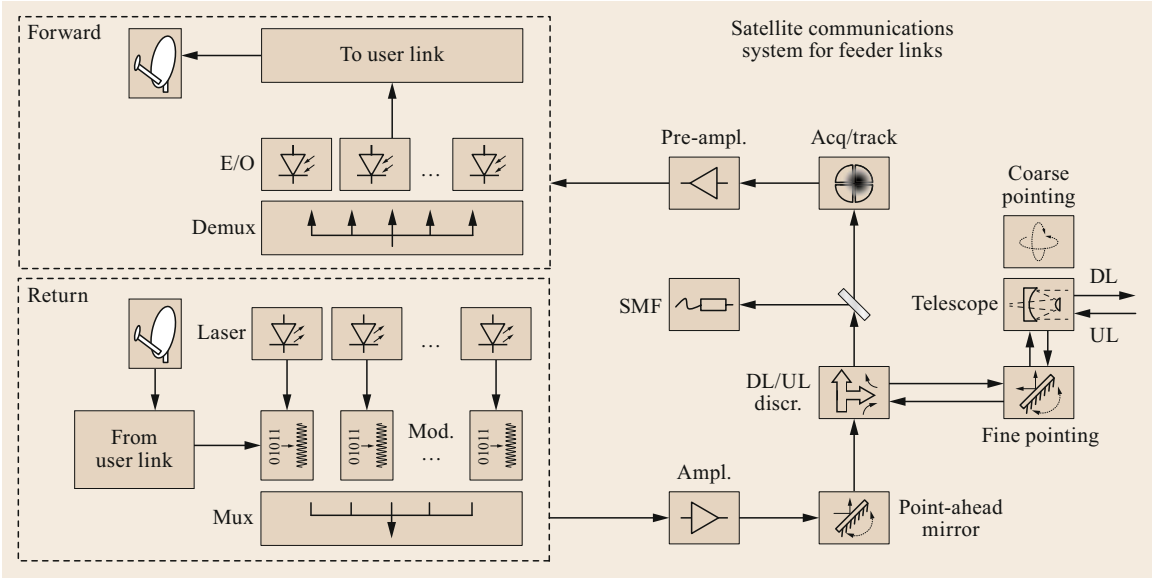


Fig. 34.33 Main components of the space terminal of a very-high-throughput communications satellite based on an optical feeder link

user link. The constellation points of the DVB signal can be mapped to the optical feeder link signal. The band expansion depends on the modulation order of the optical signal, but it is lower than in the digital transparent option. Data is transmitted in baseband, and error-correction algorithms can be used to protect the transmitted information. At the satellite, data should demap the constellation points of the feeder link into the APSK DVB format, and the wavefront must be generated. However, there is no need to access the DVB frame.

Finally, a fully regenerative system would be the most efficient approach, enabling optimization of the physical layer. This option requires full manipulation of the contents of the DVB signal, and it would require the signal to be completely generated aboard the satellite. That means that this is the most costly option in terms of signal processing, and therefore on mass, power consumption and heat dissipation. However, although this option would exploit the full potential of the satellite link, it is not expected for the coming generations of satellite terminals.

Figure 34.33 shows a block diagram with the main components of the space terminal of a very-high-throughput satellite terminal based on an optical feeder link. The telescope with the coarse and fine-pointing units carry out the single-mode fiber coupling. The received signal is pre-amplified, de-multiplexed, converted to the electrical domain, processed if required, and sent back after frequency conversion to the user link (here assumed to be in the Ka-band). In the return channel, data from the user link is modulated on multiple lasers, each one at a different center wavelength, multiplexed, amplified and after correction of the point-ahead angle, coupled into the telescope system and sent back to the ground station through the feeder link.

The forward link is the most challenging because of the following aspects:

- The atmospheric channel in the feeder uplink introduces signal fluctuations due to the combined effect of scintillation and beam wander, which produces a high dynamic range in the received power. Methods to minimize the signal fluctuations are implemented in the ground segment.
- Low-noise and high-sensitivity pre-amplifiers are a key technology for robust communications. The amplifiers must deal with the large dynamic range.
- Frequency up-conversion for the Ka-band user signals is required and represents an important part of the communications payload. Direct conversion from baseband is not applied, so as to avoid nonlinearities. An intermediate frequency such as C-band can be used to simplify the onboard up-conversion. However, this approach limits the capacity of the optical channel, becoming very inefficient. Another alternative is to perform the frequency conversion using optical technologies, based on coherent heterodyne reception [34.81]. This approach requires further optical components aboard the satellite, such as lasers, modulators and receivers, but it is quite straightforward to be integrated in an analog trans-

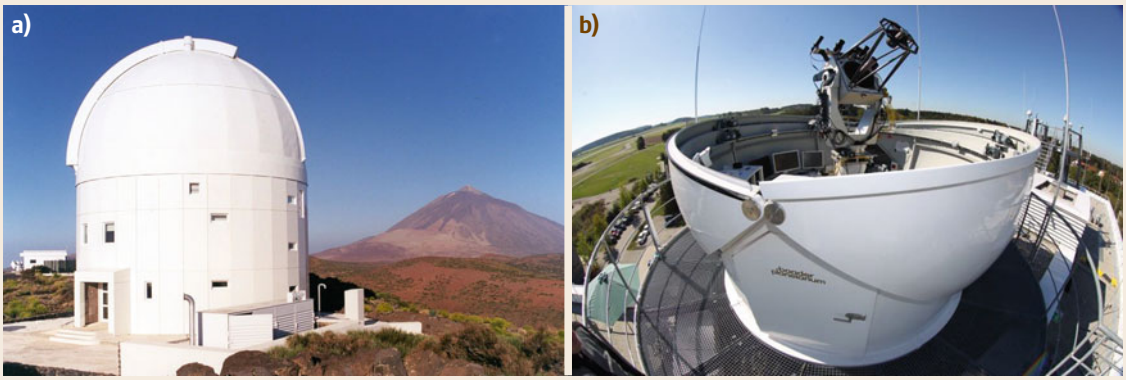


Fig. 34.34a,b ESA optical ground station in Tenerife, Canary Islands (credit ESA) **(a)**. DLR optical ground station in Oberpfaffenhofen **(b)**

parent option, since the signal can be directly converted in the Ka-band without further steps in the optical domain, simplifying the onboard system a great deal.

- Onboard processing is the key for future digital satellite systems, especially when dealing with throughputs of Tb/s. Heat dissipation, power consumption and mass are design drivers of future space-qualified high-speed processors. This technology is still not available for such high-data-rate applications, but it may evolve in the coming years.

For a relay scenario, the approach is similar, having only one optical channel and one RF channel, with the return channel being the high-speed carrier of the data. In this case, the users are LEO satellites or HAPs (high-altitude platforms), and they would make use of an inter-satellite link to transmit the data to the relay GEO satellite, and the feeder link may be implemented in the Ka-band, as mentioned in Sect. 34.4.2. The main technical challenges are similar in this case, especially when increasing the data rate. The communications payload is smaller, however, than a system supporting feeder links, and since there is an operational system already in space, new generations of such systems seem closer in time.

34.4.4 Ground Segment

The main element in the ground segment is the telescope, whose major developments come from astronomy, where telescope diameters have reached 10 m, as in the GranTeCan in the Canary Islands, the Keck 1 and 2 in Hawaii, or the Hobby–Eberly telescope in Texas. The new generation of telescopes will reach the 30 m class, for example ESO's Extremely Large Telescope in Chile or the Thirty-Meter Telescope (TMT), with construction planned to start in 2020. For optical communications, small-class telescopes have been used. Experiments with GEO satellites have been performed with the 1.016 m telescope of the ESA Optical Ground Station (ESA-OGS) in Tenerife, Canary Islands (Fig. 34.34a), and the 27 cm aperture of the Transportable Adaptive Optics Ground Station (TAOGS) (Fig. 34.35).

The ESA-OGS in Tenerife (Fig. 34.34a) is located in the *Observatorio del Teide* in Tenerife, Canary Islands (Spain). The telescope is mounted on an equatorial mount, and it has a Coudé path that allows the light to be redirected from the telescope to an optical table. The telescope was built in 2001 to support ground-to-GEO satellite links with the optical terminal aboard the Artemis satellite within the SILEX project [34.56].

The transportable adaptive optics ground station (TAOGS) (Fig. 34.35a) was developed by DLR and Tesat Spacecom to support ground-to-satellite optical links with the optical terminals aboard Alphasat and the EDRS satellites. This ground station is equipped with an adaptive optics system that performs single-mode fiber coupling for data reception. All the equipment is installed in a container, allowing for ease of transport.

The DLR optical ground station in Oberpfaffenhofen (DLR-OP-OGS) (Fig. 34.34b) has a telescope on an azimuth-elevation mount with a 40 cm Ritchey–Chrétien telescope. The station will include an 80 cm telescope with a Coudé path, which will be operational by 2020. The TOGS station (Fig. 34.35b) enables having a compact system, folding the 60 cm telescope into a carbon fiber box. All these stations are typical examples of ground segment infrastructure for GEO satellite communications.

For the downlink, the large telescope diameter has a beneficial impact on the link budget, as the receiver gain increases with the receiver diameter, as shown

Fig. 34.35a,b DLR transportable adaptive optics ground station **(a)**. DLR transportable optical ground station **(b)**

in (34.17). This means that a telescope with a larger-receiver may allow higher data rates in the downlink. However, an adaptive optics system is typically required because the atmospheric turbulence limits the optical signal coupling, as explained in more detail in the next subsection.

For the uplink, the transmitter size is also limited by the turbulence, and especially by the pointing accuracy, which is limited by the beam wander, as described by (34.35). When transmitting and receiving through different apertures, as in the DLR-OP-OGS, TAOGS and TOGS, the beam wander cannot be minimized, and therefore the divergence must cope with all the beam wander movements to ensure that the uplink reaches the satellite most of time, as shown in (34.40). The root-mean-squared value of the beam wander can be on the order of tens of microradians

$$\sqrt{\sigma^2_{\text{pointing}}} = 0.73 \left(\frac{\sqrt{2}\lambda}{D_\text{T}}\right)\left(\frac{D_\text{T}}{\sqrt{2}r_0}\right)^{5/6} . \qquad (34.40)$$

Transmitting and receiving from the same aperture allows for sampling of the tip/tilt in the right location, enabling a reduction in the beam wander and therefore a reduction in the transmitted beam divergence (involving an increase in the transmitting aperture) and better power efficiency [34.82]. The main drawback of transmitting and receiving from the same aperture is the isolation between the two directions. Circular polarization is typically used in satellite lasercom because it renders the system insensitive to rotations of the optical terminal when pointing or during satellite movement. Therefore, left-hand and right-hand circular polarization are frequently used for downlink and uplink separation; however, typical isolation of the polarization splitters/beam combiners falls around 20–30 dB. In order to increase the extinction ratio, a typical approach is to combine polarization with wavelength discrimination.

The main components of the ground segment for a very-high-throughput communications system are shown in Fig. 34.36. The telescope transmits and receives the data to/from the satellite, the coarse-pointing system points towards the satellite and keeps the pointing error small enough to allow the fine-pointing assembly to compensate for the remaining angle-of-arrival fluctuations of the signal. The wavefront sensor (WFS) and the deformable mirror (DM) are part of the adaptive optics system which compensates for the phase distortions of the atmospheric turbulence. Both the adaptive optics and pointing systems are applied for both link directions, in the downlink for fiber coupling and in the uplink for pre-compensation of the beam wander and phase distortions. In the downlink, the light is coupled into a single-mode fiber, pre-amplified, demultiplexed and converted to the electrical domain for FEC and data processing before sending the data to the network. For the uplink, the data coming from the network is converted to the optical feeder link format, modulated at each laser carrier, multiplexed, amplified and, after compensating the point-ahead angle, coupled into the telescope system to transmit the signal towards the satellite.

Some systems are required to make the ground station suitable for reliable and stable ground-to-GEO satellite communications. These systems are discussed in the following subsections:

- Adaptive optics for atmospheric turbulence compensation.
- Point-ahead angle and references for uplink pre-correction.
- Aperture diversity for turbulence mitigation.

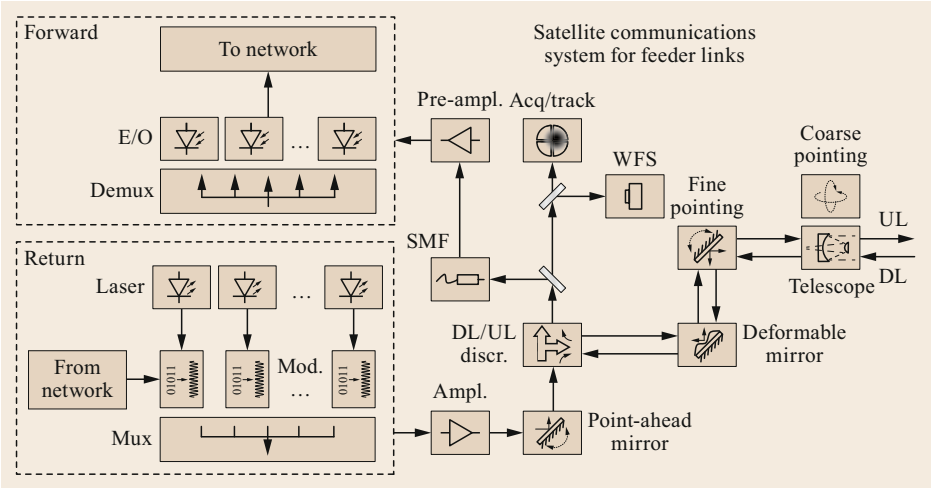


Fig. 34.36 Main components of the ground station for a very-high-throughput communications satellite based on optical feeder links

Adaptive Optics for Atmospheric Turbulence Compensation

The light collected by the telescope usually needs to be coupled into a single-mode optical fiber in order to make use of all components developed for fiber communications such as low-noise amplifiers or demultiplexers. At the ground station, the light wavefront is distorted due to the turbulence, limiting the performance of the fiber coupling. Phase distortions create intensity speckles at the focal plane of the telescope, which change size and position randomly. As a result, the coupling will usually add fading to the communications link. The coupling efficiency depends on the correlation between the received field and the coupling mode of the fiber. By correcting the phase distortions of the received field, the correlation with the coupling mode of the fiber increases, helping to stabilize the amount of coupled light. The ratio between the aperture diameter and the Fried parameter D/r_0 defines the *amount of phase distortions collected by the telescope*. For telescopes smaller than the Fried parameter, only tip/tilt correction is required, since no phase aberrations are on the receiver aperture collection area. But this implies low telescope gains, which has an impact on high-data-rate links and GEO link distances, making such an approach impracticable. By increasing the telescope diameter, adaptive optics is required, and its complexity grows with increase of the telescope diameter.

The same requirements apply for coherent reception, where the incoming light is mixed with a local oscillator. To maintain high heterodyne efficiency, the received wavefront needs to match the local oscillator in order to demodulate the signal. Therefore, an adaptive optics system is required to correct the phase distortions and maintain stable coupling or coherent reception. This kind of systems is widely used in astronomical telescopes to boost imaging performance towards the diffraction limit [34.83]. The conditions for lasercom links, however, are different. While astronomy usually works around the zenith from astronomical sites (high locations), in communications, lower elevation angles are targeted (current EDRS satellites are at around 30° elevation from central Europe), and ground stations are also located at lower altitudes. Only corrections of small fields of view are necessary in communications, because the tracking keeps the counter-partner aligned, but the turbulence requirements are more demanding than in astronomy.

The main elements that constitute an adaptive optics system are the tip/tilt mirror, the wavefront sensor (WFS) and the deformable mirror (DM). They are shown in Fig. 34.37. As explained earlier, the tip/tilt mirror compensates the angle-of-arrival fluctuations due to the atmosphere and any tracking errors, which cause spot shifts in the focal plane. The WFS estimates the phase of the received wavefront, and a control computer computes the signals to drive the DM. The DM surface shape is formed by a set of actuators, and it is adapted to the received beam in order to conjugate (compensate) its phase distortion. For the concerned scenarios, the atmosphere can reach frequencies on the order of kHz, which places very strong requirements on the adaptive optics system closed-loop speed.

An adaptive optics system cannot perfectly compensate for the phase distortions, due to the limited number of actuators of the DM, for example, or the limited bandwidth of the control loop. As an example, from these two aspects, one can derive requirements for the control loop and for the required number of actuators. The number of actuators is determined directly from the square of the ratio between the beam diame-

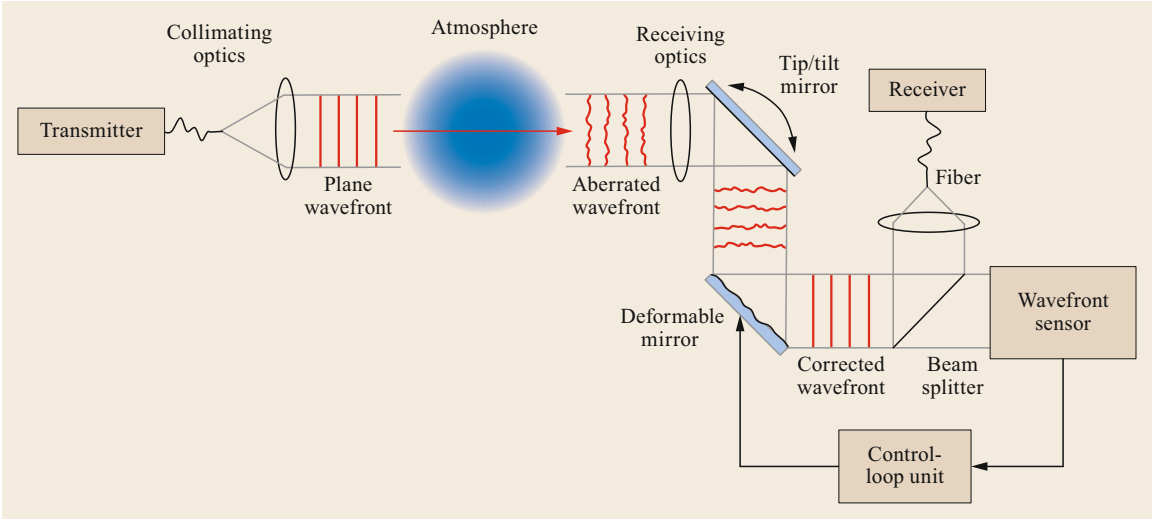


Fig. 34.37 Block diagram of an ideal adaptive optics system

ter and the actuator space, which has an impact on the residual phase error as a result of the limited resolution by fitting the received phase [34.83].

Recalling the definition of the Greenwood frequency in (34.31), which defines the characteristic frequency of the atmosphere, further requirements for the adaptive optics closed loop can be defined. Similarly to (34.33), the residual phase error due to the limited bandwidth of the closed loop is defined by (34.41) [34.84]

$$\sigma^2_{\theta,\text{loop}} = \left(\frac{f_\text{G}}{f_{\text{AO},3\,\text{dB}}}\right)^{5/3} . \qquad (34.41)$$

Another important element of the adaptive optics system is the WFS, which estimates the phase of the received light. Typically, Shack–Hartmann sensors are used as WFS given their relatively simple hardware configuration. This kind of sensor is used chiefly in astronomy, usually working around the zenith, where the turbulence strength is lower [34.85]. This imposes some limitations for communications where turbulence can become moderate to strong. Shack–Hartmann sensors are based on the measurement of the phase gradient by means of an array of lenses placed at the pupil. The focus of each lens is imaged by a camera, and the phase is reconstructed by means of a centroid- or correlation-based algorithm. The main limitation of such an approach appears when turbulence increases. In this case, some of the sub-apertures do not receive enough light due to scintillation, introducing errors in the reconstruction. Furthermore, the appearance of a rotational component of the phase under strong turbulence has been studied, with the conclusion that it cannot be observed by sensors based on the measurement of the phase gradient, such as Shack–Hartmann or curvature sensors [34.86]. As an alternative, interference-based WFS, iterative methods or a combination of two sensors are options to be considered to achieve a more resistant wavefront estimation under strong turbulence [34.87, 88].

Point-Ahead Angle and References for Uplink Pre-Correction

In Fig. 34.38, there is a representation of a bidirectional link between a satellite and a ground station. The ground station receives the downlink and tracks the incoming light. Because of atmospheric turbulence, a tip/tilt mirror is required for compensating for the angle-of-arrival fluctuations. In the meantime, the satellite moves, and the ground station needs to point ahead by a certain angle, as explained in Sects. 34.1.2 and 34.3.3. This point-ahead angle for a GEO satellite is around 18 μrad. The pointing direction of the uplink will fluctuate due to the atmospheric turbulence, the so-called beam wander, as explained in Sect. 34.2.2 and described by (34.28). Using the measurement of the angle-of-arrival fluctuations, the uplink could ideally be compensated against beam wander by applying the same fluctuations as a pre-compensation. Unfortunately, due to the point-ahead angle, the uplink and downlink travel through different atmospheric paths. The atmospheric effects are correlated, however, within a certain cone angle, called the isoplanatic angle, as represented by the green triangle in Fig. 34.38. The isoplanatic angle takes into account the correlation of all phase distortions. There is a similar definition for only the tip/tilt effects, called isokinetic angle. How large both angles are depends on the turbulence conditions.

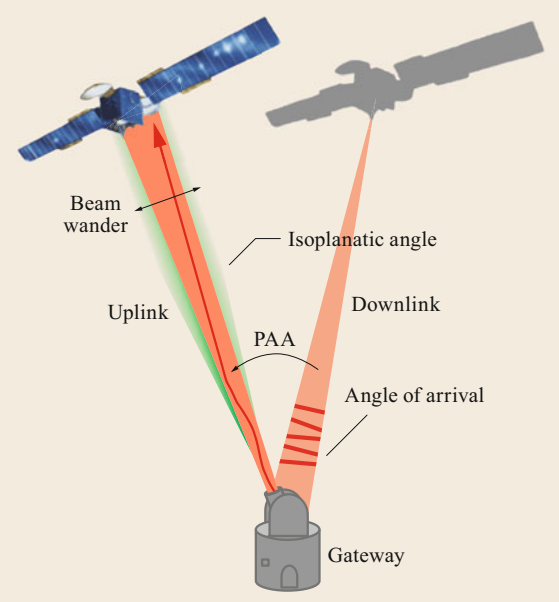


Fig. 34.38 Isoplanatic angle and point-ahead angle in satellite bidirectional links

As a rule of thumb, the isokinetic angle is about 1.5–2 times larger than the isoplanatic angle, meaning that higher aberration orders of the phase distortions have smaller coherence angles, and thus shorter coherence times. This is, however, not an ON/OFF process. Being outside the coherence cone of the atmosphere means an increase in the decorrelation between the two paths, and therefore higher residual beam wander, which leads to higher intensity fluctuation at the satellite: the larger the separation between the uplink and downlink, the larger the decorrelation, until the two paths are completely independent. The most straightforward way to mitigate beam wander is to increase the divergence, as noted by (34.35). The main drawback is the decrease in the mean power at the satellite due to a larger divergence beam. In the end, this is a link-budget optimization between the beam divergence, the probability of outages and the received power, as applied in [34.79].

Another alternative is to have a reference in the direction of the uplink, as used in astronomy with laser guide stars [34.89]. Laser guide stars based on Rayleigh scattering have been studied since the early 1990s in astronomy [34.90, 91]. The main drawback of Rayleigh scattering is that this effect happens mainly in the first km of the propagation, limiting the adaptive optics performance due to the short distance at which the reference signal is located [34.92]. The most promising technology is laser guide stars based on the sodium atom excitation in the mesosphere, sodium layer of the atmosphere, at a height of about 90 km [34.93]. By transmitting a laser at 589 nm, the sodium atoms in this layer are excited and produce an *artificial star* that can be used as a reference for adaptive optics. In astronomy, this laser is transmitted within the isoplanatic angle of the observation direction in order to apply adaptive optics on the observed object. In communications, this technique could be applied in the direction of the uplink in order to apply pre-correction adaptive optics [34.79].

As described in Fig. 34.37, adaptive optics corrects for the phase distortions in the downlink in order to enable single-mode fiber coupling. The same system can be used for transmitting a laser in the other direction, pre-distorting the phase. The goal is to decrease the intensity fluctuation produced by the atmospheric phase distortions. Since the isoplanatic angle may be smaller than the point-ahead angle, a laser guide star could provide a reference in the uplink direction that could be used for correcting the phase. The main point to be solved remains in the tip/tilt correction. By observing the laser guide star from the transmitter direction, no tip/tilt can be measured because the light travels up and down through the same path. In order to use the laser guide star for tip/tilt correction, an off-axis observation is required, as proposed in [34.94]. Currently, there are experiments targeting a demonstration of real-time compensation of the beam wander by means of laser guide stars [34.95].

Aperture Diversity for Turbulence Mitigation

The integration of an adaptive optics system at the ground station is not the only approach for minimizing the impact of turbulence. The use of transmitter diversity and receiver diversity is also a possibility, and it has been applied at both the pupil and focal plane.

At the pupil plane, spatial diversity exploits the decorrelation between different optical paths through the atmosphere in transmission or reception. As rule of thumb, for transmitted beams, the atmospheric paths are assumed to be uncorrelated when they are half a meter apart. In this case, the transmitted beams overlap after some km of transmission, so that the receiver on the satellite receives only the combination of all the beams, and hence fluctuations average out. Statistically, scintillation decreases linearly with the number of transmitted beams, when each path is assumed to be independent of the other.

This approach works properly when no data is modulated on the laser and the signals come from independent laser sources, such as for uplink beacons. If only one single-mode source is used—for example, by splitting the signal between transmitters—the combination of the signals at the satellite can lead to interference patterns, because the path difference between them may

lie within the coherence length of the laser. Applying a delay between transmitters may be an option, for example by transmitting the light through some km of fiber.

In the SILEX ground station, 4× transmitter diversity was implemented (four transmitted uplink lasers separated by around half a meter) using one single multimode laser, but the results were not the expected [34.96]. In the case of multimode lasers, the visibility of the interference pattern is a periodical function of the path difference between laser signals; therefore, delay between transmitters must be accurately selected [34.82].

In the case that data is modulated on the laser carrier, the bandwidth of the signal increases. An overlap of even only a part of the bandwidth will lead to strong interference. In this case, the signals may be separated in polarization or the wavelength for example. This can be done because the chromatic dispersion around the communications wavelength is very low, and the polarization remains unchanged by transmitting through the atmosphere. But neither approach is really advantageous. Usually, polarization (in combination with the wavelength) is used for the separation between link directions. And wavelength separation (diversity) requires the data spectrum times the number of transmitted beams, which is unfeasible for very high throughputs. More elaborate approaches such as generating single sidebands may be an alternative to place two transmitters in the same bandwidth [34.97], but its application for coherent communications schemes still needs to be investigated, since most of these approaches require noncoherent detection. An Alamouti transmitter diversity scheme would be the best option, being compatible with coherent modulation schemes, but the receiver complexity at the satellite increases.

For the reception, diversity can be applied by an array of telescopes. This approach has been designed and tested mainly for deep-space scenarios, especially to avoid very large monolithic mirrors—for example, in the lunar link within the LADEE mission, where the LLGT telescope was developed. In this case, photon-counting technology was used, and the four received signals were combined in the electrical domain.

If the telescope apertures are small enough, adaptive optics could be avoided, and only tilt correction is needed. However, signals need to be combined either electrically or optically, which increases the hardware complexity at the receiver side, by aligning the phases of the received signals optically or by increasing the number of receivers. In [34.98], receiver diversity is shown by combining four coherent receivers electrically. The combined signal shows the reduction of the atmospheric fluctuations. This assumes, however, that the aperture diameter of each telescope is large enough to be within the signal sensitivity, which limits the minimum diameter of the telescope, which is again a trade-off with the expected turbulence, since we are assuming that no adaptive optics is needed.

The same diversity approach could be followed at the focal plane by, for example, setting a fiber bundle. In the presence of turbulence, the intensity is distributed randomly at the focal plane: several speckles change position and intensity. A fiber bundle can collect more power and combine the signals optically or electrically, following the same approach as before. Such approaches have been developed to increase the FOV of lasercom systems [34.99]. The same idea can be applied using a multimode fiber. In [34.100], a multimode fiber connected to multiple single-mode fibers was developed and demonstrated. In this case, the phase distortions may excite several modes in the fiber, and each one is coupled in different single-mode fibers. Another idea is to bring the optical power distributed among the different modes back to the main mode, since most of components developed for fiber communications, in particular those related to DWDM, are based on single-mode fiber.

34.5 Future Optical Satellite Networks

A satellite network can be defined as a set of satellites at different orbits, combining GEO, MEO and LEO, for exploiting the visibility of the higher orbits and the short delays of the lower orbits. Although many past studies have sought to define satellite network architectures and their traffic management, current systems are focused on only one orbit, mainly LEO and GEO. Systems that have been announced thus far will complement the satellite services with further GEO satellites and constellations and LEO and MEO, focusing mainly on providing Internet access.

The following sections provide a short description of current satellite systems, their main applications and finally some considerations on the network architecture.

34.5.1 Current and Upcoming Satellite Systems

Looking at the current systems in the different orbits, in GEO, the satellite communications market is mainly focused on video broadcasting and Internet access, and on Ka-band and Ku-band technology. Some examples of

HTS satellites in service are Ka-Sat with 70 Gb/s and Eutelsat 172B, based on Ku-band technology and operating in Asia-Pacific; ViaSat1 and ViaSat2 with 140 and 350 Gb/s; Inmarsat Global Xpress, which are four GEO satellites offering global coverage; Intelsat (Epic Series); and GSAT-19 covering India.

The MEO orbit is used primarily by navigation systems such as GPS, Galileo, Beidou (also with GEO satellites) and Glonass. This orbit is also used by the O3b constellation for broadband communication. Sixteen satellites provide 16 Gb/s per satellite, with 700 km beam footprints. A MEO constellation is proposed by Laserlight Communications [34.101], who is planning to deploy an all-optical constellation with 12 MEO satellites in collaboration with Optus.

LEO constellations are traditionally used for mobile satellite service, with both global access, as in the case of the fully inter-satellite and inter-orbit meshed IRIDIUM network, or regionally with bent-pipe transponders, as in the Globalstar and Orbcomm networks.

Currently, LEO satellite constellations offer relatively low data throughput. Examples are as follows:

- Iridium next-generation comprises 66 satellites providing L-band communication to mobile users (128 kb/s), up to 1.5 Mb/s to Iridium Pilot marine terminals, and high-speed Ka-band service up to 8 Mb/s to fixed/transportable terminals.
- The Globalstar second-generation constellation will consist of 24 satellites and offer mobile voice communications and low-data-rate transfer. This system is used in monitoring areas such as oil and gas, government, mining, forestry, commercial fishing, military applications or transportation.
- Orbcomm is 100% dedicated to M2M communications. The constellation comprises 50 satellites.

New developments are in the beginnings of 2020 taking place in LEO constellations:

- LeoSAT is a satellite constellation with around 100 satellites planned, which aims at deploying broadband services with user access rates from 50 Mb/s up to 1.2 Gb/s, with optical inter-satellite links and user/gateway links in the Ka-band, and onboard processing on each satellite.
- Starlink is a satellite constellation developed by SpaceX with the support of Samsung, with around 12 000 satellites planned, with optical inter-satellite links and ground receivers with phased-array antennas.

On the other side, not all satellite constellation proposals would become successful. For example one of the first companies of the new space era, Oneweb, who planned to have about 900 operation satellites by 2019, reported bankruptcy in the beginning of 2020, after the launch of 74 satellites. By now it is difficult to anticipate what will happen with this growing new market, but it could have an impact on the time line of future developments on optical technologies in space.

34.5.2 Applications for Future Optical Satellite Networks

A variety of applications would profit from a global satellite network, where optical frequencies can play a key role, increasing data throughput and freeing up currently allocated RF frequencies. The main applications are divided into terrestrial, maritime and aeronautical.

Broadband Internet access for private users and for industry is the main terrestrial application, which is currently strongly pushed in Europe. In its Digital Agenda for Europe, the European Commission set a target of 30 Mb/s per home in the EU in 2020. A study of broadband coverage in the EU [34.102] found that coverage outside the large cities is only 30% in most EU countries. Fiber optics to serve the needs of areas with sparse populations is an expensive infrastructure. Satellite communications may be a complementary solution to the currently existing infrastructure. In this case, the requirement is 30 Mb/s or more per user. For such a scenario, optical feeder links seem to present a good alternative given the massive nonregulated spectrum.

5G targets very low latency and high capacity using high-frequency bands, but the coverage at these frequencies is extremely limited, and they are foreseen to be as close as possible to the end user to minimize latency. The coverage decreases with the latency and inversely with the capacity. The bands are assigned to the new 5G applications and the legacy to support 2G, 3G and 4G [34.103]. Therefore, current deployment of the 5G network would benefit from a satellite system that allows new stations to be set outside the optical network coverage. The 5G network is targeted to revolutionize the industry with the so-called Industry 4.0 monitoring and telecontrol of production. A particular interest of optical communications in this field is the use of the optical frequencies itself which avoid potential frequency overlaps.

A tactile Internet would also be a benefit of broadband satellite communications. This concept includes applications such as online gaming, virtual reality, robotics and telepresence, with latency constraints down to 1 ms and capacity constraints up to 1 Gb/s, or factory automation and heath care, with latency constraints down to 10 ms and capacity constraints up to

Table 34.6 Latency in ms for terrestrial Internet connections and round-trip satellite connections

Terrestrial fiber network [34.106]	
Transatlantic	73.2
Europe	11.2
North America	36.6
Intra-Japan	9.3
Transpacific	102.1
Asia-Pacific	97.9
Latin America	131.5
EMEA to Asia-Pacific	130.6
Satellite round-trip	
LEO	18.4
MEO	93.5
GEO	272.1

100 Mb/s [34.104, 105], and these are only some examples.

Current latency measurements of the terrestrial network latency, provided by [34.106], show values of around 11 ms within Europe, below 75 ms for transatlantic communications or around 100 ms for transpacific communications, as shown in Table 34.6. These values can be compared with the round-trip latencies for the satellites in Table 34.6, assuming ground-to-satellite links at 10° elevation. As expected, LEO satellites can offer lower latency due to their proximity to the Earth; however, the ultra-low-latency constraints of the tactile Internet cannot be respected by any satellite connection, or indeed by current terrestrial infrastructure.

A mobile edge cloud close to the end users should provide a cloud-based platform for all applications with very low latency. In this architecture, the latency constraints are set in the connection between the users and the mobile edge cloud, by optimizing the communication's physical layer:

> Caching and in a more general category, information centric networking, can be assumed as one of the promising candidate technologies to design a paradigm shift for latency reduction in next generation communication systems. [34.105]

Satellite communications may play a role in the connection between the core network and the mobile edge cloud.

For maritime applications, tracking and monitoring of vessels, people and goods, and sharing information between vessels is becoming essential, and reliable worldwide communication for vessels is required. Polar areas, particularly in the Arctic above 76° N, are especially poorly covered. Future maritime transport of goods will require better communication services to enable the receipt of updated information on the route (e.g., ice, currents, weather) and to keep track of vessels and goods. In the future, autonomous transport or remote control of vessels will help to meet the needs of the growing goods transport.

The main use cases are as follows [34.107]:

- e-Navigation is a project of the International Maritime Organization (IMO) towards a future digital concept for the maritime sector. This project includes on-route information updates (destinations, waypoint and route optimization, map updates, weather information, ice information or currents information), low-cost onboard communication services for crew entertainment and the automated information system (AIS) satellite-based data exchange system using the commonly carried very high frequency (VHF) equipment for complementing the terrestrial network access. In this respect, the VHF data exchange system (VDES) will also extend the coverage in Arctic regions. The expected traffic is on the order of 400 (kB/h)/ship. This system is contingent on the approval of the WRC19.
- Arctic, and polar regions in general, are not well covered by current communication systems. Communication using GEO satellites is theoretically possible up to 81° N, but typically only up to 76° N. Inmarsat satellites serve the Arctic up to 76° N except for an area around 120° E of the Laptev Sea in Russia and around 120° W of the Beaufort Sea in Canada. The polar orbiting satellites of Iridium and Cospas-Sarsat serve the whole of the Arctic. IMO has procedures in place to possibly recognize satellite systems in addition to the systems provided by Inmarsat and Cospas-Sarsat.
- Autonomous ships: Unmanned merchant ships on intercontinental voyages with advanced sensor systems to detect and avoid obstacles, and enable advanced onboard control, positioning and navigation systems to determine and control the exact location, speed and course as well as route. In this case, a latency of less than one second is required so that ships can be remotely controlled in real time. A bandwidth of up to 4 Mb/s is required to send radar and video pictures.
- The Global Maritime Distress and Safety System (GMDSS) is a radio system whose techniques and frequencies are defined by the ITU and for which mandatory equipment carriage requirements have been adopted by the IMO for commercial vessels. A satellite network may provide alternatives to high-frequency (HF) (3–30 MHz) communications, where the traditional means of long-distance communication for ships are still used as a backup to GEO satellite services.

In aeronautical communications, modernization of the current air traffic management (ATM) system is anticipated. The ATM system manages all aircraft in controlled airspace. The system uses analog double-sideband amplitude modulation (DSB-AM) deployed in the VHF band between 118 and 137 MHz to communicate with the pilots. This modulation is very spectrum-inefficient, and voice communications cannot cope with increasing air traffic needs. As predicted by the EUROCONTROL statistics and forecast service [34.108], air traffic will increase by 50% by 2035, which will bring the current ATM system to its limits, especially in the densest flight regions like Europe and the United States. In the 1990s, the International Civil Aviation Organization (ICAO) standardized the VHF data link (VDL) standards, a digital system with 25 kHz bandwidth, enabling data communications. However, the link capacity is well below the requirements of aeronautical communications.

ICAO recommended the use of the L-band between 960 and 1164 MHz, but ensuring the coexistence of already existing systems. In this frame, the L-band Digital Aeronautical Communications System (LDACS) was developed based on orthogonal frequency-division multiplexing (OFDM) together with adaptive coding and modulation, exploiting the 500 kHz bandwidth available in the L-band. The communication ranges from 561 kb/s, using strong coding and robust modulation, to 2.6 Mb/s, using higher modulation orders and weak coding [34.109]. LDACS standardization is currently under way in ICAO, and was planned to start in 2018.

A satellite-based communication solution for the European ATM system is driven by ESA in Iris, the ARTES Satellite Communication for Air Traffic Management program, in partnership with Inmarsat. High-capacity digital data links via satellite could enable the transfer of information on latitude, longitude, altitude and time to monitor and adjust the aircraft route efficiently when necessary, for example due to a change in weather conditions. The target is 2028, when Iris will support the service around the globe.

In summary, aeronautical applications can profit from direct broadband access to a global network supporting onboard WLAN (passenger aircraft) or real-time transmission of reconnaissance data (governmental and disaster relief missions). Similarly, maritime applications will profit from bandwidth enhancements with regard to existing narrowband and medium-bandwidth services such as INMARSAT's BGAN services. Land applications are expected to include broadband fixed and nomadic access as well as intermediate- and low-bandwidth mobile access. Industry 4.0 and IoT (Internet of Things) applications are expected to be supported as well, for example, connecting dispersed sensors, control and command of remote installations such as satellite ground terminals, and fleet management.

34.5.3 Network Architecture

To date, optical satellite communication has been developed for point-to-point data transmission, either through direct LEO downlinks or through data relay over GEO. But space lasercom can potentially have a huge impact on enlarging the ground network infrastructure to space, achieving global coverage of areas lacking ground infrastructure with broadband access.

From the radio-frequency systems point of view, satellite networks have long been investigated, and schemes have been proposed that combine different constellations in LEO and MEO [34.110], or combine GEO, MEO and LEO [34.111]. Routing in satellite networks has been studied defining several layers. For example, in [34.111], MEO satellites manage traffic in the LEO constellation within their footprint, and GEO satellites monitor and manage the entire system, informing the gateways about traffic congestion or inter-satellite link failures. The design of the switching and monitoring approach among layers to guarantee a desired QoS (quality of service) depends on the number of layers (constellations and satellites per orbit), the number of inter-satellite links and the available throughput. The design of such a system may need to consider a combination of radio-frequency and optical links to cope with the many possible applications.

LEO satellite constellations have been considered for some time given their advantage of short delays, more favorable power budgets and closer distances than MEO and GEO. For satellite-to-ground optical communications, this last aspect is at first glance no longer a clear advantage because of the limited pointing accuracy due to the larger point-ahead angles at LEO and MEO, which prevent beam wander compensation, as discussed in Fig. 34.38. As a rule of thumb, one may need a beam divergence 5–10 times as large for LEO communications as for GEO. The main drawback is therefore a reduction in the received power, leading to similar transmitted power requirements for LEO and GEO as with current technologies. The advantage of having LEO closer is lost because of the point-ahead angle. Furthermore, LEO satellite payloads will be smaller because of the larger number of satellites; thus the constellation may rely on low-cost deployment. This means that telescopes aboard the satellite will tend to be smaller, having penalties on receiver gain as well. In the case of satellite-to-ground radio-frequency links, multiple gateways must be active to achieve high throughput.

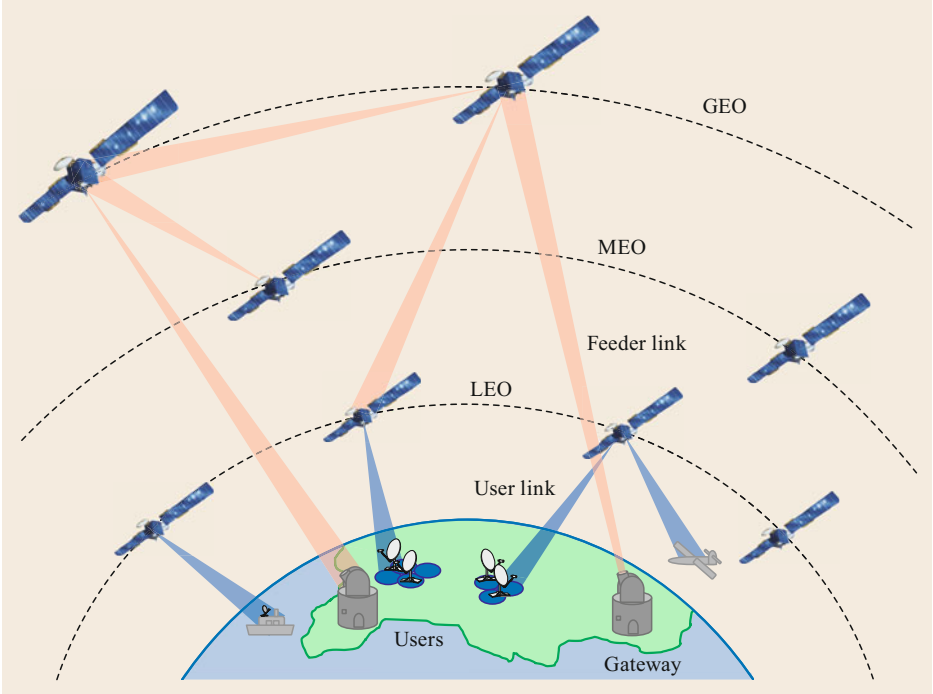


Fig. 34.39 Optical satellite communications network

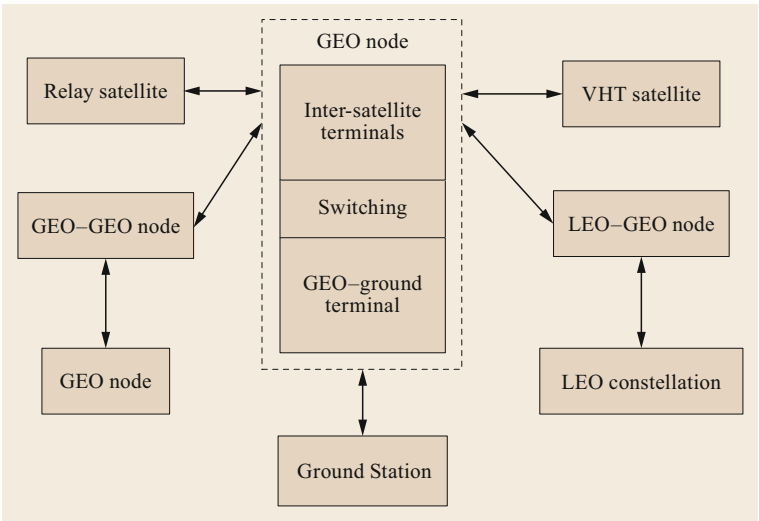


Fig. 34.40 Conceptual block diagram for an optical satellite network

GEO satellites are primarily more expensive in the development and deployment, because of the larger distance to be covered in launch and the higher requirements due to radiation. The main advantage of GEO satellites is the large coverage that achieves one satellite, and due to the properties of the atmosphere, the small divergences of the optical beams can be tightened narrower than in LEO or MEO. Most likely, an optical global network will be a combination of the three orbits (as represented in Fig. 34.39), optimizing the QoS and combining several applications, maybe even including navigation services in MEO or in LEO, as proposed by Stanford University [34.112].

A system combining a variety of applications may need to combine a mesh configuration with satellite constellations. Figure 34.40 shows a concept for a satellite network. The development of ad hoc satellite platforms, for backhaul and switching between ground and GEO for example, would enable increased data throughput of the feeder links. In this case, the GEO

platform could incorporate signal regeneration and optical switching towards other application-oriented satellites. This is especially interesting for onboard signal processing, for example for error correction algorithms protecting the data across atmospheric turbulence. GEO satellites at shorter distances can be easily connected with optical links under limited power requirements and transport of high data volumes. Application-oriented satellites, for example based on data relay or VHT communications, could have simplified payloads, fostering their own applications. Other dedicated platforms may connect other GEO nodes at large distances, for example, between Europe and Asia or to LEO constellations. Optical frequencies are the best candidate for such networks: they are more efficient in mass and power, they are more resistant to interferences and it solves the RF spectrum bottleneck. However, a combination of the optical and RF technologies is required to satisfy the requirements of such a vast variety of applications.

References

34.1 K.E. Wilson, J.R. Lesh, T.Y. Yan: GOPEX: A laser uplink to the Galileo spacecraft on its way to Jupiter, Proc. SPIE (1993), https://doi.org/10.1117/12.149251

34.2 K. Araki, Y. Arimoto, M. Shikatani, M. Toyoda, M. Toyoshima, T. Takahashi: Performance evaluation of laser communication equipment onboard the ETS-VI satellite, Proc. SPIE (1996), https://doi.org/10.1117/12.238434

34.3 G.D. Fletcher, T.R. Hicks, B. Laurent: The SILEX optical interorbit link experiment, Electron. Commun. Eng. J. **3**(6), 273–279 (1991)

34.4 J. Horwath, N. Perlot, M. Knapek, F. Moll: Experimental verification of optical backhaul links for high-altitude platform networks: Atmospheric turbulence and downlink availability, Int. J. Satell. Commun. Netw. **25**(5), 501–528 (2007)

34.5 M. Toyoshima, H. Takenaka, Y. Shoji, Y. Takayama, Y. Koyama, H. Kunimori: Results of Kirari optical communication demonstration experiments with NICT optical ground station (KODEN) aiming for future classical and quantum communications in space, Acta Astronaut. **74**, 40–49 (2012)

34.6 V. Cazaubiel, G. Planche, V. Chorvalli, L. Le Hors, B. Roy, E. Giraud: LOLA: A 40000 km optical link between an aircraft and a geostationary satellite, Proc. SPIE (2006), https://doi.org/10.1117/12.2308161

34.7 R. Fields, C. Lunde, R. Wong, J. Wicker, D. Kozlowski, J. Jordan, B. Hansen, G. Muehlnikel, W. Scheel, U. Sterr, R. Kahle, R. Meyer: NFIRE-to-TerraSAR-X laser communication results: Satellite pointing, disturbances, and other attributes consistent with successful performance, Proc. SPIE (2009), https://doi.org/10.1117/12.820393

34.8 J. Horwath, C. Fuchs: Aircraft to ground unidirectional laser-communications terminal for high-resolution sensors, Proc. SPIE (2009), https://doi.org/10.1117/12.808869

34.9 D.M. Boroson, B.S. Robinson, D.V. Murphy, D.A. Burianek, F. Khatri, J.M. Kovalik, Z. Sodnik, D.M. Cornwell: Overview and results of the lunar laser communication demonstration, Proc. SPIE (2014), https://doi.org/10.1117/12.2045508

34.10 H. Zech, F. Heine, D. Tröndle, S. Seel, M. Motzigemba, R. Meyer, S. Philipp-May: LCT for EDRS: LEO to GEO optical communications at 1,8 Gbps between alphasat and sentinel 1a, Proc. SPIE (2015), https://doi.org/10.1117/12.2196273

34.11 A. Carrasco-Casado, H. Takenaka, D. Kolev, Y. Munemasa, H. Kunimori, K. Suzuki, T. Fuse, T. Kubo-Oka, M. Akioka, Y. Koyama, M. Toyoshima: LEO-to-ground optical communications using SOTA (Small Optical TrAnsponder) – Payload verification results and experiments on space quantum communications, Acta Astronaut. **139**, 377–384 (2017)

34.12 B. Moision, B. Erkmen, E. Keyes, T. Belt, O. Bowen, D. Brinkley, P. Csonka, M. Eglington, A. Kazmierski, N. Kim, J. Moody, T. Tu, W. Vermeer: Demonstration of free-space optical communication for long range data links between balloons on Project Loon, Proc. SPIE (2017), https://doi.org/10.1117/12.2253099

34.13 D. Giggenbach, J. Poliak, R. Mata-Calvo, C.F.N. Perlot, R. Freund, T. Richter: Preliminary results of terabit-per-second long-range free-space optical transmission experiment THRUST, Proc. SPIE (2015), https://doi.org/10.1117/12.2193902

34.14 J. Yin, Y. Cao, Y.-H. Li, S.-K. Liao, L. Zhang, J.-G. Ren, W.-Q. Cai, W.-Y. Liu, B. Li, H. Dai, G.-B. Li, Q.-M. Lu, Y.-H. Gong, Y. Xu, S.-L. Li, F.-Z. Li, Y.-Y. Yin, Z.-Q. Jiang, M. Li, J.-J. Jia, G. Ren, D. He, Y.-L. Zhou, X.-X. Zhang, N. Wang, X. Chang, Z.-C. Zhu, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, J.-W. Pan: Satellite-based entanglement distribution over 1200 kilometers, Science **356**(6343), 1140–1144 (2017)

34.15 N. Perlot: Turbulence-induced fading probability in coherent optical communication through the atmosphere, Appl. Opt. **46**(29), 7218–7226 (2007)

34.16 K. Saucke, C. Seiter, F. Heine, M. Gregory, D. Tröndle, E. Fischer, T. Berkefeld, M. Ferienčik, I. Richter, R. Meyer: The Tesat transportable adaptive optical ground station, Proc. SPIE (2016), https://doi.org/10.1117/12.2218275

34.17 P. Conroy, J. Surof, J. Poliak, R.M. Calvo: Demonstration of 40GBaud intradyne transmission through worst-case atmospheric turbulence conditions for geostationary satellite uplink, Appl. Opt. **57**(18), 5095–5101 (2018)

34.18 X. Liu, T.H. Wood, R.W. Tkach, S. Chandrasekhar: Demonstration of record sensitivities in optically preamplified receivers by combining PDM-QPSK and M-Ary pulse-position modulation, IEEE J. Lightwave Technol. **30**, 406–413 (2012)

34.19 S. Dolinar, K.M. Birnbaum, B.I. Erkmen, B. Moision: On approaching the ultimate limits of photon-efficient and bandwidth-efficient optical communication. In: *Int. Conf. Space Opt. Syst. Appl. (ICSOS)* (2011) pp. 269–278, https://doi.org/10.1109/ICSOS.2011.5783682

34.20 C.E. Shannon: Communication in the presence of noise, Proc. IRE **37**(1), 10–21 (1949)

34.21 D.M. Boroson: Channel capacity limits for free-space optical links, Proc. SPIE (2008), https://doi.org/10.1117/12.785859

34.22 A. Al-Habash, L.C. Andrews, R.L. Phillips: Mathematical model for the irradiance probability density function of a laser beam propagating through turbulent media, Opt. Eng. **40**(8), 1554–1562 (2001)

34.23 J.A. Greco: Design of the high-speed framing, FEC, and interleaving hardware used in a 5.4 km free-space optical communication experiment, Proc. SPIE (2009), https://doi.org/10.1117/12.826309

34.24 G.P. Anderson, S.A. Clough, F.X. Kneizys, J.H. Chetwynd, E.P. Shettli: *AFGL Atmospheric Constituent Profiles (0–120 km)*, Environmental Research Papers, Vol. 954 (Air Force Geophysics Laboratory, Hanscom 1986)

34.25 B. Mayer, S. Shabdanov, D. Giggenbach: *Electronic Data Base of Atmospheric Attenuation Coefficients*, DLR-Internal Document (DLR, Oberpfaffenhofen 2002)

34.26 D. Giggenbach, F. Moll, C. Fuchs, T. de Cola, R. Mata-Calvo: Space communications protocols for future optical satellite-downlinks. In: *62nd Int. Astronaut. Congr. (IAC)* (2011) p. 2957

34.27 C. Fuchs, F. Moll: Ground station network optimization for space-to-ground optical communication links, J. Opt. Commun. Netw. **7**(12), 1148–1159 (2015)

34.28 S. Poulenard, A. Mège, C. Fuchs, N. Perlot, J. Riedi, J. Perdigues: Digital optical feeder links system for broadband geostationary satellite, Proc. SPIE (2017), https://doi.org/10.1117/12.2255987

34.29 L. Andrews, R. Phillips: *Laser Beam Propagation Through Random Media*, 2nd edn. (SPIE, Bellingham 2005)

34.30 V.I. Tatarskii: *The Effects of the Turbulent Atmosphere on Wave Propagation* (Israel Program for Scientific Translations, Jerusalem 1971)

34.31 D.P. Greenwood: Bandwidth specification for adaptive optics systems, J. Opt. Soc. Am. **67**(3), 390–393 (1977)

34.32 G.A. Tyler: Bandwidth considerations for tracking through turbulence, J. Opt. Soc. Am. A **11**(1), 358–367 (1994)

34.33 R. Barrios, F. Dios: Exponentiated Weibull distribution family under aperture averaging for Gaussian beam waves, Opt. Express **20**(12), 13055–13064 (2012)

34.34 K. Kiasaleh: On the probability density function of signal intensity in free-space optical communications systems impaired by pointing jitter and turbulence, Opt. Eng. **33**(11), 3748–3757 (1994)

34.35 M. Toyoshima, S. Yamakawa, T. Yamawaki, K. Arai, M.R. Garcia-Talavera, A. Alonso, Z. Sodnik, B. Demelenne: Long-term statistics of laser beam propagation in an optical ground-to-geostationary satellite communications link, IEEE Trans. Antennas Propag. **53**(2), 842–850 (2005)

34.36 Union of Concerned Scientists: Satellite database, https://www.ucsusa.org/resources/satellite-database (2019)

34.37 M. Toyoshima, T. Takahashi, K. Suzuki, S. Kimura, K. Takizawa, T. Kuri, W. Klaus, M. Toyoda, H. Kunimori, T. Jono, Y. Takayama, K. Arai: Results from phase-1, phase-2 and phase-3 Kirari optical communication demonstration experiments with the NICT optical ground station (KODEN). In: *25th AIAA Int. Commun. Satellite Syst. Conf., Seoul* (2007)

34.38 M. Gregory, F. Heine, H. Kämpfner, R. Meyer, R. Fields, C. Lunde: TESAT laser communication terminal performance results on 5.6Gbit coherent inter satellite and satellite to ground links. In: *Int. Conf. Space Opt. (ICSO), Rhodes Island* (2010)

34.39 Q. Zhang, G. Wang: Influence of HY-2 satellite platform vibration on laser communication equipment: Analysis and on-orbit experiment. In: *3rd Int. Symp. Space Opt. Instrum. Appl.*, Springer Proceedings in Physics, Vol. 192, ed. by H. Urbach, G. Zhang (Springer, Cham 2017)

34.40 B.V. Oaida, M.J. Abrahamson, R.J. Witoff, J.N.B. Martinez, D.A. Zayas: OPALS: An optical communications technology demonstration from the International Space Station. In: *IEEE Aerospace Conf., Big Sky* (2013), https://doi.org/10.1109/AERO.2013.6497167

34.41 H. Takenaka, A. Carrasco-Casado, M. Fujiwara, M. Kitamura, M. Sasaki, M. Toyoshima: Satellite-to-ground quantum-limited communication using a 50-kg-class microsatellite, Nature Photon. **11**, 502–508 (2017)

34.42 T. Kuwahara, K. Yoshida, Y. Tomioka, K. Fukuda, H. Kunimori, M. Toyoshima, T. Fuse, T. Kubooka, J. Kurihara, Y. Takahashi: Laser data downlink system of micro-satellite RISESAT. In: *27th Ann. AIAA/USU Conf. Small Satell.* (2013), Paper SSC13-II-5

34.43 C. Schmidt, C. Fuchs: *The OSIRIS Program at DLR* (DLR, Oberpfaffenhofen 2018)

34.44 C. Schmidt, M. Brechtelsbauer, F. Rein, C. Fuchs: OSIRIS payload for DLR's BiROS satellite. In: *Proc. Int. Conf. Space Opt. Syst. Appl. (ICSOS)* (2014)

34.45 W. Chen, J. Sun, X. Hou, R. Zhu, P. Hou, Y. Yang, M. Gao, L. Lei, K. Xie, M. Huang, R. Li, H. Zang, Y. Wan, E. Dai, Y. Xi, W. Lu, S. Wei, L. Liu, J. Li: 5.12Gbps optical communication link between LEO satellite and ground station. In: *IEEE Int. Conf. Space Opt. Syst. Appl. (ICSOS)* (2017), https://doi.org/10.1109/ICSOS.2017.8357402

34.46 M. Toyoshima, K. Araki: A study of a ground-to-satellite optical communication link with multi-

ple uplink laser beams. In: *Int. Laser Sens. Symp. (ILSS)* (1999), Paper P2-16
34.47 G.T. Belle, A.B. Meinel, M.P. Meinel: The scaling relationship between telescope cost and aperture size for very large telescopes, Proc. SPIE (2004), https://doi.org/10.1117/12.552181
34.48 P. Martelli, S.M. Pietralunga, D. Nicodemi, M. Martinelli: Accurate sensitivity in optically preamplified direct detection, Opt. Lett. **29**(13), 1473–1475 (2004)
34.49 A.C. Clarke: Extra-terrestrial relays: Can rocket stations give world-wide radio coverage?, Wirel. World **35**(10), 305–308 (1945)
34.50 M. Shimizu, K. Shiratama, Y. Ohgushi, M. Shikatani, Y. Arimoto, T. Aruga: Point-ahead mechanism for ETS-VI optical ISL experiment, Proc. SPIE (1990), https://doi.org/10.1117/12.18224
34.51 K.E. Wilson: An Overview of the GOLD Experiment Between the ETS-VI Satellite and the Table Mountain Facility, TDA Progress Report 42–124 (1995)
34.52 National Reconnaissance Office: *Press-Release – NRO GeoLITE Satellite Successfully Launched* (NRO, Chantilly 2001)
34.53 B. Laurent: SILEX overview on the European optical communications programme, Acta Astronaut. **37**, 417–423 (1995)
34.54 T. Tolker-Nielsen, G. Oppenhauser: In-orbit test result of an operational optical intersatellite link between ARTEMIS and SPOT4, Proc. SPIE (2002), https://doi.org/10.1117/12.464105
34.55 T. Tolker-Nielsen, J.-C. Guillen: SILEX: The first European optical communication terminal in orbit, ESA Bull. **96**, 42–44 (1998)
34.56 A. Alonso, M. Reyes, Z. Sodnik: Performance of satellite-to-ground communications link between ARTEMIS and the Optical Ground Station, Proc. SPIE (2004), https://doi.org/10.1117/12.565516
34.57 Y. Fujiwara, M. Mokuno, T. Jono, T. Yamawaki, K. Arai, M. Toyoshima, H. Kunimori, Z. Sodnik, A. Bird, B. Demelenne: Optical inter-orbit communications engineering test satellite (OICETS), Acta Astronaut. **61**(1–6), 163–175 (2007)
34.58 Y. Takayama, T. Jono, Y. Koyama, N. Kura, K. Shiratama, B. Demelenne, Z. Sodnik, A. Bird, K. Arai: Observation of atmospheric influence on OICETS inter-orbit laser communication demonstrations, Proc. SPIE (2007), https://doi.org/10.1117/12.736789
34.59 A. Löscher: Atmospheric influence on a laser beam observed on the OICETS – ARTEMIS communication demonstration link, Atmos. Meas. Tech. Discuss. **3**(3), 2035–2054 (2010)
34.60 F. Heine, G. Mühlnikel, H. Zech, S. Philipp-May, R. Meyer: The European Data Relay System, high speed laser based data links. In: *7th Adv. Satell. Multimed. Syst. Conf./13th Signal Proc. Space Commun. Worksh. (ASMS/SPSC)* (2014) pp. 284–286
34.61 Airbus: SpaceDataHighway, https://www.airbus.com/space/telecommunications-satellites/space-data-highway.html (2018)
34.62 M.T. Knopp, D. Giggenbach, R.M. Calvo, C. Fuchs, K. Saucke, F. Heine, F. Sellmaier, F. Huber: Connectivity services based on optical ground-to-space links, Acta Astronaut. **148**, 369–375 (2018)
34.63 S. Bobrovskyi, R. Barrios, D. Giggenbach, F. Moll, F. Sellmaier, F. Huber: EFAL – EDRS feeder link from Antarctic latitudes – System architecture and operations concept. In: *SpaceOps 2014 Conf.* (2014), https://doi.org/10.2514/6.2014-1822
34.64 CCSDS: Optical Communications Working Group, https://cwe.ccsds.org/fm/Lists/Charters/DispForm.aspx?ID=79 (2018)
34.65 D.J. Israel, B.L. Edwards, J.W. Staren: Laser communications relay demonstration (LCRD) update and the path towards optical relay operations. In: *IEEE Aerosp. Conf.* (2017), https://doi.org/10.1109/AERO.2017.7943819
34.66 D.J. Israel, B.L. Edwards, J.D. Moores, S. Piazzolla, S. Merritt: The laser communications relay demonstration experiment program. In: *23rd Ka Broadband Commun. Conf.* (2017), available from: http://hdl.handle.net/2060/20170010177
34.67 Y. Chishiki, S. Yamakawa, Y. Takano, Y. Miyamoto, T. Araki, H. Kohata: Overview of optical data relay system in JAXA, Proc. SPIE (2016), https://doi.org/10.1117/12.2217552
34.68 Y. Satoh, Y. Miyamoto, Y. Takano, S. Yamakawa, H. Kohata: Current status of Japanese optical data relay system (JDRS). In: *IEEE Int. Conf. Space Opt. Syst. Appl. (ICSOS)* (2017), https://doi.org/10.1109/ICSOS.2017.8357398
34.69 M. Toyoshima, T. Fuse, A. Carrasco-Casado, D.R. Kolev, H. Takenaka, Y. Munemasa, K. Suzuki, Y. Koyama, T. Kubo-oka, H. Kunimori: Research and development on a hybrid high throughput satellite with an optical feeder link – Study of a link budget analysis. In: *IEEE Int. Conf. Space Opt. Syst. Appl. (ICSOS)* (2017), https://doi.org/10.1109/ICSOS.2017.8357424
34.70 Y. Munemasa, T. Fuse, T. Kubo-oka, H. Kunimori, D.R. Kolev, A. Carrasco-Casado, H. Takenaka, Y. Saito, P.V. Tinh, K. Suzuki, Y. Koyama, M. Toyoshima: Design status of the development for a GEO-to-ground optical feeder link, HICALI, Proc. SPIE (2018), https://doi.org/10.1117/12.2291881
34.71 A. Carrasco-Casado, P. X. Do, D. Kolev, T. Hosonuma, K. Shiratama, H. Kunimori, P. V. Trinh, Y. Abe, S. Nakasuka, M. Toyoshima: Intersatellite-Link Demonstration Mission between CubeSOTA (LEO CubeSat) and ETS9-HICALI (GEO Satellite). In: IEEE Int. Conf. Space Opt. Syst. Appl. (ICSOS) (2019), https://doi.org/10.1117/12.2218275
34.72 E. Lutz: Towards the terabit/s satellite – Interference issues in the user link, Int. J. Satell. Commun. Netw. **34**(4), 461–482 (2015)
34.73 ITU-T G.694.1: Spectral grids for WDM applications – DWDM frequency grid (2012)
34.74 J. Poliak, R. Mata Calvo, F. Rein: Demonstration of 1.72 Tbit/s optical data transmission under worst-case turbulence conditions for ground-to-geostationary satellite communications, IEEE Commun. Lett. **22**(9), 1818–1821 (2018)
34.75 ADVA: DLR and ADVA set new world record for optical free-space data transmission,

https://www.adva.com/en/newsroom/press-releases/20180510-dlr-and-adva-set-new-world-record-for-optical-free-space-data-transmission (2018)
34.76 A. Carrasco-Casado, A. Biswas, R. Fields, B. Grefenstette, F. Harrison, S. Sburlan, M. Toyoshima: Optical communication on CubeSats – Enabling the next era in space science. In: *IEEE Int. Conf. Space Opt. Syst. Appl. (ICSOS)* (2017), https://doi.org/10.1109/ICSOS.2017.8357210
34.77 R.F. Kalman, J.C. Fan, L.G. Kazovsky: Dynamic range of coherent analog fiber-optic links, J. Lightwave Technol. **12**(7), 1263–1277 (1994)
34.78 H. Al-Raweshidy, S. Komaki (Eds.): *Radio Over Fiber Technologies for Mobile Communications Networks* (Artech House, London 2002)
34.79 S. Dimitrov, B. Matuz, G. Liva, R. Barrios, R. Mata-Calvo, D. Giggenbach: Digital modulation and coding for satellite optical feeder links. In: *7th Adv. Satell. Multim. Syst. Conf./13th Signal Proc. Space Commun. Worksh. (ASMS/SPSC)* (2014), https://doi.org/10.1109/ASMS-SPSC.2014.6934537
34.80 S. Poulenard, A. Mège, C. Fuchs, N. Perlot, J. Riedi, J. Perdigues: Digital optical feeder links system for broadband geostationary satellite, Proc. SPIE (2017), https://doi.org/10.1117/12.2255987
34.81 M. Sotom, B. Benazet, A.L. Kernec, M. Maignan: Microwave photonic technologies for flexible satellite telecom payloads. In: *35th Eur. Conf. Opt. Commun. (ECOC)* (2009)
34.82 R. Mata Calvo, P. Becker, D. Giggenbach, F. Moll, M. Schwarzer, M. Hinz, Z. Sodnik: Transmitter diversity verification on ARTEMIS geostationary satellite, Proc. SPIE (2014), https://doi.org/10.1117/12.2036554
34.83 J.W. Hardy: *Adaptive Optics for Astronomical Telescopes* (Oxford Univ. Press, Oxford 1998) p. 43
34.84 G.A. Tyler: Bandwidth considerations for tracking through turbulence, J. Opt. Soc. Am. A **11**(1), 358–367 (1994)
34.85 J.D. Barchers, D.L. Fried, D.J. Link: Evaluation of the performance of Hartmann sensors in strong scintillation, Appl. Opt. **41**(6), 1012–1021 (2002)
34.86 K. Murphy, D. Burke, N. Devaney, C. Dainty: Experimental detection of optical vortices with a Shack-Hartmann wavefront sensor, Opt. Express **18**(15), 15448–15460 (2010)
34.87 M.A. Vorontsov: Decoupled stochastic parallel gradient descent optimization for adaptive optics – Integrated approach for wave-front sensor information fusion, J. Opt. Soc. Am. A **19**(2), 356–368 (2002)
34.88 C.E. Carrizo, R.M. Calvo, A. Belmonte: Intensity-based adaptive optics with sequential optimization for laser communications, Opt. Express **26**(13), 16044–16053 (2018)
34.89 R.R. Parenti, R.J. Sasiela: Laser-guide-star systems for astronomical applications, J. Opt. Soc. Am. A **11**(1), 288–309 (1994)
34.90 L.A. Thompson, C.S. Gardner: Experiments on laser guide stars at Mauna Kea Observatory for adaptive imaging in astronomy, Nature **328**(6127), 229–231 (1987)
34.91 L.A. Thompson, S.W. Teare: Rayleigh laser guide star systems – Application to the University of Illinois seeing improvement system, Publ. Astron. Soc. Pac. **114**, 1029–1042 (2002)
34.92 D.L. Fried: Focus anisoplanatism in the limit of infinitely many artificial-guide-star reference spots, J. Opt. Soc. Am. A **12**(5), 939–949 (1995)
34.93 D. Bonaccini, W.K.P. Hackenberg, M.J. Cullum, E. Brunetto, T. Ott, M. Quattri, E. Allaert, M. Dimmler, M. Tarenghi, A. van Kesteren, C. Dichirico, B. Buzzoni, P. Gray, R. Tamai, M. Tapia: ESO VLT laser guide star facility, Proc. SPIE (2002), https://doi.org/10.1117/12.454802
34.94 R. Ragazzoni: Robust tilt determination from Laser Guide Stars using a combination of different techniques, Astron. Astrophys. **319**, L9–L12 (1997)
34.95 R. Mata Calvo, D. Bonaccini Calia, R. Barrios, M. Centrone, D. Giggenbach, G. Lombardi, P. Becker, I. Zayer: Laser guide stars for optical free-space communications, Proc. SPIE (2017), https://doi.org/10.1117/12.2256666
34.96 J. Romba, Z. Sodnik, M. Reyes, A. Alonso, A. Bird: ESA's bidirectional space-to-ground laser communication experiments, Proc. SPIE (2004), https://doi.org/10.1117/12.560934
34.97 A. Mustafa, D. Giggenbach, J. Poliak, S. ten Brink: Spectrally efficient transmitter diversity scheme for optical satellite feeder links employing multiple signal sidebands. In: *Photonic Netw., 18. ITG Symp.* (2017)
34.98 D.J. Geisler, T.M. Yarnall, C.M. Schieler, M.L. Stevens, B.S. Robinson, S.A. Hamilton: Experimental demonstration of multi-aperture digital coherent combining over a 3.2-km free-space link, Proc. SPIE (2017), https://doi.org/10.1117/12.2256581
34.99 D.V. Hahn, D.M. Brown, N.W. Rolander, J.E. Sluz, R. Venkat: Fiber optic bundle array wide field-of-view optical receiver for free space optical communications, Opt. Lett. **35**(21), 3559–3561 (2010)
34.100 T.A. Birks, I. Gris-Sánchez, S. Yerolatsitis, S.G. Leon-Saval, R.R. Thomson: The photonic lantern, Adv. Opt. Photonics **7**(2), 107–167 (2015)
34.101 Laserlight: https://www.laserlightcomms.com/
34.102 Point Topic: *Broadband Coverage in Europe 2011* (European Commision, Brussels 2012)
34.103 Ericsson: *5G Deployment Considerations* (Ericsson, Stockholm 2018)
34.104 ITU: *The Tactile Internet*, ITU-T Technology Watch Report (ITUT, Geneva 2014)
34.105 I. Parvez, A. Rahmati, I. Guvenc, A.I. Sarwat, H. Dai: A survey on low latency towards 5G: RAN, core network and caching solutions, IEEE Commun. Surv. Tutor. **20**(4), 3098–3130 (2017)
34.106 Verizon: IP latency statistics, http://www.verizonenterprise.com/about/network/latency/ (2018)
34.107 S. Plass, F. Clazzer, B. Fritz, I. Yasrine, M. Maurizio: Maritime communications – Identifying current and future satellite requirements and technolo-

gies. In: *20th Ka Broadband Commun. Navig. Earth Obs. Conf.* (2014), available from: https://elib.dlr.de/90468/

34.108 Eurocontrol: Forecasting, https://www.eurocontrol.int/forecasting (2019)

34.109 M. Schnell, U. Epple, D. Shutin, N. Schneckenburger: LDACS: Future aeronautical communications for air-traffic management, IEEE Commun. Mag. **52**(5), 104–110 (2014)

34.110 C. Chen, E. Ekici: A routing protocol for hierarchical LEO MEO satellite IP networks, Wirel. Netw. **11**(4), 507–521 (2005)

34.111 F. Long: *Satellite Network Robust QoS-Aware Routing* (Springer, Berlin, Heidelberg 2014)

34.112 T. Reid, A. Neish, T. Walter, P. Enge: *Navigation from Low Earth Orbit* (SLAC, Menlo Park 2016)

Alberto Carrasco-Casado

Space Communications Laboratory
NICT – National Institute of Information and Communications Technology
Tokyo, Japan
alberto@nict.go.jp

Alberto Carrasco-Casado received his BS Degree and MS Degree in Telecommunication Engineering in 2005 and 2008, respectively. His received an MS Degree in Space Research in 2009 and a PhD in Electric, Electronic, and Automation Engineering in 2015. In 2015, he joined NICT (National Institute of Information and Communications Technology) in Tokyo, Japan. His research interests focus on satellite laser communications and quantum cryptography.

Ramon Mata-Calvo

Institute of Communications and Navigation
DLR – Deutsches Zentrum für Luft-und Raumfahrt (German Aerospace Center)
Oberpfaffenhofen-Wessling, Germany
ramon.matacalvo@dlr.de

Ramon Mata-Calvo received his double-Degree on Telecommunications Engineering in 2005, from Universitat Politècnica de Catalunya (UPC) and Politecnico di Torino, where in 2009 he received his PhD. in Electronics and Communications Engineering. In 2009, he joined the DLR's Institute of Communications and Navigation. Since 2013 he is leading the group "Advanced Optical Technologies", dealing with satellite optical free-space communications.

35. Visible Light Communications

Xin Lin, Tomokuni Matsumura

Current wireless communications need to fulfill two important requirements according to different applications. The first is to achieve high-speed and long-distance data transmission, and the second is to realize ubiquitous and short-range information services. As a carrier for wireless communications, radio, infrared radiation, and visible light (VL) are complementary transmission media, and different applications call for the use of one medium or the others. Radio and infrared radiation are favored for the first requirements. For the second requirement, however, visible light communication (VLC) is an advancing field, which has received much attention. VLC transmits data by intensity modulation of light sources, such as light emitting diodes (LEDs) and laser diodes (LDs), which are faster than the response of the human eye. VLC merges lighting and data communications in applications such as indoor lighting, signboards, wireless local area networks (W-LAN), streetlights, vehicles, traffic signals, underwater signals, and so on. This chapter describes the methods and application of VLC and covers topics such as light sources, receivers, VLC technologies, and current applications including ubiquitous indoor information services, visible-light wireless LAN (VLW-LAN), and underwater visible-light wireless communication (UVLWC).

35.1 Overview of Visible Light Communication (VLC)

The rapid advance and diffusion of portable information terminals in work and living environments is accelerating the introduction of various wireless communication methods. The desire for ubiquitous and short-range information services, and underwater large-capacity data transmission, has motivated recent interest in visible light communication (VLC) to satisfy these requirements. This chapter aims to introduce visible light (VL) optical wireless communications, including light sources, photodetectors, transmission channel characterization and methods for optical intensity modulation. The chapter, apart from emphasizing the main enabling technologies of VLC, also incorporates recent technical developments such as visible laser diodes (LD), image sensors (IS), visible-light wireless local area networks (VLW-LAN), and adaptive underwater visible-light wireless communications (UVLWC). Figure 35.1 shows the main enabling technologies and application fields related to VLC. The research and development of VLC involves not only optical devices, such as light sources and photodetectors, but also techniques for optical carrier modulation and data transmission. The results of research and development of VLC can be used not only for terrestrial ubiquitous information services

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_35

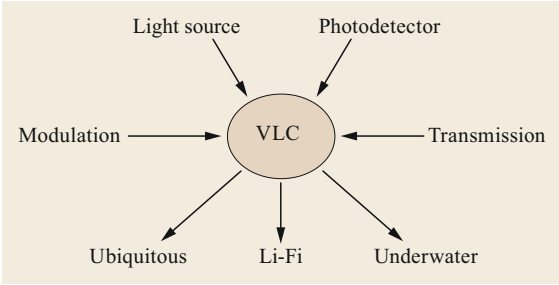


Fig. 35.1 Main enabling technologies and application fields related to VLC

and VLW-LANs, but also for underwater communications.

35.1.1 Historical Development of VLC

It is of interest to precede a description of VLC techniques with a brief overview of the historical development of the field. Some forms of communication using VL beams were in use in ancient times. The use of smoke and reflected sunlight during the day and the use of fire during the night were perfected by the Greeks to convey messages over long distances, although they were rudimentary compared to today's optical communication methods [35.1].

The first successful practical scheme for VLC can be attributed to the great inventor A. G. *Bell* [35.2]. In 1880, Bell invented the photophone, a device in which speech could be transmitted on a beam of light as shown in Fig. 35.2. In this scheme, a narrow beam of sunlight via a plane mirror and a lens was focused onto a flexible mirror. Original sound waves via a speaking tube striking the mirror caused it to vibrate with the intensity of the reflected light varying proportionally to the strength of the sound. A selenium (Se) detector whose resistance varied according to the intensity of incident light then detected the light [35.3]. In this way, Bell managed to send a voice signal over a distance of about 200 m. Though Bell's photophone did not reach commercial fruition due to the lack of a dependable light source and of a low-loss transmission medium, it shows that the medium for information transmissions can use not only long-wavelength radio waves but also short-wavelength VL waves, and it forms an important cornerstone for VLC. In Bell's own words: "I regard the photophone as the greatest invention I have ever made, greater than the telephone" [35.4].

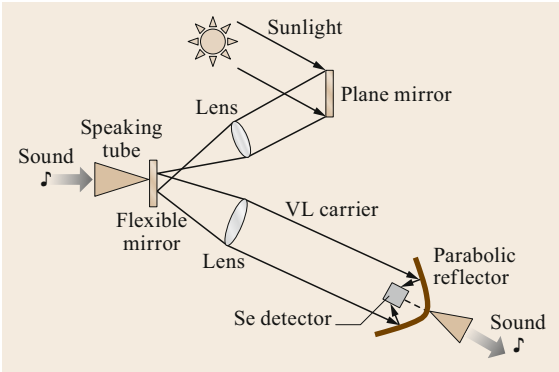


Fig. 35.2 Schematic of the Bell's photophone

Practical VLC had to wait until the arrival of the semiconductor diode source of the light emitting diode (LED). In the 1990s, the success of blue LED chips made white LED lighting possible. VLC reached a new height when LED lighting was applied successfully to indoor data transmission [35.5–8]. In 2011, the IEEE (Institute of Electrical and Electronic Engineers) developed the first standard for the use of VLC for wireless personal area networks—IEEE Std 802.15.7^{TM}-2011—covering topics such as network topologies, addressing, collision avoidance, acknowledgment, performance quality indication, dimming support, visibility support, colored status indication and color-stabilization [35.9]. It has driven VLC to make tremendous progress.

35.1.2 Comparison Between VL and Radio Media

In wireless communication systems using electromagnetic waves as a medium, the transmitter superimposes information on a carrier signal. At the receiver end, this information is retrieved from the carrier and processed. Theoretically, increasing the carrier frequency therefore increases the transmission bandwidth. Figure 35.3 shows the spectrum ranges of electromagnetic waves. Light propagates in the form of electromagnetic waves in free space. The region of optical wavelengths

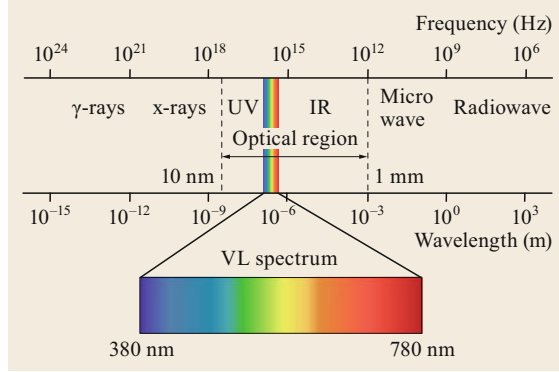


Fig. 35.3 Frequencies and wavelengths of electromagnetic waves

Table 35.1 Comparison between radio and VL wireless links

Property of medium	Radio	VL	Implication for VL
Bandwidth regulated?	Yes	No	– Approval not required – Worldwide compatibility
Passes through walls?	Yes	No	– Short range – More easily secured
Passes through water?	Extremely low frequency only	Yes	– Simpler underwater wireless communication
Main noise?	Other users	Background light	– Difficult to operate outdoors
Input represent	Amplitude	Power	– High power requirement – Simpler intensity modulation

stretches from about 10 nm of ultraviolet (UV) radiation to about 1 mm of far infrared (IR) radiation. The VL portion of this spectrum lies in the 380–780 nm region. The corresponding frequency of the VL region is typically 10^{14} Hz, which can thus support wideband modulation for optical communication.

The use of VL radiation as the medium for wireless communication offers several significant advantages over radio:

- VL emitters are available at low cost because LED lighting can be used as a sending source.
- The VL spectral region offers a virtually unlimited bandwidth that is unregulated worldwide.
- VL radiation is absorbed by dark objects, diffusely reflected by light-colored objects, and directionally reflected by shiny surfaces. It can also penetrate through transparent and translucent objects such as glass, water and so on, but not through walls or other opaque barriers. These inherent characteristics of VL make VLC transmissions easy to secure against casual eavesdropping, and it prevents interference between links operating in different rooms.

However, the VL medium is not without drawbacks compared to radio:

- Because VL cannot penetrate walls, communication from one room to another requires the installation of light access points that are interconnected via a wired backbone.
- In many indoor environments, there exists stronger ambient light noise, arising from sunlight, incandescent lighting, and fluorescent lighting, which induces noise in a VL receiver.
- At present, intensity modulation with direct detection (IM/DD) is the most practical information transmission technique for VLC. IM/DD must employ relatively high transmit power for high signal-to-noise ratio (SNR) of a direct-detection receiver. However, when using LED lighting as a sending light source, too-strong power (brightness) will make people feel physiological discomfort. Often, the transmitter power level may be limited by concerns over power consumption and eye safety.

The properties of radio and VL wireless links are compared in Table 35.1 [35.10]. Radio and VL are complementary transmission media, and different applications favor the use of one medium or the other. Radio is favored for applications in which user mobility must be maximized or where transmission though walls or over long range is required and may be favored when transmitter power consumption must be minimized. For indoor ubiquitous, short-range underwater links, an international compatibility is required, however, and VLC is an attractive possibility.

35.2 Visible Light (VL) Sources

Optical sources for VLC should satisfy some desirable properties in terms of intensity, radiation pattern, emission wavelength, spectral characteristics and response time. Semiconductor diode sources of LED and LD types are a natural choice for VLC systems due to low power consumption, long life, compact size, rapid response (on/off time of devices), and regular spectral characteristics with single-wavelength emission.

An LED is basically an incoherent source. It emits radiation over a wide angle and contains a broad spectrum of wavelengths. On the other hand, an LD is a highly coherent source, emits in a narrow range of angles, and has a narrow spectrum and faster response time than LEDs. However, in terms of operating current requirement, cost and reliability, LEDs scores over LDs.

35.2.1 Visible LED

Basic Principles

The LED is a two-lead semiconductor light source. It is a positive-negative (p-n) junction diode under for-

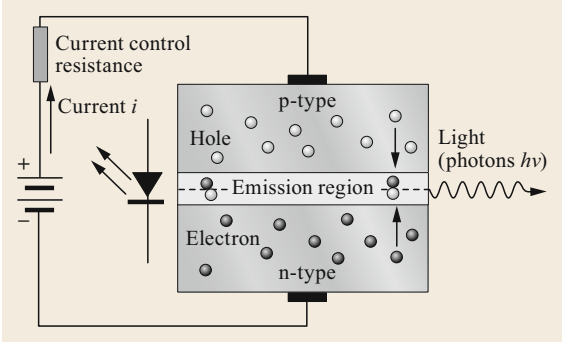


Fig. 35.4 LED with a p-n junction

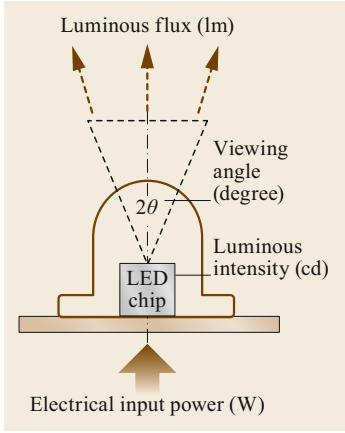


Fig. 35.5 The interrelationship of brightness units for LED

ward bias that emits light when activated, and consists of a chip of semiconducting material doped with impurities to create a p-n junction. When a suitable voltage is applied to the leads, electrons are able to recombine with electron holes within the device, releasing energy in the form of photons. This effect is called electroluminescence. The wavelength of the light emitted, i.e., the color of the light, is determined by the energy bandgap of the materials forming the p-n junction. The materials used for the LED have a direct bandgap with energies corresponding to near-infrared, visible, or near-ultraviolet light. The inner workings of an LED are shown in Fig. 35.4.

LED development began with infrared and red devices made with gallium arsenide (GaAs); their emission intensity is in the milliwatt order. Advances in materials science have enabled devices with ever-shorter wavelengths, emitting visible light in a variety of colors, and now, white light LEDs by mixture of monochromatic blue LEDs and phosphor material have been able to achieve very high brightness, over the order of a watt. Table 35.2 shows some available colors with wavelength range, voltage drop, and material [35.11].

Brightness Units

Brightness units used in LED specifications are candelas/millicandelas (cd/mcd), lumens/millilumens (lm/mlm), and watts/milliwatts (W/mW). Figure 35.5 shows their interrelationship in an LED.

The units cd/mcd represent luminous intensity at a specific forward current and the units are usually expressed as a value relative to the viewing angle. The larger the viewing angle, the more light flows given the same intensity, since light is not always evenly dispersed, and the output of light is determined by the location of the beholder. For example, 1 cd over 120° viewing angle is a lot brighter than 1 cd over 20° viewing angle.

The units lm/mlm are total luminous flux flowing from the LED; it is the power of light as perceived by the human eye with respect to the wavelength of light being emitted. If the luminous intensity (cd) and half viewing angle θ (degrees) are known, first convert θ from degrees to steradian (sr) by using $\text{sr} = 2\pi(1 - \cos\theta)$, and then luminous flux can be given by $\text{lm} = \text{cd} \times \text{sr}$ or $\text{mlm} = \text{mcd} \times \text{sr}$. lm/mlm units usually refer to the total light output of the device such as an LED lamp at the rated current.

The units W/mW represent the electrical power provided to the LED. The wattage of an LED lamp or/and incandescent lamp is how much power consumption that particular lamp will draw, and is not necessarily equal to the power output of the lamp. That is why a smaller wattage rating on an LED lamp but can provide a higher lumen output than incandescent lamps; LEDs save more energy and are also brighter.

Table 35.2 Luminous colors corresponding to some semiconducting luminescent materials

Luminous colors	Wavelength range (nm)	Drive voltage (V)	Luminescent material
Infrared	> 760	< 1.63	GaAs, AlGaAs
Red (R)	610–760	1.63–2.03	AlGaAs, GaAsP, AlGaInP, GaP
Green (G)	500–570	1.9–4.0	GaP, AlGaInP, AlGaP, InGaN, GaN
Blue (B)	450–500	2.48–3.7	ZnSe, InGaN, SiC, Si
Ultraviolet (UV)	< 400	3.0–4.1	InGaN, AlN, AlGaN, AlGaInN
White	Broad spectrum	2.8–4.2	– Color mixing by RGB LED – UV LED + RGB phosphor – Blue LED + yellow phosphor

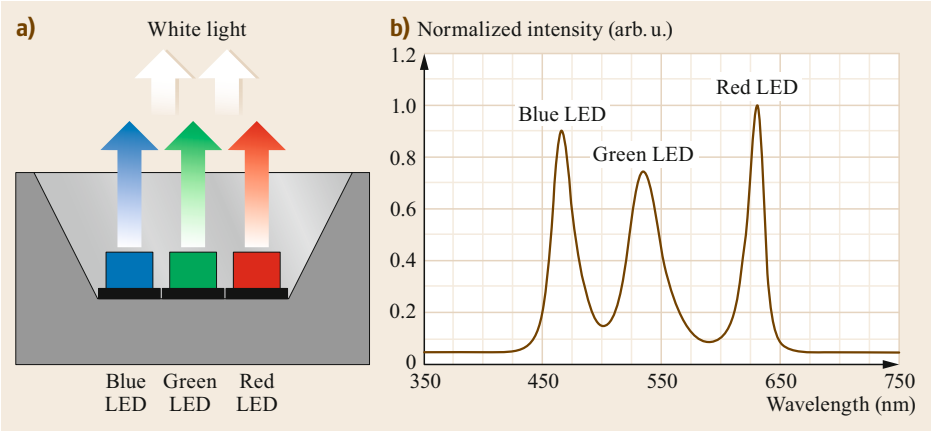


Fig. 35.6a,b Multicolor white LED unit formed by single RGB LED chips: **(a)** formation method and **(b)** combined spectral distribution, FWHM (full width at half maximum) spectral bandwidth is approximately 24–27 nm for each color chip

White LED

In all visible light LEDs, white LED is the most important light source for VLC and illuminations. There are two primary techniques to produce white LEDs, which generate high-intensity white light [35.12]. Note that the *whiteness* of the light produced by these methods is essentially engineered to suit the human eye, and depending on the situation it may not always be appropriate to think of it as *white light*.

The first technique is to use a single LED chip that emits three primary colors of red (R), green (G), and blue (B), and then mix three colors to form white light, as shown in Fig. 35.6a. This type of LED unit is called RGB LED (sometimes also referred to as full-color LED or multicolor LED). Because this method needs electronic circuits to control the blending and diffusion of different colors, and because the individual color LEDs typically have slightly different emission patterns, this leads to variation of the color depending on direction. In addition, each LED chip in this type of LED unit is a narrow band source, that makes gaps in spectrum (Fig. 35.6b), and emission power decays exponentially with rising temperature, resulting in a substantial change in color stability. So RGB LEDs cannot provide good quality white lighting. Nonetheless, this method has many applications because of the flexibility of mixing different colors. For example, for VLC, it is an effective source to implement wavelength division multiplexing (WDM), which is an optical communication technology for transmitting large-capacity signals.

The other technique is to use a phosphor material to convert monochromatic light from a blue or UV LED to form broad-spectrum white light. This type of LED unit is called phosphor-based or phosphor-converted white LED. The luminous efficacy of phosphor-based LED units depends on the spectral distribution of the resultant light output and the original wavelength of the single LED chip that is used to form the white LED unit. The luminous efficacy is a measure of how well a light source produces visible light. For example, the luminous efficacy of a typical yellow phosphor-based white LED unit ranges from three to five times the luminous efficacy of the original single blue LED because of the human eye's greater sensitivity to yellow than to blue. Currently, this method is the most popular method for making high-intensity white LEDs due to the simplicity of manufacturing compared to the method of multicolor LEDs. Figure 35.7 shows a yellow phosphor-based white LED unit; the formation method and the broadened optical spectrum are shown in Fig. 35.7a,b, respectively.

For a yellow phosphor-based white LED, the thickness of the yellow phosphor layer and the wavelength of the blue LED chip influence its color temperature, and different color temperatures lead to different spectral distributions [35.13], as shown in Fig. 35.8. Note that when this white LED is chosen and used as a VLC light source, its color temperature is an important parameter that must be considered. Because only the single blue LED is employed for data transmission in VLC, the phosphor effect makes the speed and quality of the communication attenuate when the color temperature is below about 4000 K. The color temperature is a measure of visible light color that is emitted by a light source, using a unit of measure for absolute temperature in kelvin.

White LEDs can also be made by coating near-ultraviolet LED chips with a mixture of phosphors that emit red, blue, and green light. This method is less efficient than blue LED chips with yellow phosphor, but because several phosphor layers of distinct colors are applied, the emitted spectrum is broadened, effectively raising the color rendering index (CRI) value of a given LED [35.14]. The CRI is most useful as a measure of a light source's color characteristics. In general

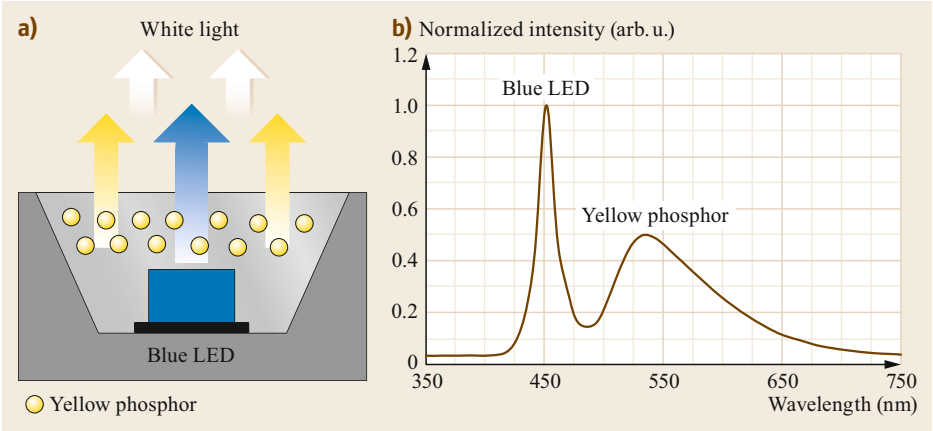


Fig. 35.7a,b Phosphor-based white LED unit formed by blue LED chip and yellow phosphor: **(a)** formation method and **(b)** combined spectral distribution

terms, CRI is a measure of a light source's ability to show object colors *realistically* or *naturally* compared to a familiar reference source, either incandescent light or daylight.

Output and Modulation Characteristics

The output and modulation of LED sources are two primary characteristics for LED-based VLC. For an injected current i of a p-n junction LED, N carriers ($N = i/e$, e is elementary electric charge) are generated. If η is the quantum efficiency, i.e., the fraction of charges that recombine radiatively, and it describes how many electrons are actually generated in a photoelectric process, then ηN photons will be produced. If the bandgap is E then the emitted photons will have an energy $E = h\nu$ (h is Planck's constant and ν is frequency of the photon). So the optical output P_o and the injected current i are linearly related as

$$P_o = \eta NE = \eta \frac{i}{e} E \,. \tag{35.1}$$

However, at higher levels of injected currents, harmonic distortions may occur. It is therefore important to design electronic modulation circuits that suitably linearize the optical output and thus keep the intermodulation distortion to a minimum. When light intensities from an LED are modulated by an electrical signal, the optical output at low modulation frequencies is constant. However, at high frequencies, due to the delay in the recombination process, the output power falls off. Hence the modulation frequency f for the communication system has a limit at which the optical output power $P_o(f)$ is reduced by -3 dB to half of the maximum value $P_{o|max}$. Thus, the modulation response of an LED is described by

$$P_o(f) = \frac{P_{o|max}}{\sqrt{1 + (2\pi f \tau)^2}} \,, \tag{35.2}$$

where τ is the carrier lifetime. In semiconductor physics, the carrier lifetime is defined as the average time it takes for a minority carrier to recombine. For semiconductor photo elements such as LEDs and LDs, the carrier lifetime is used as the time constant of the exponential decay of carriers. It includes radiative and nonradiative lifetimes, and is given by

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}} \,, \tag{35.3}$$

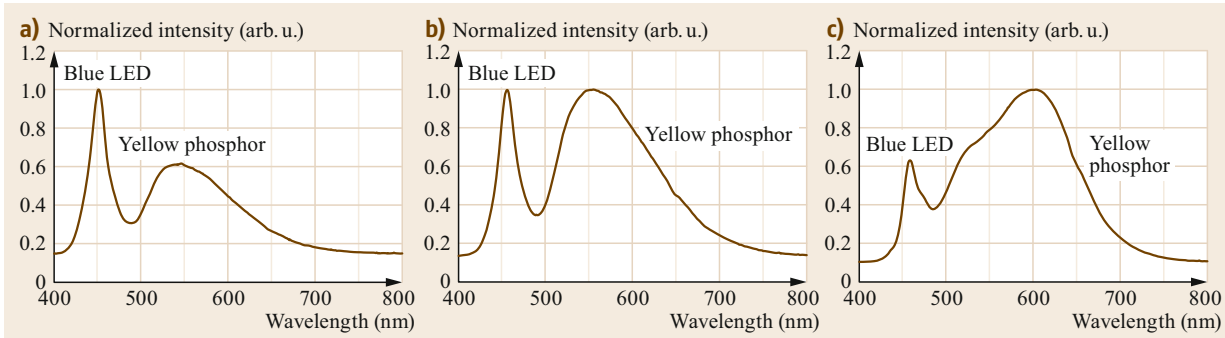


Fig. 35.8a–c Dependence of the color temperature on the spectral distribution in yellow phosphor-based white LED. Color temperature: **(a)** 5700, **(b)** 4000, **(c)** 3000 K

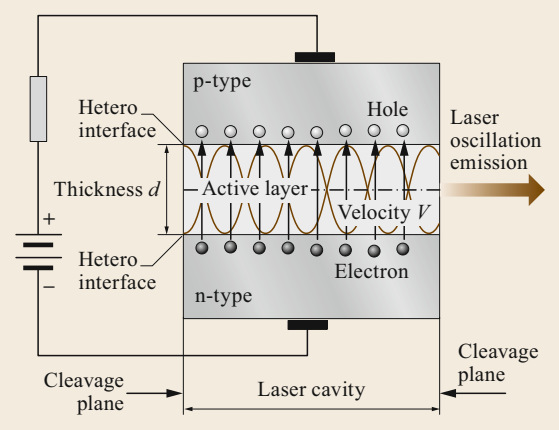


Fig. 35.9 Laser-oscillation principle of LDs

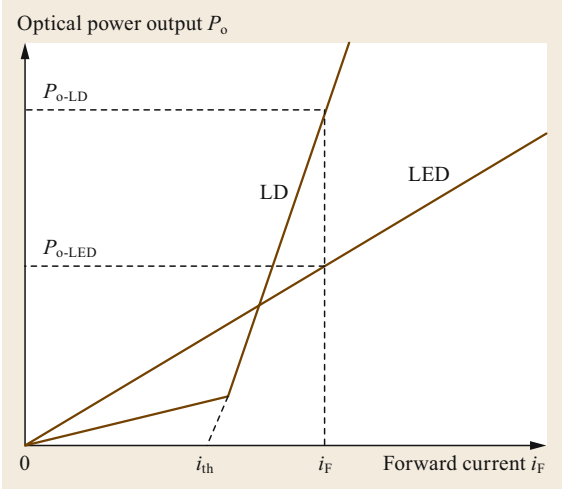


Fig. 35.10 Dependence of optical power output on forward current for LEDs and LDs

where τ_r and τ_{nr} represent the radiative and nonradiative lifetimes, respectively [35.15].

35.2.2 Visible LD

A visible LD can also act as a VLC source, and its active medium is formed by a p-n junction of a semiconductor diode similar to LEDs. The semiconductor material of LDs is also the same as that of LEDs, and the emission wavelength has also developed from the infrared to the red region, and then to the blue region. The difference between LDs and LEDs is that LDs make LEDs have a laser-oscillate structure i.e., a laser cavity, as shown in Fig. 35.9. In the LD structure, shown in Fig. 35.9, to satisfy the laser-oscillation condition both ends of the cavity are processed and become a cleavage plane. The active layer, which is used for light amplification, has a structure capable of totally reflecting light.

The optical power output against the forward current for LEDs and the LDs is plotted in Fig. 35.10. From this figure, one can infer that an LD is a threshold device and the linear part of the response curve is small; an LED on the other hand has good linear response and is best suited for analog modulation. An LD also has a modulation frequency characteristic that can be described by (35.2). In this case τ is given by

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}} + \frac{2V}{d}\,, \tag{35.4}$$

where V is the recombination velocity and d is the active layer thickness [35.15]. The reorganization is a process in a semiconductor when electrons and holes disappear together and release energy.

Viewed against the incoherent and wide spectrum in spatial emission of LEDs, LDs offer far superior performance. Hence, for long-distance and high-bit-rate systems, LDs are the obvious choice. However, in terms of cost, eye safety, complexity in drive circuit, reliability, temperature dependence and drive current requirements, LDs compare poorly with LEDs. Table 35.3 presents a comparison between LEDs and LDs [35.16]. LDs are much more expensive than LEDs, but offer many nearly ideal characteristics for communications:

Table 35.3 Comparison between LEDs and LDs

Characteristics	LED	LD
Spectral width	25–100 nm	$< 10^{-5}$–5 nm
Modulation bandwidth	Tens of kilohertz–tens of megahertz	Tens of kilohertz–tens of gigahertz
E/O conversion efficiency	10–20%	30–70%
Color gamut	100%	130–140%
Polarized light use	No	Yes
Directivity	180°	0°
Eye safety	Generally considered eye-safe	Must be rendered eye-safe, especially for visible LD
Cost	Low	Moderate to high

1. Very high electro-optic conversion (E/O) efficiencies of 30–70%
2. Wide modulation bandwidths, which range from hundreds of MHz to more than 10 GHz
3. Very narrow spectral widths (spectral widths ranging from several nm to well below 1 nm are available).

To achieve eye safety with an LD requires that the laser output is passed through some element that destroys its spatial coherence and spreads the radiation over a sufficiently extended emission aperture and emission angle.

As laser lighting by visible LDs has made great progress in recent years, this may be an important opportunity for VLC.

35.3 VL Detectors

For current VLC, even if light is used as a carrier wave, *all-optical* processing cannot be achieved (in the future, advances in optical materials or optical elements may enable an all-optical communication system). Light is only used to transmit signals. For signal detection and processing, optic-electro conversion (O/E) is still indispensable. The photodiode (PD) is used for this purpose. PDs convert the information transmitted via optical carriers to corresponding electrical signals, which are then processed further to perform the desired function. To transform optical power into electrical power, i.e., to detect the optical signal, a variety of physical effects can be utilized. The optimum choice is the use of the solid-state PD, because it can not only operate at low voltages and offer high speed and long lifetime, but also because it can be integrated to form an optoelectronic integrated circuit thus realizing a monolithic optical receiver of a VLC system.

Currently, two types of silicon PD are widely available for VLC: positive-intrinsic-negative (p-i-n) PD (PIN-PD) and avalanche photodiode (APD). Additionally, a PD array (also called an image sensor, IS) is often used for a VLC system requiring low speed, outdoor application, and multiple source reception.

35.3.1 Positive-intrinsic-negative Photodiode (PIN-PD)

A silicon PIN-PD is the ideal solid-state photodetector and is employed in most receivers of VLC systems at present due to its fast response, high quantum efficiency and capability to handle high electric fields and modulation frequencies.

The structure and principle of a PIN-PD is shown in Fig. 35.11. In the p-i-n structure, the intrinsic i region is sandwiched between the p and the n layers. When the incident photon has energy greater than or equal to the bandgap energy E_g of the semiconductor material, electron–hole pairs are generated. In a well-designed PIN-PDs, the photocarrier generation process occurs mainly in the depletion region of the p-i-n junction where the incident light is largely absorbed. As a result of the high electric field present in this region, the electrons and holes separate and drift in opposite directions. When the carriers drift through the high field region, a photogenerated current I_p is induced in the load, developing an output voltage across the load resistor R_L. The thickness of the i region can be tailored so that most of the light absorption takes place within it, thus minimizing the diffusion component of the current. Also, to obtain fast response time, the capacitance of the junction can be reduced due to the large i region [35.17].

35.3.2 Avalanche Photodiode (APD)

An APD is essentially a p-i-n device that is operated at very high reverse bias voltage V_R, so that photogenerated carriers create secondary carriers by impact ionization, resulting in internal electrical gain [35.17]. The structure of an APD for VL (i.e., short wavelength) is shown in Fig. 35.12. The photoabsorption region i(p^-) is sandwiched the p^+ and the avalanche regions, which consists of a p and a p^+ layers. (p^+ and n^+ denote the heavily doped, and p^- denote the lightly doped.) Contrary to the near-infrared APD, photon $h\nu$ is structured to enter from the p^+ layer.

By controlling the reverse bias voltage V_R, an arbitrary current multiplication factor M can be obtained. However, it is important to note that since the reverse breakdown voltage V_B of the APD has a large temperature dependency, the multiplication factor also strongly depends on the temperature. The current multiplication factor is given by Miller's empirical formula

$$M = \frac{1}{1-(V_R/V_B)^m}\,, \tag{35.5}$$

where m is a constant determined by the APD, and $V_R < V_B$. Normally, the operation region of an APD is $M = 30–100$ [35.18].

APDs are favored in optical receivers with direct detection when there is little ambient-induced shot noise (shot noise is the inherent electronic noise that occurs in photoelectric devices, originating from the discrete

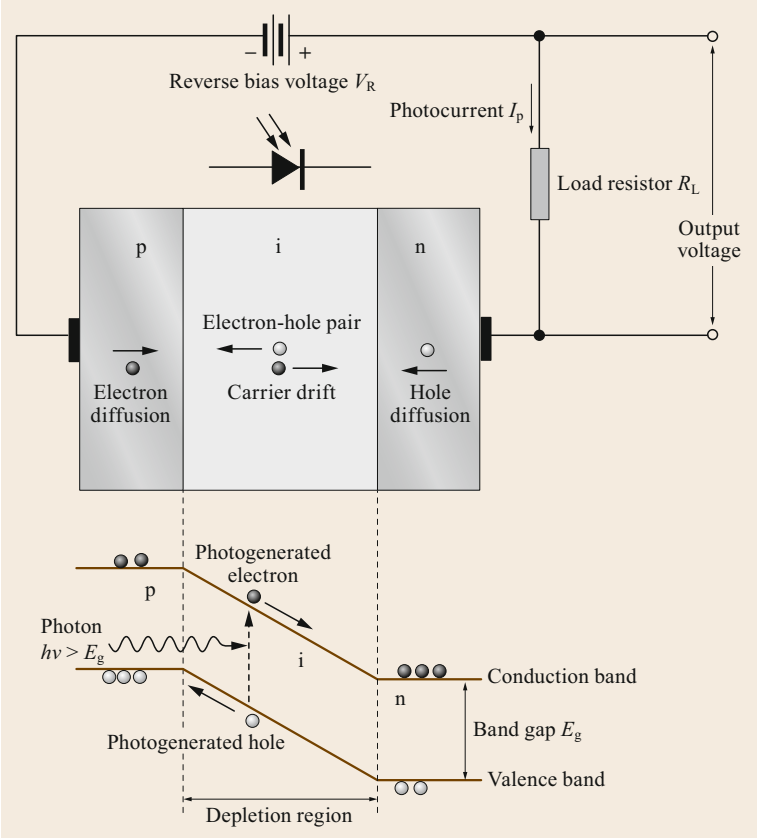


Fig. 35.11 Structure and principle of a PIN-PD

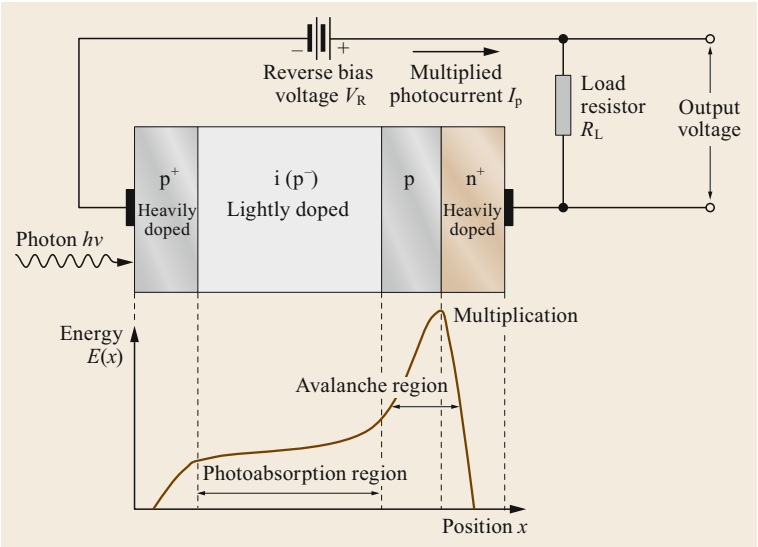


Fig. 35.12 Structure and principle of an APD

nature of electric charge and associated with the particle nature of light), because its internal gain helps overcome the thermal noise from its electronic circuit, increasing the receiver SNR. APD-based receivers can lead to good VL link performance when ambient light is weak. When ambient-induced shot noise is dominant, however, use of an APD results in a net decrease in SNR, because the random nature of an APD's internal

Table 35.4 Comparisons of characteristics for IS, PIN-PD and APD

	IS	PIN-PD	APD
Primary advantages	Parallel effect (spatial separation)	– Low cost – Low bias requirement – Fast response	Current multiplication effect
Primary drawbacks	Low speed	Mixed effect for multiple sources (spatial interference)	– High cost – High bias requirement
Implication for applications	For outdoor and multisource VLC systems	For general purpose VLC systems	For weak-ambient-light VLC systems

gain increases the variance of the shot noise by a factor greater than the signal gain.

For a receiver of VLC, despite the higher dark current and the excess noise due to multiplication problems in APDs, in general they can provide 3 to 5 dB higher sensitivity than PIN-PDs above 100 Mbps. In the case of lower bit rates, low leakage currents of PIN-PDs result in better sensitivity. Additional drawbacks of APDs include their high cost, requirement for high bias, and their temperature-dependent gain. Therefore PIN-PDs are preferred in most VLC systems at present.

35.3.3 Image Sensor (IS)

An IS is essentially a PD-array device that consists of a PD array chip for light detection and an integrated circuit (IC) chip for signal processing [35.19], as shown in Fig. 35.13.

In an IS-based VLC system, the transmitted data are received by extracting the luminance corresponding to the VLC transmitter from the captured image. Its receiver can spatially separate multiple or noise sources due to the IS's parallel effect of its array structure, simultaneously demodulate multiple sources, remove noise sources such as sunlight, and detect VL signals even in daytime or outdoor environments. For example, as shown in Fig. 35.13, single PD/APD mixes and receives light from multiple sources. That is, it is affected by noises from background light or other signal sources. However, in the IS-based system, multiple sources can

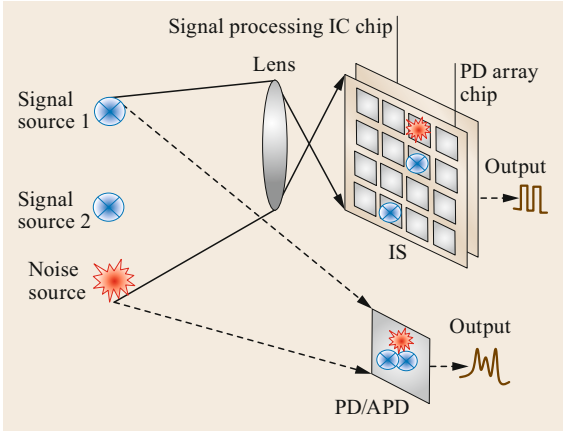


Fig. 35.13 Comparison between PD and IS for removing noise sources

be simultaneously received by passing through a lens, and parallel communication can be performed when multiple sources are separated by the signal processing IC chip [35.20].

In addition, in contrast to the PD-based VLC system, the IS-based system does not require precise adjustment of the optical axis between the transmitter and the receiver. However, since the IS needs to process images from multiple sources in parallel, image processing speed is slower than for the PD. Comparisons of characteristics for IS, PIN-PD and APD are given in Table 35.4.

35.4 VLC Techniques

As described in Sect. 35.1.2, the baseband IM/DD method is the most practical technique for VLC at present. A physical model of a baseband IM/DD-based VLC method is as shown in Fig. 35.14. It consists of an IM sending source, a spatial transmission channel that includes noise from the ambient environment, O/E conversion and the electrical circuit, and a DD module structured around a photodetector and signal extracting/processing circuits.

The purpose of any communication system is to accurately extract the signals sent by the transmitter. Theoretically, the greater the carrier power, the higher the accuracy of the received signal. However, a good communication system should also have low power

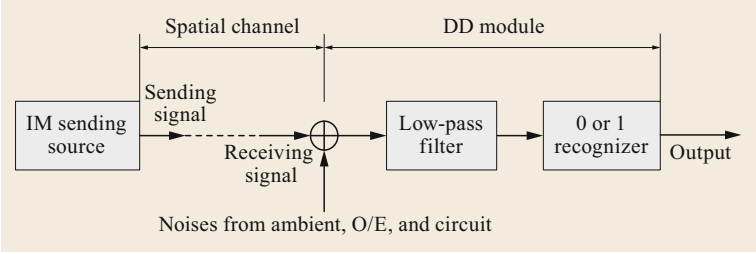


Fig. 35.14 Physical model of baseband IM/DD-based VLC method

consumption, that is, having as minimal carrier power as possible to transmit data. The key factors affecting the performance of these factors in the IM/DD-based VLC system are optical modulation methods, spatial channels and transmission quality.

35.4.1 Optical Modulation Methods

Baseband Intensity Modulation (BIM)

On-off keying (OOK) and pulse-position modulation (PPM) are the most common schemes of BIM [35.10]. The three most important criteria for evaluating BIM techniques are [35.21]:

- The average received optical power required to achieve a desired bit error rate (BER)
- The bandwidth requirement of the receiver
- Flicker efficiency (i.e., reduction of flicker caused by optical modulation).

OOK using no-return-to-zero (NRZ) pulse (OOK-NRZ) is the simplest scheme among all BIM techniques suitable for VLC systems. PPM with L levels (LPPM) is also widely used in illumination-light-based VLC systems. Their coding method and occupied bandwidth are shown in Fig. 35.15a,b, respectively, and Table 35.5 compares their modulation characteristics in terms of bandwidth, power, and flicker efficiency [35.22]. Flicker is an optical phenomenon where the light source repeats light and dark phases moderately enough to be perceived by the human eyes, and it happens when the light source is modulated at low speed by light intensity.

OOK-NRZ has good bandwidth efficiency. From Table 35.5, for a given bit rate R_b, LPPM requires more than OOK-NRZ by a factor $L/\log_2 L$, e.g., 4-PPM requires two times more bandwidth than OOK-NRZ (also see Fig. 35.15b). Therefore, the OOK-NRZ scheme has a higher data transmission rate, and it is almost bit rate-independent for data rates in excess of 100 Mbps.

LPPM is a preferred modulation scheme for devices where low-power consumption is required [35.23], such as portable and underwater optical transmitters due to its high power efficiency. Low flicker [35.24] makes it suitable to be employed in indoor lighting-based VLC

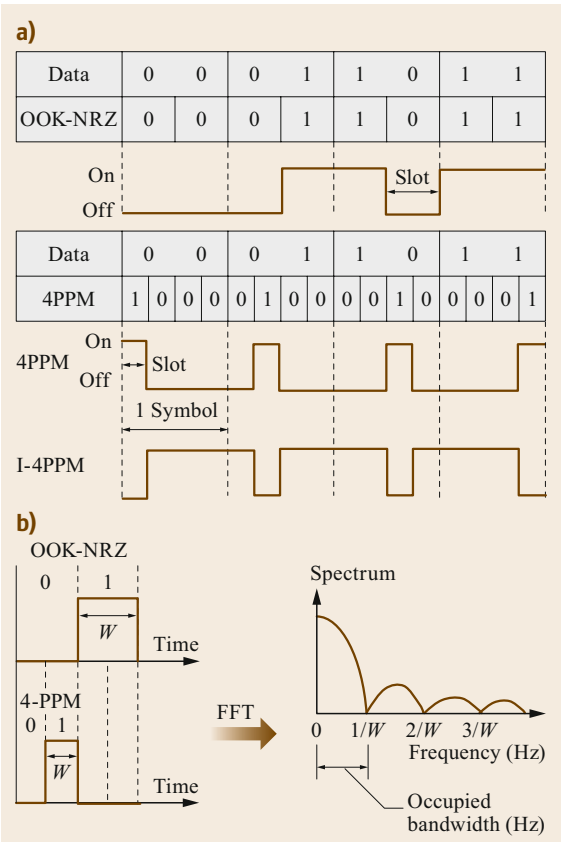


Fig. 35.15a,b Coding methods of OOK-NRZ and LPPM/I-LPPM: **(a)** waveforms and **(b)** occupied bandwidth. FFT: Fast Fourier transform

systems. LPPM yields an average power requirement that decreases steadily with increasing L. When L is larger than 2, the LPPM exhibits a higher efficiency than OOK-NRZ does. However, an increase in L (i.e., in power efficiency) causes an increase in the bandwidth requirement.

In order to reduce the amount of light loss due to light intensity modulation, the inverted-LPPM (I-LPPM) waveform at the bottom of Fig. 35.15a can be used. I-LPPM is logical inverted LPPM, such that each symbol has three-quarters of the amount of all light and

Table 35.5 Comparisons of modulation characteristics for OOK-NRZ and *L*-PPM

	Bandwidth requirement	Power requirement	Influence of flicker
OOK-NRZ	$R_b/2$ (R_b is bit rate)	P_a (P_a is average received power)	Large
*L*PPM	$(R_b/2)(L/\log_2 L)$	$P_a/\sqrt{(L/2)(\log_2 L)}$	Small

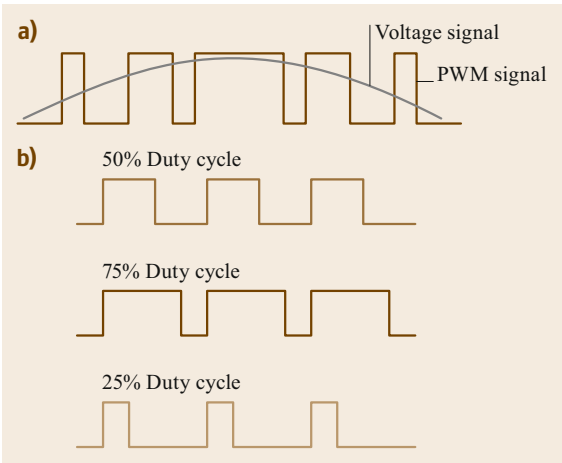


Fig. 35.16a,b Method of PWM: **(a)** waveforms and **(b)** examples of duty cycle

*L*PPM only has one quarter. The consideration to ensure as much of the light as possible is invariant (i.e., light illumination is invariant) is also important for a lighting-based VLC system. Hence, I-*L*PPM is approved by JEITA (Japan Electronics and Information Technology Industries Association) as a standardized BIM technique for VL identification (ID) systems [35.25] and VL beacon systems [35.26].

Pulse-width modulation (PWM) is also an IM scheme, which modulates light intensity by changing the duty cycle of the pulse wave, as shown in Fig. 35.16a. Duty cycle is measured in percentage. The percentage duty cycle specifically describes the percentage of time a digital signal is on over an interval or period of time. If a digital signal spends half of the time on and the other half off, this digital signal has a duty cycle of 50% and resembles an ideal square wave. If the percentage is higher than 50%, the digital signal spends more time in the high state than the low state and vice versa if the duty cycle is less than 50%, as shown in Fig. 35.16b.

The PWM method can be used to control brightness (i.e., dimming) of LEDs by adjusting the duty cycle. When an LED lamp is used as the sending source in a VLC system, illumination dimming may be required. In this case, how to incorporate dimming while not corrupting the communication link is an important subject. *Ntogari* et al. proposed a method of combining PWM and discrete multitone (DMT), which can ensure dimming does not influence the data transmission [35.27].

Besides that, the optical duobinary modulation (ODM) signal [35.28–30] has three intensity levels: 0, 1, and 2. Level 0 and 2 produce 100% transmission with opposite optical phases, and level 1 produces 0% transmission. ODM has a narrow spectral width and large dispersion tolerance due to multilevel encoding. Hence, it is a better scheme to transmit high-speed optical signals over bandwidth-limited channels. However, also due to the multilevel encoding, the identification range of the signal becomes narrower, with the result that the BER is increased.

In general, ODM is a better choice for uncompensated single-mode fiber communications, since it is more resilient to chromatic dispersion. The ODM can also be suggested and used for WDM-based VLC systems for which long distance or/and high speed are required.

Other Methods

Color-shift keying (CSK) is a new VLC modulation scheme approved as part of IEEE 802.15.7-VLC in 2011 [35.9]. The CSK signal is generated by three-color VL sources, and transmits data by using the color coordinates, as shown in Fig. 35.17. Figure 35.17 shows the color codes of CSK on a $x-y$ chromaticity diagram. The chromaticity diagram is specified by CIE1931 (Commission International de l'Eclairage). The color of light is represented by the plane coordinates of x, y. The light wavelength is written around the curve (i.e., spectral locus) of the chromaticity diagram. Each monochromatic is represented by these wavelengths, with wavelengths shown in nanometers. The horizontal axis x and the vertical axis y are values of chromaticity. The circle mark on the spectral locus indicates a code of data, for example, the monochromatic wavelength of 510 nm corresponds to code 001.

In the CSK scheme, the communication rate is determined by the number of constellations on the chromaticity diagram, and the speed increases with increasing constellations. It is suitable for video-based VLC systems that have image-processing or color-control requirements [35.31].

Orthogonal frequency division multiplexing (OFDM) is a digital multicarrier parallel modulation scheme, which extends the concept of single subcarrier modulation by using multiple subcarriers within the same single channel to realize large-capacity data transmission. In OFDM, data are transmitted in parallel on

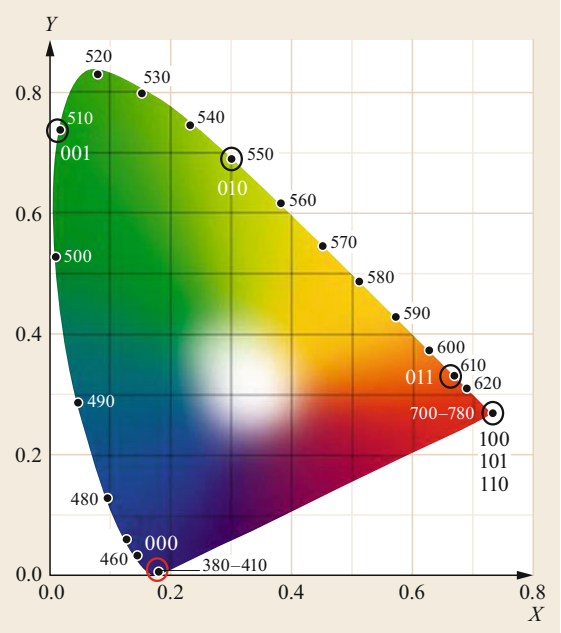


Fig. 35.17 Color codes of CSK on $x-y$ chromaticity diagram

a number of different frequencies, and each frequency channel is separated from the others by a frequency guard band to reduce intersymbol interference between adjacent channels. Currently, OFDM is used in most broadband wired and wireless communication systems. However, in these radio frequency transmission systems, the high peak-to-average power ratio (PAPR) in OFDM is usually considered a disadvantage due to nonlinearities of the power amplifier.

Recently, a number of studies have shown that OFDM is also a promising technology for optical wireless communications [35.32, 33]. *Afgani* et al. demonstrated theoretically that the high PAPR in OFDM can be exploited constructively in an LED-based VLC system to intensity modulate the light from LEDs to realize high-speed VL wireless links [35.34]. However, the nonlinear behavior of LEDs in the transmitter is a limiting factor of system performance. Optical modulation signals with large PAPR suffer uneven distortion. To achieve good performance in optical systems OFDM must be adapted in various ways.

DMT is a variant of OFDM. Its basic idea is also to split the available bandwidth into a large number of subchannels. The study of *Qian* et al. shows that DMT can be well suited to nonlinear LED-based VLC systems [35.35]. DMT has good system performance since it is an orthogonal linear transformations scheme; it can spread the nonlinear effects evenly to each data symbol. Also, DMT can allocate data so that the throughput

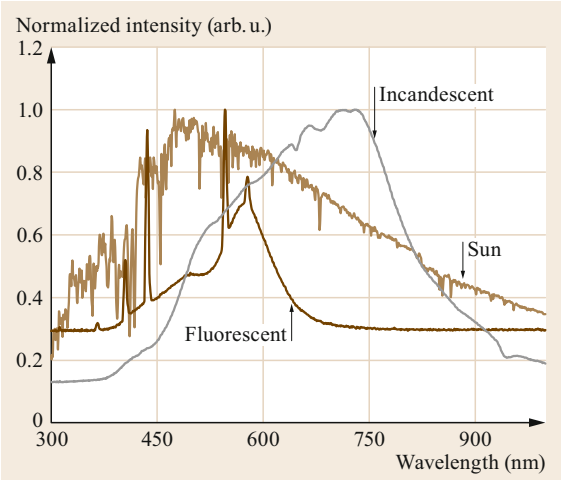


Fig. 35.18 Light intensity spectra of some VL sources

of every single subchannel is maximized. If some subchannel cannot carry any data, it can be turned off and the use of available bandwidth is optimized. *Vucic* et al. have realized a high-speed LED-based VLC of 513 Mbps using DMT modulation [35.36].

35.4.2 Spatial Channel

The spatial channel of an IM/DD-based VLC system consists of a communication environment and a receiving electronic circuit, as shown in Fig. 35.14. Hence, the noise output P_N of the spatial channel is a Gaussian noise pattern, having a total variance that is the sum of contributions from shot and thermal noises

$$P_N = \sigma^2_{total} = \sigma^2_{shot} + \sigma^2_{thermal} \,. \tag{35.6}$$

The shot noise is caused by *light*, and is the sum of ambient light and internal O/E of the optical detector

$$\sigma^2_{shot} = 2ei_a(\lambda)B + 2eRP_SB \,. \tag{35.7}$$

In the first term (i.e., the part of ambient light) of (35.7), e is electric charge, B is the bandwidth of the detection circuit, and i_a is the current due to ambient light. Many environments contain intense ambient VL radiation arising from sunlight, incandescent and fluorescent lamps, and other sources. The optical intensity spectra of some VL sources are shown in Fig. 35.18; i_a is the function of the light-source wavelength, and its value is obtained by numerical integration at a range of the light-source wavelength. These VL from the ambient environment can also be received at an average power much larger than the signal light, even when optical filtering is employed. The resulting direct current (DC) photocurrent causes shot noise.

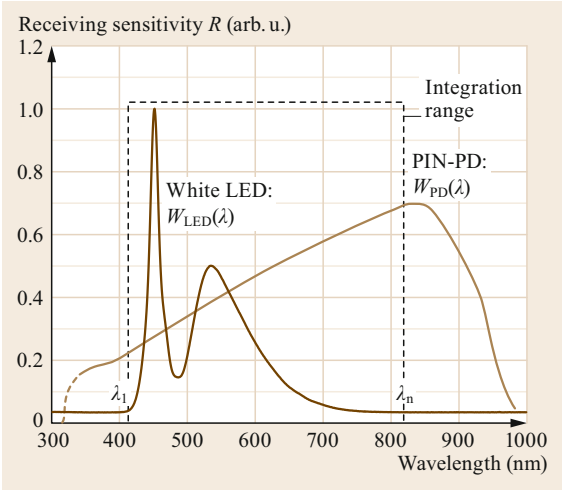


Fig. 35.19 Example of receiving sensitivity of a white-LED-based system

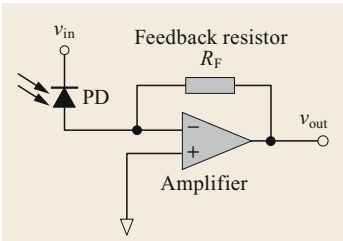


Fig. 35.20 The transimpedance amplifier circuit for a VLC receiver

In the second term (i.e., the part of O/E) of (35.7), P_S is received signal-light power, and R is total sensitivity of the system. When using a white LED as the sending source, R can be given by an integration of the wavelength distribution of the white LED W_{LED} and the PD W_{PD}

$$R = \frac{1}{n}\int_{\lambda_1}^{\lambda_n} W_{LED}(\lambda) W_{PD}(\lambda) d\lambda \,, \tag{35.8}$$

where n is the sampling number of the wavelength. The integration range λ_1 to λ_n is generally 380–780 nm, as shown in Fig. 35.19.

The thermal noise (or Johnson noise) is caused by external electronic circuits of the optical detector, i.e., from *heat*. The transimpedance amplifier circuit is popularly used in the receiver of VLC system, as shown in Fig. 35.20 (v_{in} and v_{out} indicate the input and output voltage). This approach can control the receiving sensitivity and bandwidth by employing a feedback resistor R_F. The thermal noise with R_F is given by

$$\sigma^2_{thermal} = \frac{4kTB}{R_F} \,. \tag{35.9}$$

Here, k is Boltzmann's constant and T is absolute temperature of the ambient environment. Because of the thermal noise of the feedback resistor the receiver noise level is higher and hence the sensitivity is degraded. However, sensitivity can be improved by increasing the value of R_F. On the other hand, in order to extend the bandwidth R_F has to be low.

In addition, for indoor VLC, multipath effects caused by reflected light from walls or plural lamp sources is possible. Figure 35.21a shows the reflected-light multipath effect. LED lamps are generally installed on the ceiling and the illumination light has larger diffusion range. In this case, the receiver receives not only direct light from the LED lamp, but also the reflected light from walls, which have propagation delay time. The influence of the direct light depends on performance of the VLC system, and the reflected light depends on the reflectance of walls and the length of the reflection path. Despite this, some studies have shown [35.21, 37] that the influence of the direct light is much larger than reflected light. Therefore, the influence of reflected light can generally be ignored.

Fig. 35.21b shows the plural-sources multipath effect. In this case, the appropriate configuration of plural LED lamps can avoid or mitigate the interference from

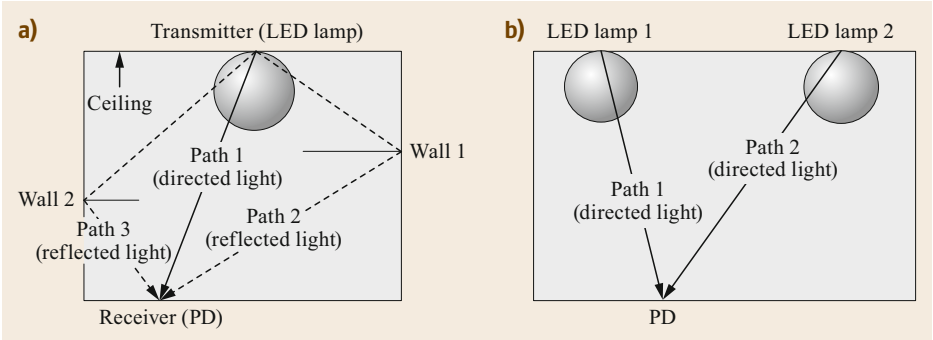


Fig. 35.21a,b The multipath effect is caused by: **(a)** reflective light from walls and **(b)** plural lamp sources

multipath sources. For example, by leaving enough distance between adjacent lamps, one receiver can only receive signals from one lamp. On the other hand, the inherent robustness of OFDM against multipath effects can make it a good choice of modulation method in this case [35.38].

35.4.3 Transmission Quality

The transmission quality and capacity of the system depends on the SNR at its receiver. The SNR is determined by the modulation scheme and characteristics of the spatial channel. In a VLC system, the SNR is the ratio of the received average power P_a (watt: W) with the receiving sensitivity of receiver R (ampere/watt: A/W) to the total noise output P_N (A^2); the receiver's sensitivity directly effects the theoretical received power. The SNR is expressed as

$$\mathrm{SNR} = \frac{(RP_a)^2}{P_N} . \tag{35.10}$$

In addition, the BER, the most important value for evaluating digital transmission quality, can be calculated by the SNR. When the light is modulated by using the OOK-NRZ and LPPM schemes, the BERs are given by (35.11) and (35.12), respectively [35.21, 39]

$$\mathrm{BER}_{\text{OOK-NRZ}} = Q\left(\sqrt{\mathrm{SNR}}\right) \tag{35.11}$$

and

$$\mathrm{BER}_{L\text{PPM}} = \frac{L}{2} Q\left(\sqrt{\frac{L\log_2 L}{2}}\sqrt{\mathrm{SNR}}\right) , \tag{35.12}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{-y^2/2}\mathrm{d}y \quad (x \geq 0) . \tag{35.13}$$

When $L = 2$ in (35.12), $\mathrm{BER}_{\text{OOK-NRZ}} = \mathrm{BER}_{L\text{PPM}}$, that is, 2PPM is equivalent to OOK-NRZ. This can also be derived from Table 35.5. Power requirements can also readily be derived from the BER expressions. From Table 35.5, LPPM requires a factor of $\sqrt{(L/2)(\log_2 L)}$ less power than OOK-NRZ to obtain a particular BER performance.

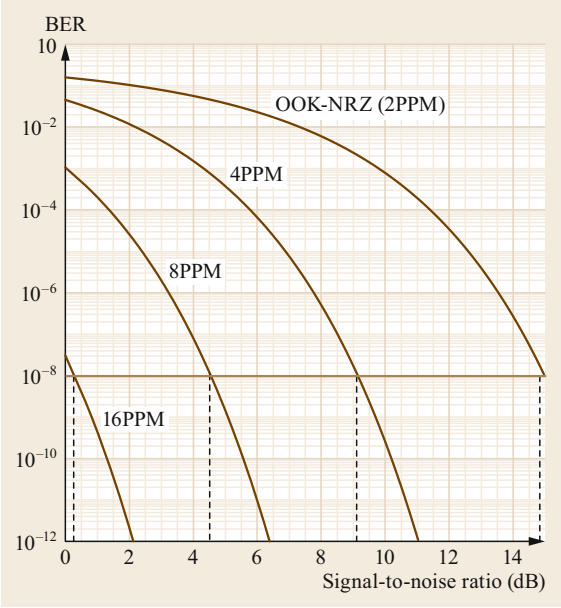


Fig. 35.22 BER versus SNR for OOK-NRZ and LPPM

Figure 35.22 shows the simulation results of BER versus SNR for OOK-NRZ and LPPM ($L = 4, 8, 16$). For LPPM, the SNR requirement to reach a certain BER (e.g., 10^{-8}) is decreased with increasing L, and has higher data transmission efficiency than the OOK-NRZ. However, note that an increase in the L causes a decrease in the communication speed.

The data transmission capacity C of a VLC system, i.e., the maximum transmission rate (bit/s: bps) with a maximum permissible BER can be defined by Shannon's information theory, and is given by

$$C = B\log_2(1 + \mathrm{SNR}) . \tag{35.14}$$

Shannon's information theory means that for a wireless communication system, if an appropriate modulation method (with good B and SNR) is used, then data can be transmitted using maximum communication capacity of this system.

35.5 Current Applications

At present, most VLC systems use LEDs as the light source of the transmitter because of the low cost and eye safety compared to VL LDs. Hence, the current mainstream VLC is LED-based illumination-light communication (ILC). ILC can not only be used for indoor information services, such as location-based services (LBSs) [35.40, 41], audio information guides [35.42], VLW-LAN [35.21, 43], etc., but also can be used for

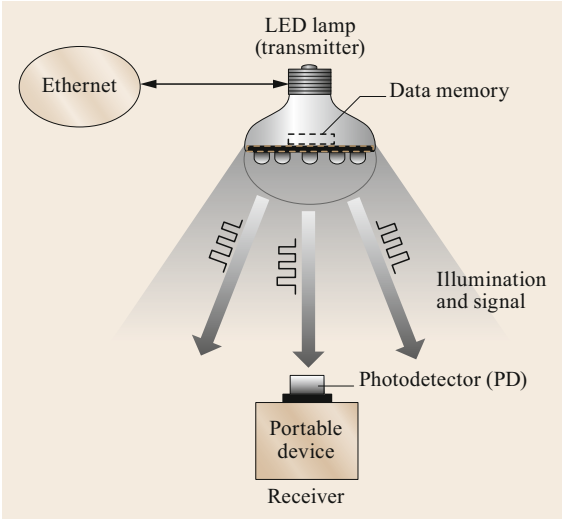


Fig. 35.23 Indoor one-to-one VLC usage model

data transmission outdoors, such as among signboards, streetlights, vehicles, traffic signals, and so on [35.44, 45]. Particularly, the IS receiving is the preferred technique employed in outdoor environments due to its effect in suppressing noise. In short, VLC is a ubiquitous communication technology. In addition, VL wave with a wavelength of 400–650 nm is an ideal underwater carrier, which has received much attention recently [35.46, 47]. Its attractive features are low attenuation rate underwater compared to radio waves, and high data rate compared to acoustic waves.

In this section, as the VLC typical usage models of current applications, the indoor one-to-one system, the VLW-LAN system, and UVLWC will be discussed.

35.5.1 Indoor VLC

Fig. 35.23 illustrates a general scheme of an indoor VLC one-to-one link. The transmitter is an LED-lamp-based fixed device. The receiver is a portable device with a photodetector, such as a tablet terminator, a mobile phone and so on. The LED lamp (i.e., the transmitter) establishes an illumination-light link to the portable receiver. The contents for data transmission can be taken by accessing the resources on an Ethernet or storage in a built-in data memory. Typical applications include LBS, ID recognition systems, and audio information guides.

Figure 35.24 is an example of one-to-one link for a LED-based acousto-optic wireless communication system. The transmitter is a LED lamp using the yellow phosphor-based white LED. A light modulation and control circuit is used to transmit audio information. It is a baseband IM/DD-based system with

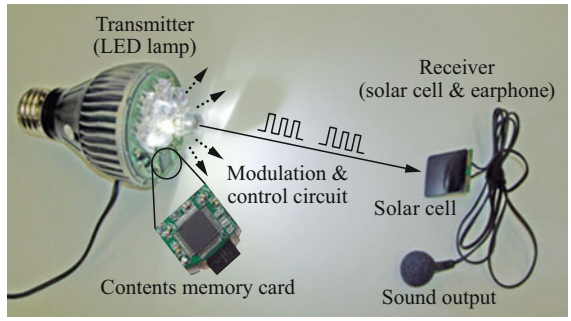


Fig. 35.24 LED-based acousto-optic wireless communication system

I-4PPM scheme. The sound contents are stored in a micro memory card. The receiver is a compact batteryless acousto-optic device consisting of a solar cell, which is used for detecting signal and providing power to the receiver, and an earphone (or a speaker) for sound output.

In addition to the one-to-one link, the VL links also can be employed to achieve direct, peer-to-peer communication between plural portables and/or plural fixed terminals. In the future, this type of ad hoc interconnection is well suited to an indoor ubiquitous Internet of things (IoT).

35.5.2 Visible-light Wireless LAN (VLW-LAN)

The LED-lamp-based VLW-LAN system is shown in Fig. 35.25. It is an LED-based multiple-access optical network. The LED lamp is used as an optical access point (OAP) that can establish multiple links to available portable terminals, and to bridge between VL links and Ethernet LAN or achieve access to resources on an Ethernet.

In general, the link of a VLC system has an uninterrupted line-of-sight (LOS) path between the transmitter and the receiver. The LOS links may employ three designs, which are directed, nondirected, and hybrid links according to the degrees of directionality of the transmitter and receiver. Directed links employ directional transmitters and receivers, which must be aimed in order to establish a link. Nondirected links employ wider-angle transmitters and receivers, alleviating the need for such pointing. Hybrid links combine transmitters and receivers having different degrees of directionality.

A LED-based VLW-LAN system establishes a nondirected downlink and uplink because both transmitter and receiver are wider-angle units. A nondirected link may be more convenient to use for mobile terminals, since it does not require aiming of the transmitter and receiver. When several nondirected transceivers are located in proximity to each other, they naturally form a shared bus topology, making

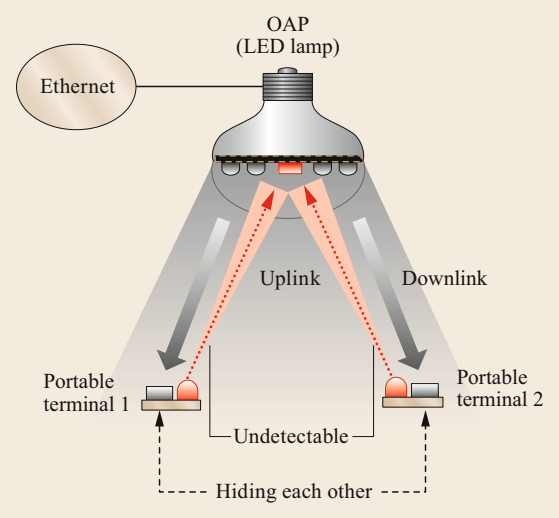


Fig. 35.25 LED-lamp-based VLW-LAN system

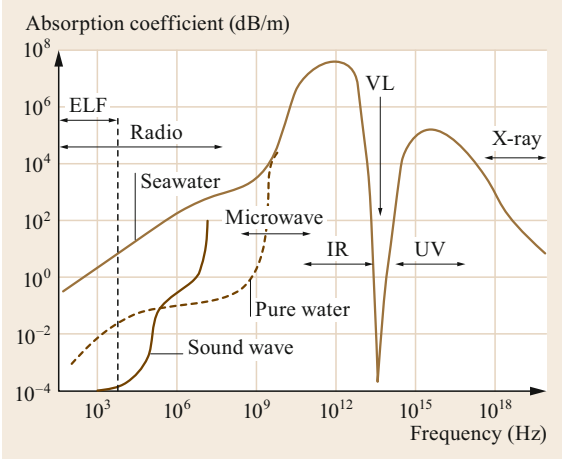


Fig. 35.26 Absorption coefficients of electromagnetic waves and sound waves in water

nondirected links suitable for multiple-access LAN. However, a nondirected link has a higher loss than the other two types, requiring higher transmit power and a receiver with a larger light-collection area.

For the VLW-LAN system, *hidden nodes* may be present, i.e., receivers that are undetectable to other receivers because of the inherent directivity of light propagation, as shown in Fig. 35.25. A request to send/clear to send/data/acknowledge (RTS/CTS/DATA/ACK) protocol [35.48] can be used to solve the problem of the hidden nodes. In addition, carrier-sense multiple access with collision detection (CSMA/CD) or with collision avoidance (CSMA/CA) protocols [35.48] can be used to control and access multiple terminals.

As a complementary technique to wireless LAN WiFi (wireless-fidelity), the VLW-LAN is also called LiFi (light-fidelity). Its attractive features are saving of frequency resources, ensuring security between adjacent wireless LANs, and the fact that it can create an underwater optical wireless sensor network (UOWSN) because visible light can penetrate water. In addition, *ubiquitous* is a keyword in the current information age, and the OAP used in the VLW-LAN is a ubiquitous LED lamp, hence it is possible to make a common illumination environment become a ubiquitous information-service environment.

35.5.3 Underwater Visible-light Wireless Communication (UVLWC)

Comparisons of the absorption coefficients and carrier characteristics of underwater radio, sound, and VL waves are shown in Fig. 35.26 [35.49] and given in Table 35.6.

The sound-wave communication technique is used in almost all present commercial underwater data transmission systems as it can propagate well in seawater and can reach far distances (up to several kilometers). Unfortunately, sound-wave communication systems do not enable large-capacity links, such those needed for image data transmissions, because they are inherently low speed and a Doppler effect is caused by this low speed. Moreover, when the communication system is dynamic, severe frequency-dependent dispersion may arise even at short ranges.

An alternative to sound-wave communication is using radio waves at extremely low frequencies (ELF: < 10 kHz) due to their low absorption coefficient in pure seawater, as shown in Fig. 35.26. Despite this, radio-wave-based underwater communication still presents some intrinsic drawbacks. In fact, the major obstacle in using radio for underwater communication is the severe attenuation due to the conducting nature of seawater. In particular, the attenuation is very high for

Table 35.6 Comparisons of carrier characteristics of underwater radio, sound, and VL waves

	Radio wave	Sound wave	VL wave
Distance	< 1 m in ELF	> 10 km	15–100 m (depending on the seawater turbidity)
Speed	Extremely low speed	Low speed (< 1500 m/s)	High speed (large capacity)
Problem	Very low speed so cannot use terrestrial technique	Low speed and Doppler effect	Spectral attenuation

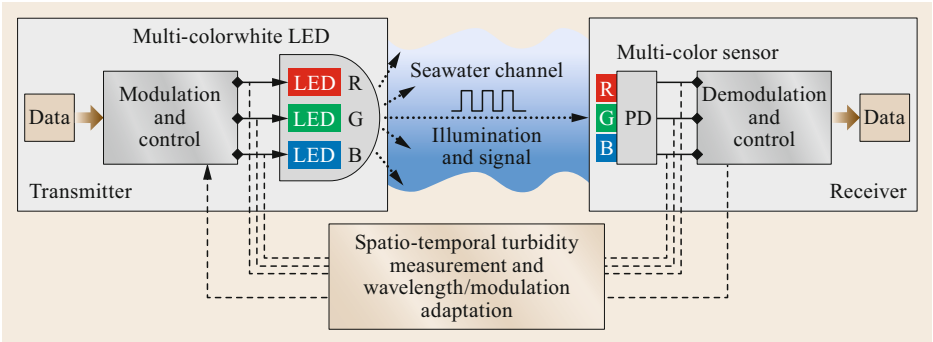


Fig. 35.27 Adaptive UVLWC system

high-frequency radio waves and, since the current terrestrial technology for wireless communication is often based on high frequencies in the order of Gbps, it is practically impossible to use terrestrial techniques in underwater applications.

VL-based underwater communication techniques are being considered as a possible solution to this because of their large data capacity due to the inherent wide bandwidth. Moreover, seawater exhibits a window of reduced absorption in the visible-spectrum range; particularly light wavelengths between 400–650 nm, where water is relatively transparent to light and absorption takes its minimum value, as shown in Fig. 35.26. Nevertheless, for different marine environments and turbidities, VL will undergo wavelength-dependent attenuation in seawater. Hence, an underwater visible-light wireless communication (UVLWC) system must satisfy a large range of noise-control conditions.

Adaptive control techniques can be applied to help overcome the seawater turbidity of the spatiotemporal change to obtain efficient and reliable light propagations over larger ranges and construct a robust link. Figure 35.27 shows an adaptive UVLWC system with wavelength- and modulation-adaptation control.

The link between the transmitter and the receiver is commonly of IM/DD type. The *L*PPM scheme is used to perform modulation-adaptation control. As previously mentioned, for *L*PPM, the SNR requirement to reach a certain BER is decreased with increasing L, hence in cases of longer-distance turbid seawater channels, by adjusting L communication quality can be ensured. However, note that an increase in the L causes a decrease in the communication speed. Alternatively, in order to reach wavelength-adaptation control, a multicolor white LED can be used. Each color chip has an independent and different wavelength peak, which can act as a separate channel and is controlled for seawater turbidity adaptation. At the receiving side, a multicolor sensor is used to receive different-color light from each wavelength channel. Each color channel can continuously measure the intensity of monochromatic light in the color channel. The measurement results inform which color of light is selected for data transmission. This optical measurement can be achieved by automatically switching the three-color PDs.

The UVLWC can be used to construct an UOWSN [35.50]. In this case, the links between each subnode are VL wireless channels. The space division and visibility of VL can ensure each subnode is independent and identifiable both in space and time. These subnode data are transmitted long-distance to a land station by using optical fibers via a main node. UVLWC also can be used for short-distance and high-speed data transmissions between an autonomous underwater vehicle (AUV) or a remotely operated vehicle (ROV) and an underwater detector [35.51].

References

35.1 A. Selvarajan, S. Kar, T. Srinivas: Overview of optical fiber communications. In: *Optical Fiber Communications*, ed. by G. Keiser (McGraw-Hill, New York 2002)

35.2 A.G. Bell: The photophone, Science **os-1**(12), 130–134 (1880)

35.3 T. Aruga: *Spatial Transmission Optics* (Suiyosha, Tokyo 2000)

35.4 S. Endo: *History of Light* (Tosho, Tokyo 1977)

35.5 Nakagawa Laboratories: Illuminative light communication device, JP3827082 (2006) and US7583901 B2 (2009)

35.6 S. Haruyama: Visible light communications, IEICE Trans. **J86-A**(12), 1284–1291 (2003)

35.7 M. Nakagawa: Ubiquitous visible light communications, IEICE Trans. **J88-B**(2), 351–359 (2005)

35.8 X. Lin: Optical wireless ubiquitous information service using LED lighting, Mon. Disp. **18**(10), 46–52 (2012)
35.9 IEEE: *Standard 802.15.7-2011: Short-Range Wireless Optical Communication Using Visible Light* (IEEE, Piscataway 2011)
35.10 J.M. Kahn, J.R. Barry: Wireless infrared communications, Proc. IEEE **85**(2), 256–298 (1997)
35.11 K. Ando: *Principles of LED* (Ohmsha, Tokyo 2010)
35.12 N. Hirosaki, N. Kimura, K. Sakuma, S. Hirafune, K. Asano, D. Tanaka: White light-emitting diode lamps for lighting applications, Fujikura Tech. Rev. **109**, 1–4 (2005)
35.13 X. Lin: Chapter 9.2, *LED-Based Illumination-Light Communication Device* (Technical Information Institute, Tokyo 2014) pp. 609–614
35.14 T. Taguchi: White LED with high color rendering, ULVAC Tech. J. **53**, 14–17 (2008)
35.15 A. Selvarajan, S. Kar, T. Srinivas: Chap. 4.1.1, *Optical Fiber Communication* (McGraw-Hill, New York 2002) pp. 53–54
35.16 K. Yamamoto: Laser lighting by visible laser diode. In: *Proc. 9th Light-Tech Expo Conf., Keynote Session (Tokyo, 2017)* (2017) pp. 21–44
35.17 Hamamatsu Photonics: *Si APD, MPPC* (Hamamatsu, Hamamatsu 2017)
35.18 J. Sakai: *Optical Communication Engineering* (Kyoritsu Shuppan, Tokyo 2000)
35.19 Hamamatsu Photonics: *Image Sensors* (Hamamatsu, Hamamatsu 2017)
35.20 H. Tanaka, M. Bandai, T. Watanabe: Fundamental discussion and experiments of visible light communication using two-dimensional code, WiNF **2010**(8), 165–170 (2010)
35.21 X. Lin: Visible-light wireless communications technique using LED lighting, IEICE Tech. Rep. **115**(247), 63–68 (2015)
35.22 V. Manea, R. Dragomir, S. Puscoc: OOK and PPM modulations effects on bit error rate in terrestrial laser transmissions, Telecomunicaii **2011**(2), 55–61 (2011)
35.23 M.D. Audeh, J.M. Kahn, J.R. Barry: Performance of pulse-position modulation on measured nondirected indoor infrared channels, IEEE Trans. Commun. **44**(6), 654–659 (1996)
35.24 T. Saito, S. Haruyama, M. Nakagawa: A study for flicker on visible light communication, IEICE Tech. Rep. **106**(450), 31–35 (2007)
35.25 JEITA: *CP-1222: Visible Light ID System* (JEITA, Tokyo 2007)
35.26 JEITA: *CP-1223: Visible Light Beacon System* (JEITA, Tokyo 2013)
35.27 G. Ntogari, T. Kamalakis, J. Walewski, T. Sphicopoulos: Combining illumination dimming based on pulse-width modulation with visible-light communications discrete multitone, J. Opt. Commun. Netw. **3**(1), 56–65 (2011)
35.28 N. Fujimoto: The fastest modulation method for visible light communications, News Release of Kindai University, 15th October (2012).
35.29 K.-P. Ho: Optical duobinary modulation, Recent Pat. Eng. **4**(2), 80–85 (2010)
35.30 R. Kaur, S. Dewra: Duobinary modulation format for optical system—a review, Int. J. Adv. Res. Electr. Electron. Instrum. Eng. **3**(8), 11039–11046 (2014)
35.31 H. Mizuno, S. Choi, A. Yokoi: Performance of CSK communication systems with displays and cameras, IEICE Tech. Rep. **114**(160), 63–68 (2014)
35.32 J. Armstrong: OFDM for optical communications, J. Lightwave Technol. **27**(3), 189–204 (2009)
35.33 H. Elgala, R. Mesleh, H. Haas: Indoor broadcasting via white LEDs and OFDM, IEEE Trans. Consum. Electron. **55**(3), 1127–1134 (2009)
35.34 M.Z. Afgani, H. Haas, H. Elgala, D. Knipp: Visible light communication using OFDM. In: *2nd IEEE Intl. Conf. TRIDENTCOM 2006* (2006), https://doi.org/10.1109/TRIDNT.2006.1649137
35.35 H. Qian, S. Cai, S. Yao, T. Zhou, Y. Yang, X. Wang: On the benefit of DMT modulation in nonlinear VLC systems, Opt. Express **23**(3), 2618–2632 (2015)
35.36 J. Vucic, C. Kottke, S. Nerreter, K.-D. Langer, J.W. Walewski: 513Mbit/s Visible light communications link based on DMT-modulation of a white LED, J. Lightwave Technol. **28**(24), 3512–3518 (2010)
35.37 T. Komine, M. Nakagawa: Fundamental analysis for visible-light communication system using LED lights, IEEE Trans. Consum. Electron. **50**(1), 100–107 (2004)
35.38 Y. Tanaka, T. Komine, S. Haruyama, M. Nakagawa: Indoor visible communication utilizing plural white LEDs as lighting. In: *12th IEEE Int. Symp. PIMRC 2001* (2001) pp. F81–F85
35.39 H. Sugiyama, S. Haruyama, M. Nakagawa: Analysis and experiment of communication distance in visible light communication system, IEICE Tech. Rep. **106**(598), 25–30 (2007)
35.40 S. Horikawa, S. Haruyama, M. Nakagawa: Visible light positioning system, Tech. Rep. IEICE WBS2003-39, 41–44 (2003)
35.41 Y. Yasuda, X. Lin: Development of the facility guidance system by visible light communication, Proc. IEIEJ **I-11**, 467–470 (2013)
35.42 X. Lin, H. Itoh: LED light equipment with optical wireless communication functions. In: *Proc. World Eng. Conv* (2011)
35.43 X. Lin, K. Hirohashi, K. Ikawa: Optical wireless LAN integrated systems with illumination function, IEICE Tech. Rep. **109**(400), 43–48 (2010)
35.44 Y. Aoyama, A. Ogawa: Optical wireless communications with LED traffic light systems. In: *Proc. SITA 2006, Hakodate* (2006)
35.45 H. Watanabe, T. Nishii, S. Oshiba: Expansion of transmission distance on communication system between the traffic lights using visible light communication, IEICE Trans. B **J98-B**(2), 180–187 (2015)
35.46 X. Lin: Underwater wireless communication system of adaptation wavelength using visible light, IEICE Trans. Fundam. **E100-A**(1), 185–193 (2017)
35.47 L.J. Johnson, F. Jasman, R.J. Green, M.S. Leeson: Recent advances in underwater optical wireless communications, Underw. Technol. **32**(3), 67–175 (2014)

35.48 IEEE: *Standard 802.11: Wireless LAN* (The Institute of Electrical and Electronic Engineers, Piscataway 1998)
35.49 S. Nakao: Attenuation of electromagnetic waves in seawater, Def. Technol. **1987**(9), 22–30 (1987)
35.50 L. Ghelardoni, A. Ghio, D. Anguita: Smart underwater wireless sensor networks. In: *IEEE 27th Conv. Electr. Electron. Eng. Israel* (2012) pp. 1–5
35.51 T. Sawa, X. Lin: Research of underwater optical wireless robust communication, Jpn. J. Opt. **45**(2), 55–61 (2014)

Xin Lin

Institute of Technology
Nakagawa Laboratories, Inc.
Tokyo, Japan
linxin@optinformation.com

Xin Lin received her PhD in Optoelectronics from Shizuoka University, Japan, in 1996. During 1997–2009, she was a researcher at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. She is currently Senior Researcher at Nakagawa Laboratories, Inc., Japan. Her research interests include optical wireless communications, optical information processing, and optical neural networks. She is Senior Member of OSA and IEICE, and a member of SPIE and JSAP.

Tomokuni Matsumura

Nakagawa Laboratories, Inc.
Tokyo, Japan
tom@naka-lab.jp

Tomokuni Matsumura received his MBA degree from University of Dallas, Texas, USA (1981) and his Bachelor's in Business/Commerce from Keio University, Japan (1972). In 2003–2016, he was President at Nakagawa Laboratories, Inc, starting VLC (visible light communications) together with the Science and Technology Department at Keio University, Japan. During 2003–2016 he was also Executive Director of VLCC (Visible Light Communications Consortium). He is currently an advisor at NAKATEN.

36. Optical Communications and Sensing for Avionics

Alireza Behbahani, Mehrdad Pakmehr, William A. Stange

This chapter is a review of avionics optical fiber communication and sensing systems. Optical fibers are the most common means of optical communication. One important extension of fiber-optic communication technology for aerospace applications is the fly-by-light (FBL) approach for aircraft control systems. Inherent characteristics of the FBL approach such as its light weight, compact size, large bandwidth, and immunity to electromagnetic interference (EMI) make it an ideal technology for future flight control systems. FBL control systems offer some advantages for the new aircraft generation within more hostile military environments. In addition to their value in communication, optical fibers have the capability to sense and provide information about the environment they are exposed to. There are various types of fiber-optic sensors that can provide detailed data on parameters such as temperature, strain, pressure, vibration, and acoustic emissions. Accurate knowledge of these parameters is vital to the structural design and safe operation of various components that must often operate in high temperature or mechanically hostile environments; well-designed optical fiber sensors can provide measurement possibilities in these environments. This chapter will provide insight into a number of these optical sensing techniques, which have demonstrated significant potential for enabling accurate measurements in harsh environments.

The main characterization of fiber-optic communication systems is the use of light as the form of energy that is being modulated and the use of optical fibers to propagate that energy from the source to the destination. The main benefit of using light energy for communication is that light can be modulated easily with high-speed signals and transported over long distances in an optical fiber with minimal degradation.

36.1 Fiber-Optic Communication

This section summarizes basic concepts regarding fiber-optic communication that will be useful for the rest of the chapter, and that, for interested readers, can be found in a much more extended form in the previous chapters of this Handbook.

The basic principle by which optical fibers transmit light is that fibers are constructed with a transparent core surrounded by a transparent cladding material with a lower index of refraction. With a lower index of refraction, light potentially leaking out of the

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4_36

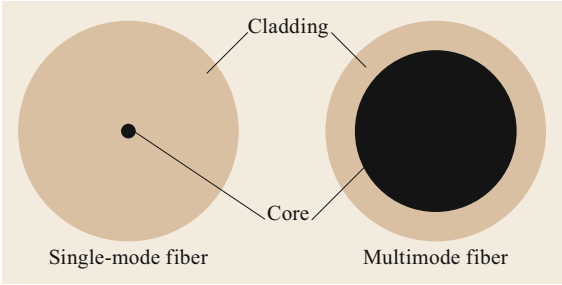


Fig. 36.1 A single-mode fiber (*left*) has a core that is very small compared with the cladding, whereas a multimode fiber (*right*) can have a large core (after [36.1])

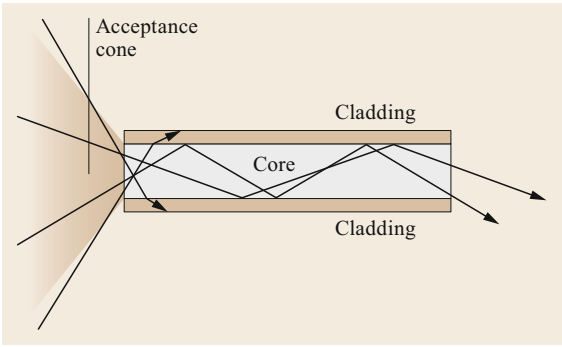


Fig. 36.2 The propagation of light through a multimode optical fiber (after [36.2])

core is preferentially bent back toward the core by the cladding. This phenomenon causes the fiber to act as a waveguide to efficiently transmit light from an input source to the end of the fiber.

The index of refraction of the fiber core determines the acceptance angle or acceptance cone. A multimode fiber has large core (Fig. 36.1) resulting in a larger pupil and allows a variety of light modes to be propagated by the fiber (Fig. 36.2). A single-mode fiber has a very small core (Fig. 36.1) that only accepts light.

Optical fibers, especially those used for communication, are mostly made from silica. Silica has been used extensively in many optical applications. This has led to significant developmental work on silica, which has improved its performance over other fiber-optic materials. Other materials, such as plastic, heavy-metal fluoride (primarily fluorozirconate and fluoroaluminate), chalcogenide glasses, polycrystalline silver halides and hollow waveguides, as well as crystalline materials like sapphire, have been used for other specialized applications.

Plastic fibers are limited by their low-temperature capability. Heavy-metal fluoride glass fibers used at mid-infrared (MIR) wavelengths have significant drawbacks when used for avionic and autonomous vehicle communication. These include brittleness, limited mechanical strength, low glass transition temperature, high fabrication cost, and chemical reactivity with water. Chalcogenide glass fibers contain one or more of the chalcogen elements (e.g., sulfur, selenium or tellurium) and other additional elements such as arsenic, germanium, antimony, gallium, etc. Like the fluoride glasses, they are weak in shear strength, which makes them too fragile for most of the desired applications. Polycrystalline silver halide fibers have a large transparency window, but they too are weak, sensitive to light, and have low melting points. Dielectric-coated, metallic hollow waveguides are excellent candidates for laser power delivery, but they have a larger diameter than most solid-core fibers, which can limit their flexibility. Furthermore, hollow waveguides have additional transmission losses with bending, which further limits their usefulness for communication. Sapphire fibers, although expensive to fabricate, offer excellent high-temperature capability and are chemically inert making them very strong candidates in those applications where their temperature capability justifies their fabrication cost.

Although fiber-optic cables are flexible and thin, unaffected by electromagnetics, and able to be made to varying degrees resistant to high temperature and harsh environments, they do have a few shortcomings.

The primary problem associated with any optical fiber, whether used for communication or for sensing purposes, is the fact that the full length of the fiber, from light source to sensing location, must be affixed to, or potentially embedded within, a structural component of the avionics or vehicle without incurring any breakage or significant degradation of the fiber. Many of the fibers typically used by the telecommunications industry would not survive the necessary application methods or the difficult environments of avionics and autonomous vehicles. Research efforts conducted by the Air Force Research Laboratories (AFRL) Turbine Engine Division in the early 1990s attempted to embed fiber-optic sensors in metal matrix composites (MMCs). These efforts demonstrated that a metal coating could be successfully applied to the outer surface of the fiber's cladding [36.3]. This metal coating was initially intended to enhance the structural integrity of the fiber to enable it to survive the MMC fabrication process. While they were not able to successfully embed a fiber within an MMC, metal-coated fibers do have some significant benefits. Chief among these benefits are increased strength and enhanced fiber durability. The use of a metal coating may provide enough strength to allow heavy-metal fluoride, chalcogenide glasses, or polycrystalline silver halide materials to be successful used. The feasibility of using these materials allows us to expand the wavelengths of light that could potentially

be used for optical communication, which may be of significant value depending on the application.

While it is successful in providing increased strength and durability to the fiber, the metal coating also provides the fiber with unique capabilities. Chief among these is the ability to transmit electric current along the length of the fiber. This capability could allow an optical fiber to serve as both a sensor as well as a power source and/or a data transmission line for conventional electronic communications. Multiplexing or layering the metallic coating could allow multiple electric transmission paths to power or relay data from several conventional sensors or different types of sensors. The ability of metal-coated fibers to transmit electricity and to provide electronic communications significantly expands the way communication between avionic components can be done.

Although it has not yet been adequately researched, a metal-coated fiber could also have the potential to serve as an antenna for wireless sensor applications, potentially enabling remote wireless sensors to transmit data to a central receiver located at the terminus of the fiber. This concept needs to be further investigated, as it could also be a significant benefit to facilitating fiber-based communication once any technical shortcomings have been addressed.

In addition to material considerations, the type of fiber needs to be considered. As discussed above, two main types of optical fiber used in fiber-optic communications include single-mode and multimode fibers. A multimode optical fiber has a larger core (typically greater than 50 μm), allowing less precise, cheaper transmitters and receivers to connect to it as well as cheaper connectors. However, a multimode fiber introduces multimode distortion, which often limits the bandwidth and length of the link. Furthermore, because of its higher dopant content, multimode fibers are usually more expensive to fabricate (fiber material is usually cheaper for single-mode fibers, but connectors could be more expensive) and exhibit higher attenuation.

Connecting two optical fibers can be done by fusion or mechanical splicing or mechanical connectors. With either of these two methods, fiber splicing requires special skills and interconnection technology due to the detailed microscopic precision required to align the fiber cores, which is crucial to the success of the splicing effort.

A more comprehensive and general overview of fiber-optic communication systems and the needed fiber-optic components is provided in [36.4, 5].

36.2 Current and Future Flight Control Systems

An overview of current flight/engine control systems and possibilities of optical fiber applications for future distributed flight/engine control systems is provided in this section.

36.2.1 Distributed Control Architecture for Engines

The current centralized flight/engine control architectures are typically characterized by a single, engine-mounted full authority digital electronic controller (FADEC), which connects directly to every control system element [36.6]. This control system architecture is based on a centralized design in which discrete sensors and actuators are directly wired to an engine-mounted electronics package. FADEC contains all the necessary circuitry to properly interface with engine control devices as well as cockpit command and data communications. The design of this centralized engine control system is primarily about minimizing the control system weight because of its effect on overall vehicle performance. Much of the design is based on legacy, the result of incremental improvement over many years. This creates constraints in the design choices for avionics engineers. A centralized engine control architecture is shown in Fig. 36.3. The electronics package of the FADEC is environmentally controlled and it assures the survivability of all of the circuitry needed for sensor and actuator operations and control algorithm processing. A description of this system is provided to set a baseline for considering the impact of a distributed architecture on communications and data flow.

In a distributed engine control system architecture, as shown in Fig. 36.4, any number of control elements are tied together through a common, standardized, *communication bus*. Sensors and effectors are replaced by nodes (typically some form of electronics) that may provide sensor data, operate actuators, or perform combinations of both. The massive wiring harness, which previously tied together the control element to interface circuitry in the engine-mounted avionics package, is replaced by a simple but robust communication structure. Weight reduction is usually the primary motivation but the standardized interface is also an enabling characteristic, which is necessary to effectively implement future advanced and modular control and prognostics and health monitoring applications.

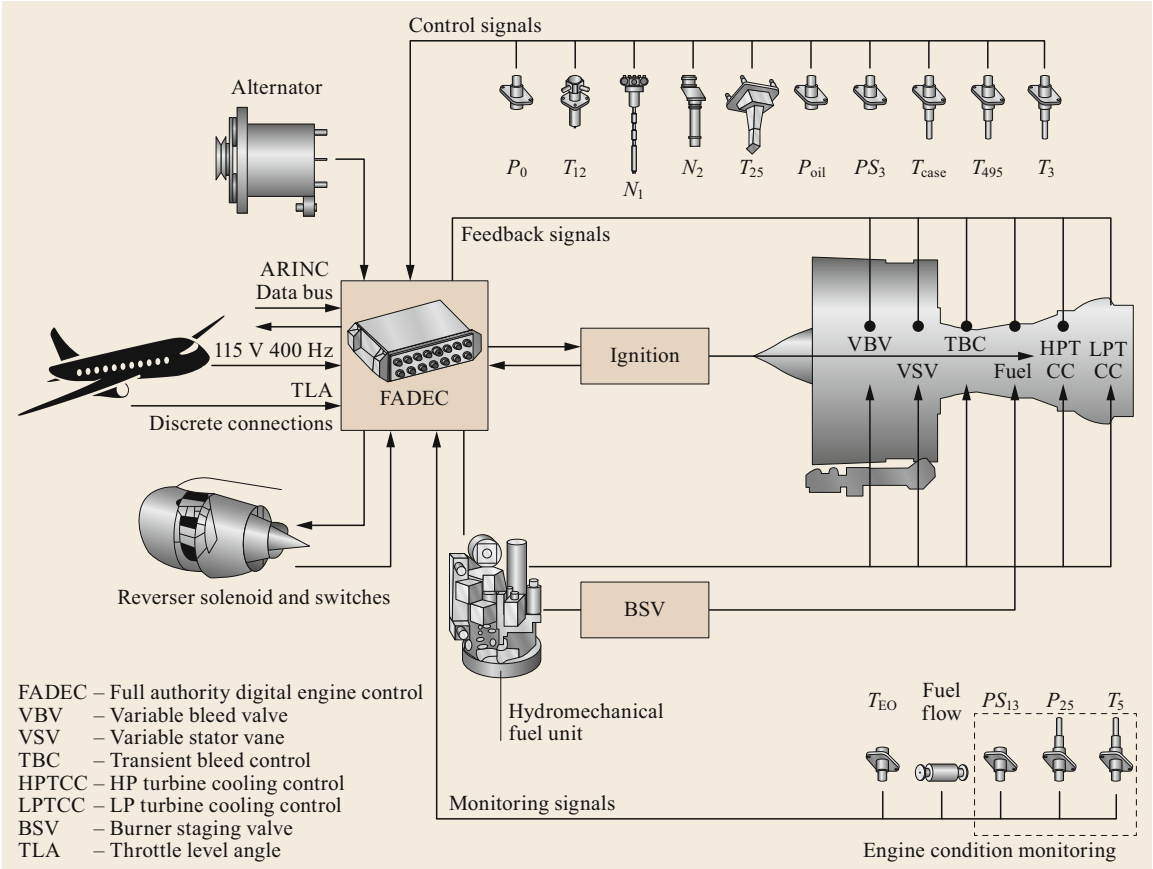


Fig. 36.3 A generic engine control system hardware connection diagram with a centralized FADEC (after [36.6])

More comprehensive discussions about distributed engine control architecture and its benefits for future aircraft systems are provided in [36.6, 8–11].

Future engine control architectures might include various fiber-optic and legacy sensors along with different engine actuators and control processors. The fully distributed system is an extension of Fig. 36.3 into discrete control elements directly connected to the communication system and is shown in Fig. 36.4. Local control nodes are not physically separate electronics enclosures but could be physically integrated into the sensors and actuators themselves. Control loop-closure functions, which are not described, are performed as virtual functions anywhere in the system where the processing capability exists to do so. Separate control law functions might be integrated into the individual actuators or they could continue to reside in a physically separate FADEC. Optical fiber technology, including optical communication and data networks and distributed optical sensing systems, could provide the possibility of developing fully distributed flight/propulsion control architectures.

36.2.2 Fly-by-Light Systems

One important extension of the fiber-optic communication technology is in aircraft flight control applications. Fly-by-light (FBL) control systems have been investigated previously by various researchers [36.7, 12–23]. FBL control systems are different from fly-by-wire (FBW) since they use optical fibers instead of wires to transmit data from the flight control computers to the actuator controller. Fiber-optic systems can transmit multiple channels of bidirectional information with lighter hardware, are intrinsically immune to electromagnetic interference, and possess a broader transmission spectrum. The FBL system transmits a redundant signal, which requires a second wire bundle in a traditional FBW system [36.23].

Bidirectional transmission of optical signals over a single optical fiber or network (without optical isolators or optical amplifiers) is equivalent to transmitting electrical signals over a coaxial or a twisted-pair wire cable. The main reason for considering bidirectional transmission over a single optical fiber instead of two

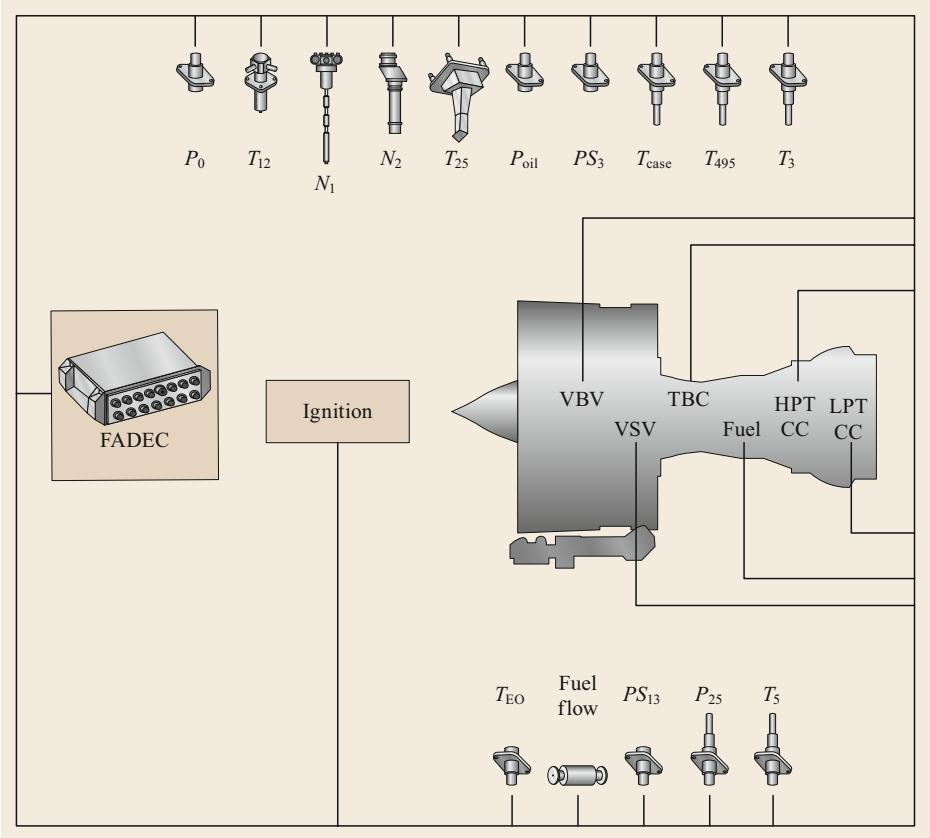


Fig. 36.4 A fully distributed control system. Each system element individually connects to the network. Each physical element can have multiple functions, some of which require real-time communication for control and others that may be less time critical (after [36.6])

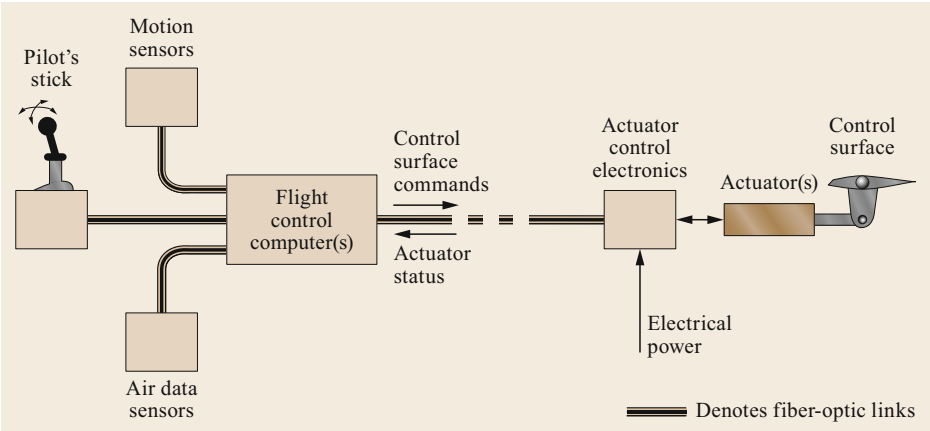


Fig. 36.5 A typical fly-by-light flight control system; redundancy is omitted for clarity (after [36.7])

unidirectional fibers is the reduction of the components/weight (fibers, optical splitters and optical amplifiers) by a factor two and the potential cost reduction by an integrated transceiver design. However, bidirectional transmission might introduce other costs and complications to the system design. A special optical component is also required to duplex the bidirectional signals at the transceiver and the crosstalk between the bidirectional signals. Other advantages of bidirectional optical communication are increased network capacity (since the fiber capacity is increased by operating at more than one wavelength) and increased reliability (since there are fewer connections or end points in the network, the fiber solution is less susceptible to connection errors).

A typical FBL system architecture schematic is shown in Fig. 36.5 [36.7]; the redundancy of the system

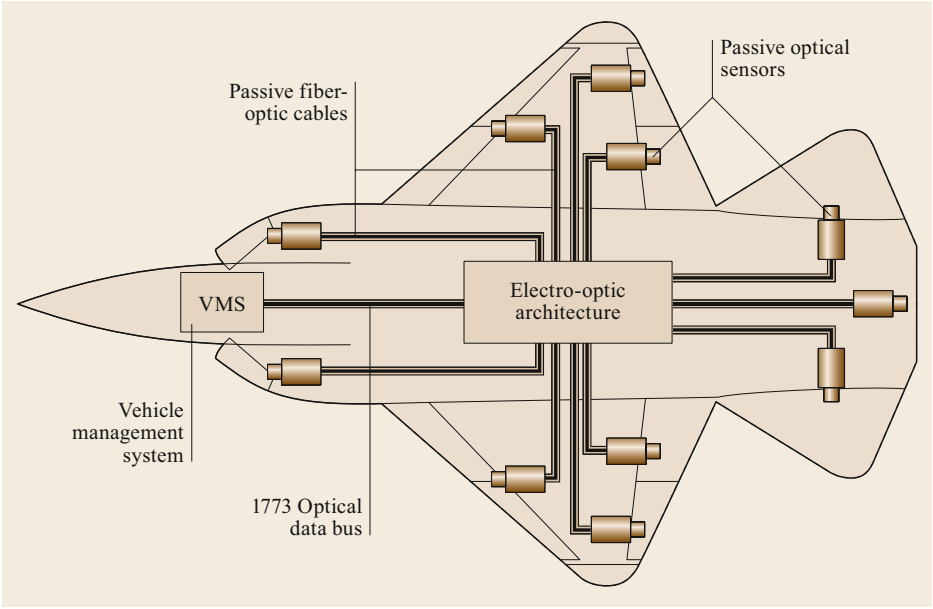


Fig. 36.6 NASA's FBL flight control system block diagram (after [36.17])

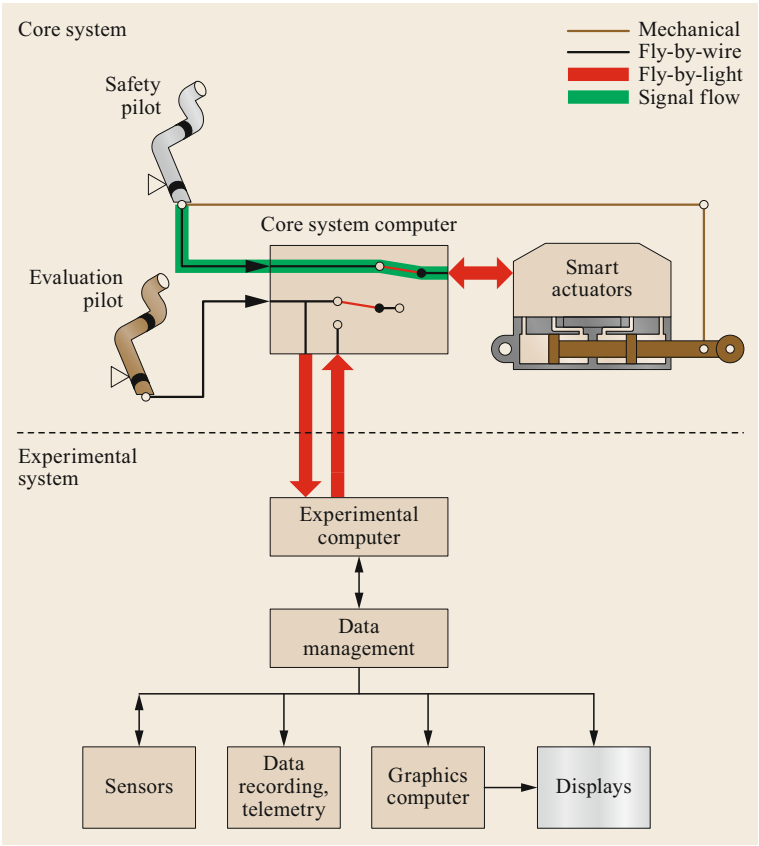


Fig. 36.7 Combined FBW/FBL system in DLR EC 135 ACT/FHS (after [36.24])

is not shown in this schematic for clarity. The fiber-optic links interconnecting the units of the flight control system eliminate the possibility of spreading electrical faults between units, as the optical fiber is an insulator. Fiber-optic links can be bidirectional and can also be used to provide the system status to the pilot's control/display panel. The bidirectional system provides the possibility to check that the data have reached their destination. Moreover, the status or 'health' of the control surface actuation system can be checked by monitoring the servo error signals and actuator hydraulic pressures [36.7].

Even though fiber-optic technology has extensive applications in communications, it has not yet seen significant applications in civil aviation due to integration, validation, reliability, and cost factors. During the 1990s, NASA designed, developed, and tested fiber-optic sensors and multiplexing electro-optic architecture for installation and flight test on a NASA-owned F-18 aircraft. This hardware was developed under the Fiber Optic Control Systems for Advanced Aircraft program, as a part of a multiyear NASA initiative to demonstrate FBL systems for application to advanced aircraft flight and propulsion control [36.14, 17]. Figure 36.6 shows a representative FBL aircraft control system developed by NASA with major components identified [36.17].

Several of the latest fighter aircraft, including the Eurofighter Typhoon, use a fiber-optic control link [36.22], and Gulfstream Aerospace recently flight-tested an FBL version of the G650 [36.23]. The active control technologies (ACT)/flying helicopter simulator (FHS) of the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt; DLR), which is based on a standard Eurocopter EC 135 type helicopter, has been modified for use as a research and test aircraft; the mechanical controls have been replaced by an FBW/FBL flight control system. The control commands are transferred by electric cables and fiber-optic cables instead of control rods [36.25]. A schematic of the integrated FBW/FBL system in DLR EC 135 ACT/FHS [36.24] is shown in Fig. 36.7. Another recent application of fiber-optic communication systems is for the F-35, Joint Strike Fighter (JSF); the JSF mission system's high speed data network is based on the American National Standards Institute (ANSI) standard for 2 Gb/s optical fiber channel. The aircraft fiber-optic cabling plant is a standard multimode 50/125-μm fiber [36.26]. More examples of applications of optical communications for flight control and FBL systems can be found in [36.24].

Future flight/engine control architectures will include various fiber-optic and legacy sensors along with advanced flight/engine actuators and new control processor technologies (Figs. 36.8 and 36.9). These new control systems will need to be able to integrate heterogeneous components—such as legacy sensors and actuators, as well as fiber-optic sensors—in a distributed architecture that uses a fiber-optic network bus as its databus. Single or multiple processors can be used for control law computations. Figure 36.9 provides a conceptual schematic of this kind of distributed optical control architecture for use in gas turbine engine systems.

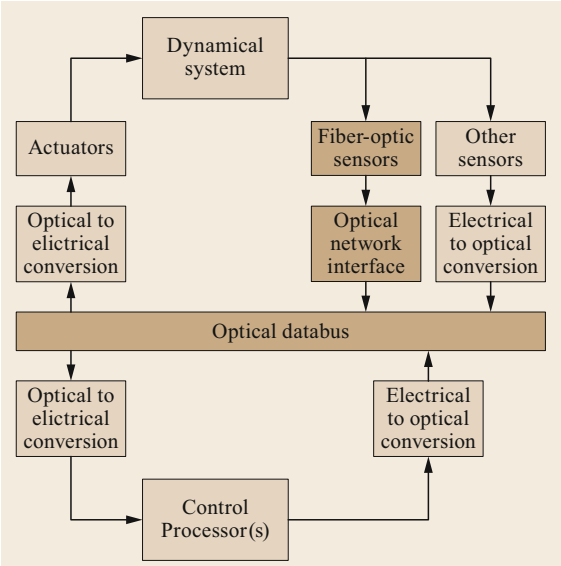


Fig. 36.8 Schematic of a simple optical network control system architecture

An optical network control system can be a spatially distributed system in which the communication between legacy sensors, fiber-optic sensors, actuators, and control processors occurs through a shared optical communication network. Some conceptual plans for this are shown in Figs. 36.8 and 36.9. In an optical network control system bandwidth, packet dropout, and network-induced delays are not typically critical issues [36.27] for proper operation of the control system. However, depending on the overall optical network control system configuration, the effects of jitter and noise on the stability of the entire system need to be studied.

The bandwidth of optical fiber networks can be further increased using wavelength-division multiplexing (WDM), a technique that uses lasers with different wavelengths to transmit data on a single optical fiber [36.27]. This technique can be broken down further into coarse WDM or dense WDM depending on the specific channel spacing. References [36.28–36] provide an overview of the development of WDM for use

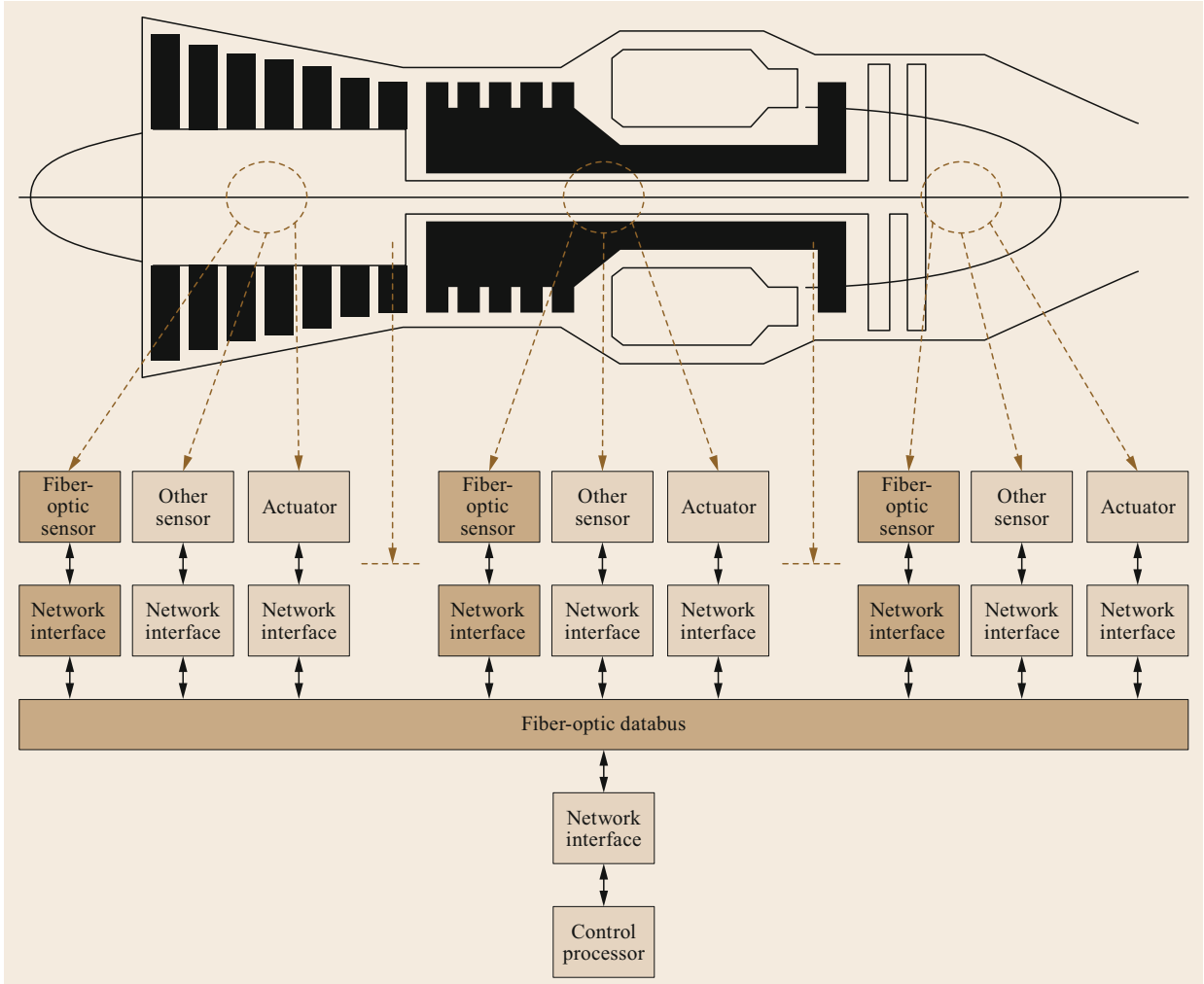


Fig. 36.9 Conceptual schematic of distributed control architecture using optical databus for gas turbine engine applications. This is a network of fiber-optic sensors, nonoptical sensors, actuators, and control processor(s) connected via a fiber-optic databus

in aerospace applications. WDM-based fiber-optic networks that can work in a temperature range of −55 to +125 °C are discussed in [36.37].

For gas turbine engine applications, high temperature conditions have a deleterious effect on communication in the networked architecture [36.10, 38]; thus, the effects of elevated temperature on the communication systems within the optical network control system framework have to be investigated. In typical distributed engine control system setups, some of the electronics employed as smart nodes (sensors and actuators) typically need to be placed in high-temperature locations, and hence processing capability in those elements could be severely affected by high temperature [36.39].

36.2.3 Avionics Full-Duplex Switched Ethernet

To implement FBL there is a need for a databus that can handle optical signals. Avionics full-duplex switched Ethernet (AFDX) is an existing avionics protocol that utilizes commercial off-the-shelf Ethernet devices (including fiber optics) with modifications for flight control environments. It was specifically designed for real-time control systems by Airbus [36.40] for flight control applications [36.41]. It is a standard that defines the electrical and protocol specifications for the exchange of data between avionic subsystems using IEEE 802.3 for the communications architecture. The AFDX data network is a specific implementation of ARINC (Aeronautical

Radio, Inc.) Specification 664 Part 7, which defines how commercial off-the-shelf networking components will be used for future-generation aircraft data networks. The six primary aspects of an AFDX data network include full duplex, redundancy, determinism, high-speed performance, and switched and profiled network.

Through the use of twisted-pair or fiber-optic cables, full-duplex Ethernet utilizes two separate pairs or strands for transmitting and receiving the data. AFDX extends the standard Ethernet to provide high data integrity and deterministic timing. The issue of deterministic communications is addressed by defining communication virtual links (VLs) between end systems with specified maximum bandwidth, bounded latency, and frame size during system design. Furthermore, a redundant pair of networks is used to improve the system integrity. More information about AFDX can be found in [36.42–44].

In order to realize and harvest the full potential of fiber-optic communication technologies for avionics and flight control applications, there is still a need to develop a better and more sophisticated optical communication databus to be implemented as an integral part of future FBL systems.

36.2.4 Network Topologies for Distributed Controls

Researches have studied a variety of topologies when connecting sensors and actuators using basic ring, star, bus, and tree topologies in the context of implementing a distributed engine control system (DECS). They have been analyzed in terms of their potential impact on network reliability, weight, and maintainability. It is also important to note the effect a given topology might have on the average message transmission length. The vast majority of comparison studies have been primarily qualitative.

The basic network topologies, which will be reviewed here, are [36.45, 46]:

(i) Ring—where each node is connected to the node just before and just after itself.
(ii) Star—where each node is connected to a central node or hub.
(iii) Bus—where each node is connected to a shared backbone medium.
(iv) Tree—where nodes are arranged in a hierarchical fashion, such that there is a root node, interior nodes, and leaves (nodes with no children).

Ring

Ring topology describes a network where each node has two connections. Usually, these connections are to the node that comes before and the node that comes after. This concept is simple to visualize in the case where the nodes are already arranged in a ring-like shape, and every node connects to its two nearest neighbors. However, the physical shape of the network could end up looking quite contorted depending on the locations of the nodes. Ring topology is considered to be weakly connected since in a fully connected network of n nodes there are $1/2\,n(n-1)$ connections while there are only n connections in a ring. Because of this, ring topology may have a large potential for weight savings. In some of the simpler ring networks, data is passed in only one direction (clockwise or counterclockwise), but in this analysis we will consider a ring that is bidirectional (i.e., data is able to be passed in either direction). This means that the failure of a single node or link in the rings will not cause an entire system to fail since there are always two paths between each node. Since the connections in a ring network do not terminate at a central location, accessibility to the network could be difficult for maintenance personnel. The average message transmission length is relatively high for ring topology since a message might have to traverse a significant portion of the ring in order to reach its final destination. Considering this issue, some research has been done on ways to improve scalability in terms of transmission time for ring networks [36.47].

Ring buses have been used in many distributed systems including various proposed IEEE 802 standards, and was adapted from token ring access developed by IBM in the 1970s. The token ring scheme is used to allow a process to add data or messages to the ring when it receives a token. With enough control, each segment of the ring may be allowed to transfer information and data simultaneously, which increases the system communication bandwidth. Ring buses are not inherently fault tolerant in transferring data/information due to their architecture, which includes a critical failure point, namely the communication channel.

For more efficient utilization of processors in a multisensor system approach, those required for a given sensor processing task can be grouped in a single ring. Results computed by the sensor ring, using a gateway node, can be passed to a data fusion processing ring. This nesting of rings is useful until all the bandwidth has been used or the system requirements have been reached [36.48].

Star

In the star configuration, each node connects directly to a central node, usually referred to as the hub. The current centralized control architecture may be viewed as a physical star topology where the FADEC represents the central node. In a distributed architecture, there may

be several stars chained together. In principle, the total number of links in a star topology is similar to that in ring topology, however, the wire length tends to be larger. This is a result of the point-to-point nature of the star topology. On average, the exterior nodes are closer to each other than to the hub. The hub has the potential to be a single point of failure for all nodes connected to it. However, the loss of any exterior node has no adverse effect on the rest of the network. In this sense, the focus on reliability (strictly in terms of network connectivity) is shifted away from the exterior nodes onto the nodes acting as hubs. Maintaining a DECS in a star topology could prove to be easier than in a ring topology since all of the connections terminate at the hubs. These hubs could be strategically placed to provide ease of access to the maintainer. Though each exterior node has only a single connection in this configuration, the hub will have many. Exterior nodes are logically connected via the hub, so there are usually only two links that a message must traverse.

Bus

For the purposes of this paragraph, the word bus will be used to refer to the physical bus network topology as opposed to the data transfer mechanism (i.e., databus) meaning, which is used throughout much of this chapter. Bus topology is characterized by a shared communication backbone. Each node is connected exclusively to the bus, instead of to other nodes, as is the case in the other topologies. Physical bus topology shares many advantages and disadvantages with star topology in terms of fixed message transmission length and maintainability. However, there is the potential to reduce the number of connectors when compared to the star case. Consider the case where there is a single hub in the center of the engine. In this case, a multitude of links converge at the location of the hub. Then consider a bus, which could be extended along the length of the engine. In this case, each node requires just one connector. Although both the bus and the star configurations present a single point of failure, the bus may be more susceptible to physical damage since it spans over a much greater portion of the engine.

Tree

Tree topology consists of a root node that is connected to each of its *children*, who in turn may have their own children. In this way, each node is connected to a single *parent* and to each of its children (except for the root, which has no parent). Trees are typically classified by the number of children they can have. For example, a binary tree is a tree in which each node may have a maximum of two children. A tree structure is inherently hierarchical, which is likely to prove beneficial as control systems become more and more hierarchical. This is the case for supervisory and control tasks where a high-level controller sends commands to low-level controllers [36.49]. The tree structure may also be useful in cases where the controller is partitioned in a functional manner. In [36.6], the authors have proposed using localized control nodes, one each for the inlet/fan, compressor, combustor, and turbine/nozzle. This gives the tree architecture a clear advantage in scalability. In other words, it can add nodes without increasing the capacity requirement of each node and without significantly increasing message transmission delays [36.50]. Performing maintenance on a network arranged in a tree formation could be difficult if the network is spread over a large area. In many ways, tree topology is comparable to a chain of stars.

36.3 Fiber-Optic Sensors

Although they have their own set of implementation issues, fiber-optic sensors offer the sensor community many of the same benefits that optical fibers deliver to the telecommunications industry. They have several potential advantages over electronic-based sensors for making detailed measurements in hostile environments. A single optical fiber with hundreds of multiplexed Bragg gratings can potentially provide measurements of multiple parameters at hundreds of locations [36.51]. Because these sensors are optically based, they are not sensitive to the problems experienced by many of the electronic-based sensors. Fiber-optic sensors are immune to electromagnetics, electrical interference, electronic noise, and shorting, and they do not require any isolation from ground. Optical fibers also allow any electronic components that are needed to convert the optical output of the fibers into digital electronic signals to be remotely located away from high-temperature areas. This capability also allows digital pooling of the collected sensor information to be done.

36.3.1 Fiber-Optic Sensor Design Issues

There are many fiber-optic sensing techniques that may be of value to future digital engine control systems. While some fiber-optic sensing methods provide single-point measurements, a number of sensing technologies can be multiplexed to allow many measurements to be

made along the length of the fiber. Taking advantage of this capability, a single optical fiber can be used to multiplex tens or hundreds of sensor locations [36.51]. Regardless of the specific sensing method being considered, optical fibers in general have several unique capabilities even though they pose their own set of application problems.

One of the problems associated with the use of fiber-optic sensors in hostile environments is their ability to accommodate the temperatures typically found in these applications. Although the extensive use of silica as the material of choice in so many optical applications has led to developments that have improved their performance over other fiber-optic materials, sapphire-based optical fibers have the most potential for use in hostile environments, due to their ability to accommodate high temperatures. Sapphire fibers can be used in temperatures approaching the melting point of sapphire, which is 2053 °C (3727 °F). In addition to their high-temperature capabilities, sapphire fibers are resistant to reactive or corrosive environments, which make them desirable candidates for use in harsh environment applications.

As discussed above, the utilization of metal-coated optical fibers provides added strength and durability to the optical fibers in addition to giving them the ability to transmit electrical current along the length of the fiber. The same benefits metal coatings provide to communication applications apply to fiber-optic sensors. One of these benefits is the availability of an expanded number of fiber-optic materials that could not have been used otherwise due to their lack of strength and fragility. This benefit expands the available spectrum of wavelengths of light that can be used. For example, chalcogenide fibers offer excellent transmission in the infrared region of the spectrum, which may enable a number of additional sensing methods to be considered for any given application. The use of a metal-coated fiber also offers the potential for an optical fiber to serve as both a sensor as well as a power source and/or a data transmission line for conventional electrically dependent sensors.

Another benefit of metal-coated optical fibers is the potential of embedding the fibers within an engine component. While the flexibility and small size of optical fibers without metal coatings have allowed them to be threaded through extremely narrow or treacherous components to provide strain and temperature measurements conducive to *smart skin* applications within some structures, metal-coated optical fibers have far greater survivability, and may be able to be embedded directly within a component while it is being fabricated. While embedding a metal-coated optical fiber is typically very desirable, there are always trade-offs. Having this capability often leads to the need to evaluate the pros and cons of each application prior to planning whether to embed the fiber, or to use surface mounting to apply it. Each method presents both advantages as well as deficiencies, which vary based on the application.

Embedding the fiber allows information to be gained from within the material and offers protection from external environmental issues such as ablation and abuse during assembly of instrumented parts into components. In certain cases, embedding even small fibers can be beneficial to stealth and signature reduction.

These benefits come with a significant increase in the cost of manufacturing the component in which the fiber is embedded. Even with a metal coating, fiber survival can be difficult to achieve given the rigors of some of the component fabrication processes. Great care must be taken to provide for fiber ingress and egress. Running the fiber out the side of a panel is not always an option since panels are typically trimmed and fitted close to other panels. An additional concern is the difficulty of repairing a sensor in the unlikely event that it is damaged.

Surface mounting reverses the positive and negative aspects found with embedding. Because the fiber is mounted on the surface, component fabrication is a separate step, as attaching the fiber need not be part of the component fabrication process, thus providing a significant reduction in production cost. Ingress and egress of the fiber is no longer an issue since the fiber is already on the surface. Part trimming can be performed before the fiber is attached to the component surface. In the (now more likely) event that the fiber is damaged, repair could be completed by repeating the process that mounted it initially, thereby eliminating the need to develop separate techniques for installation and repair.

However, problems still exist with surface mounting that could preclude its use. Because the fiber is exposed on the surface it is subject to abuse, either from the manufacturing and assembly process, or from the detrimental effects of the severe environment in which it will operate. Additionally, if information from within the component is necessary, surface mounting has an obvious disadvantage. Such would be the case if internal temperature measurements for actively cooled components in high heat flux environments were needed. It would also be the case should detailed information be required regarding the internal stresses or strain levels of a component exposed to high levels of mechanical loading.

36.3.2 Fiber-Optic Sensing Methods

Fiber-optic sensing methods can generally be classified as either single-point or distributed sensors. Research efforts conducted by AFRL's Turbine Engine Division

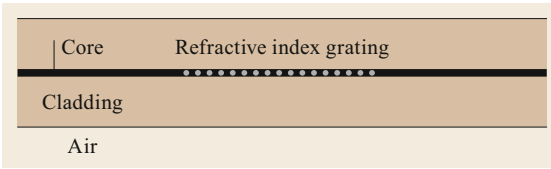


Fig. 36.10 Schematic structure of an FBG; the fiber core has a periodically varying refractive index over some length (after [36.1])

in the 1980s attempted to develop the *twin-core fiber-optic sensor* for use in turbine engines. Although this sensor was able to measure both strain and temperature, it required separate sensing sections to be fabricated along the length of the fiber in order to make distributed measurements [36.52]. The need to fabricate these separate sensing sections proved costly and reduced the number of locations that could be evaluated using this type of sensor.

One of the most versatile optical sensors is the fiber Bragg grating (FBG). FBG sensors are grating-based sensors that are regarded as probably the most mature distributed sensor, and have already been widely utilized in numerous applications. This sensing methodology is a prime candidate for supporting avionics and autonomous vehicles. An FBG sensor reflects a portion of the incoming light of a particular wavelength, called the Bragg wavelength, from a Bragg grating and allows the rest of the incoming light to pass without altering it. Figure 36.10 shows a schematic structure of an optical fiber on which an FBG sensor has been written. Fiber Bragg gratings are created by or written into an optical fiber by exposing the fiber to an ultraviolet interference pattern, which produces a periodic change in the core index of refraction.

The Bragg wavelength is defined by the refractive index grating pitch interference pattern that has been written into the fiber. This grating, and subsequently the Bragg wavelength, is affected by external environment changes, such as temperature, strain, vibration and other parameters. All these changes have an impact on the Bragg wavelength shift. Therefore, by monitoring the Bragg wavelength shift, several measurands can be monitored using FBG sensors. With the recent advancements in wavelength multiplexing technology, hundreds or even thousands of wavelengths can be multiplexed in a single optical fiber. Current technology makes it possible to multiplex tens or hundreds of FBG strain signals in one optical fiber and monitor them remotely. With their rapid development in recent years, FBG sensors have become a leading technology when compared to other competing fiber-optic sensing technologies. Among other materials, Bragg gratings can be written into optical fibers made of sapphire, thereby allowing this sensing method to be utilized at very high temperatures. Besides their wavelength multiplexing capability, FBG sensors have the advantages of low cost, compact size, and excellent linearity.

In recent years NASA has developed what they call the fiber-optic sensing system (FOSS). FOSS essentially uses FBGs in combination with optical frequency domain reflectometry (OFDR) to allow thousands of FBG sensors to be interrogated simultaneously on a single fiber. OFDR systems interpret a beat frequency between a reference arm and a sensing arm where the sensors are located. The phase difference of these interfering signals is then detected by a photodetector and processed using the Fourier transform into the time domain. Once in the time domain, a map of the reflections as a function of fiber length can be developed.

While some of the single-point fiber-optic sensing methods provide excellent measurement capabilities, their ability to make measurements only at a single location, or their need to have multiple sensing sections, makes them less desirable for avionic and autonomous vehicle applications than the optical fibers with multiple sensors distributed along their length, which provide the possibility to measure almost anywhere along the length of the fiber. These fibers are capable of *distributed sensing* and can typically provide measurements at hundreds of locations. Some of the more promising distributed fiber-optic sensing techniques are discussed below.

Optical time domain reflectometry (OTDR) based on Rayleigh scattering, Raman optical time domain reflectometry (ROTDR) based on Raman scattering, and Brillouin optical time domain reflectometry (BOTDR) based on Brillouin scattering are the most widely used time-domain distributed fiber-optic sensor methodologies.

Rayleigh scattering is caused by photons interacting with particles that are much smaller than (less than $1/15$) the wavelength of light. Atoms and ordinary molecules that have a diameter of less than 1 nm are included in Rayleigh scattering. In an optical fiber these particles can be almost any small inclusion or defect in the fiber.

Raman scattering is a change in the wavelength of light due to it being scattered by an atom or molecule. When a photon of light strikes an atom or molecule nearly all the encounters are elastic, and the photons are scattered with unchanged energy and frequency. On rare occasions, however, the molecule takes up energy from or gives up energy to the photons into rotation and vibration excitation, which are then scattered with the scattered photons having a frequency different from, and usually lower than, that of the incident photons.

Brillouin scattering is an effect caused by nonlinearities in the transmission medium, specifically by

that part of the nonlinearity that is related to acoustic phonons. When this occurs, an incident photon can be converted into a scattered photon, usually propagating in the backward direction, and a phonon. The frequency of the reflected beam is slightly lower than that of the incident beam, with the frequency difference corresponding to the frequency of the emitted phonons. The Brillouin frequency shift depends on the material composition and to some extent on the temperature and pressure of the medium.

Optical time domain reflectometry (OTDR) is in the first generation of fiber-optic distributed sensors, where Rayleigh scattering is measured as a way to determine changes in the fiber-optic structure. Rayleigh OTDR testing consists of an optical pulse being introduced into an optical fiber while the power of the Rayleigh, or Fresnel, backscattered light is measured by a photodetector as the light pulse propagates along the fiber. After sending a series of pulses and averaging them, spatial distortion of the fiber can be obtained and correlated to strain and/or temperature.

Raman optical time domain reflectometry (ROTDR) and Brillouin optical time domain reflectometry (BOTDR) have both been employed for distributed sensing applications over the past few years. Their operation mechanisms are based on the nonlinearities of optical fibers, where additional spectral components can be generated. These additional spectral components are affected by external environmental parameters. Thus, evaluating the spectral content in an appropriate way can determine the changes in external measurands.

ROTDR is based on the Raman scattering phenomena, where both anti-Stokes components and Stokes components are generated. The intensity ratio between the Stokes and anti-Stokes components can provide temperature information at any given point along the fiber, as the fiber itself is the sensing medium. Since the amplitude of the Stokes components is less dependent on temperature, ROTDR is only capable of measuring temperature rather than strain, but it has a temperature resolution of 0.2 °C (0.36 °F).

In BOTDR, light launched into an optical fiber link is partially scattered back toward the source based on the Brillouin scattering phenomena. The frequency of the scattered light is dependent on both the temperature and strain applied to the optical fiber, which allows BOTDR to measure both of these parameters.

36.3.3 Fiber-Optic Sensing Applications

Fiber-optic sensing can offer lots of benefits for various aerospace applications including harsh environments and gas turbine engines [36.52–87]. Turbine engines under development are extensively evaluated in the test cell environment where both their improvements in propulsion efficiency as well as the integrity of their structural components can be verified. The safe operation of these test cells relies on having established and dependable test cell instrumentation capabilities. The Propulsion Instrumentation Working Group (PIWG) has been established to lead the development of advanced measurement capabilities for the test cell environment [36.88]. PIWG has established a prioritized list of technologies that would be the most beneficial to the needs of the test cell instrumentation community. This ranking was done by surveying the original engine manufacturers (OEMs) and government agencies involved in turbine engine testing. Their prioritized ranking of technologies beneficial to their needs is shown in Table 36.1. The technology areas listed below could benefit from fiber-optic sensing technologies to meet anticipated instrumentation needs: hot section dynamic strain measurements (e.g., nonintrusive stress measurement system; NSMS), surface temperature mapping, high-temperature dynamic pressure, blade tip deflection, planar gas-path temperature, blade tip clearance, particulate emissions, crack detection, heat flux, gas species emissions, high-temperature static strain, acoustic emissions, thermal barrier coating (TBC) health, bearing health and other applications.

Table 36.1 Prioritized ranking of technologies beneficial to PIWG with rankings from OEMs

Hot section technology	OEM Average
Flow path temperature measurements—planar gas path	9.00
Tip clearance measurements	7.67
Metal surface temperature measurements—airfoil surface mapping	7.67
Blade tip timing NSMS in hot sections	7.00
Dynamic pressure measurements	5.67
TBC surface temperature measurements	5.67
Single-point stress and strain measurements	5.67
Wires and interconnect improvements (enabler)	5.67
Hot section component degradation	5.00
Vibration measurements	4.33
Emission species measurements	4.33
Acoustic emissions—noise	1.00

Optical fiber sensors can be beneficial for these application because they:

1. have been proven to be robust in aerospace deployment for various types of measurements,
2. have been proven to be robust in the high-temperature, high-pressure, high-shock environments of offshore oil wells, for example,

3. are electromagnetic noise free (they are both immune to electromagnetic interference (EMI) and do not generate EMI),
4. are electrically passive and safe in explosive environments,
5. allow transmission of sensed information over long distances and through difficult-to-access regions (e.g., through small holes, down oil wells, etc.),
6. are small enough to be integrated into smart materials,
7. have minimal mass,
8. are capable of being installed with less labor than comparable electronic sensors, and
9. can provide high data rates.

One of the most promising techniques for optically measuring structural temperatures in the very difficult environment of the turbine engine hot section is a technique known as thermographic phosphors.

Although thermographic phosphors utilize optical fibers in order to make their measurements, this concept is more an optical measurement technique than what would be considered strictly a fiber-optic measurement technique. Thermographic phosphors have demonstrated significant potential for determining the surface temperatures on several ceramic matrix composites (CMC) and thermal barrier coated (TBC) components. The ability to make temperature measurements on TBC components is especially important since the use of TBC allows gas turbine structural components to survive higher temperatures in the hot section of engines, while retaining acceptable component lifetimes. Because of the high levels of reflected radiation associated with TBCs and ceramic components, pyrometry, which measures the thermal black-body radiation emitted by a component to determine its temperature, is prone to extremely high error levels, which severely limits its use in the turbine hot section. Thermographic phosphors are rare-earth-doped ceramic substances that fluoresce, or emit light, when exposed to short-wavelength (typically ultraviolet) light. Because they are ceramic substances, they adhere very well to both CMCs and TBCs. Certain characteristics of the emitted light change with temperature, including color spectra, brightness, and fluorescence duration. The latter parameter is most commonly used for temperature measurement. The basic components needed to make this type of measurement consist of the phosphor coating and a method to securely apply it, a data acquisition system, an excitation source to excite the phosphor to fluoresce, an optical probe that provides illumination of the coating and collection of the fluorescence, and a data analysis protocol. Typically, a short-wavelength short-duration light emitting diode (LED) or pulsed laser source is used as the excitation source to illuminate the phosphor coating, which in turn luminesces visibly in distinct lines of fluorescence. Once the illumination ceases, the luminescence will persist for a short time, steadily decreasing. The time required for the brightness to decrease to $1/e$ of its original level is known as the decay time or lifetime. Decay time measurements are made by exciting the phosphor with a pulse of ultraviolet radiation and measuring the decay time of one of the lines of fluorescence through a narrow-band filter to eliminate as much of the background radiation as possible. Fairly detailed schematics showing how luminescence decay measurements would be made in turbine engine tests are shown in Figs. 36.11 and 36.12, while a more general overview of the phosphor thermometry method is shown in Fig. 36.13. The

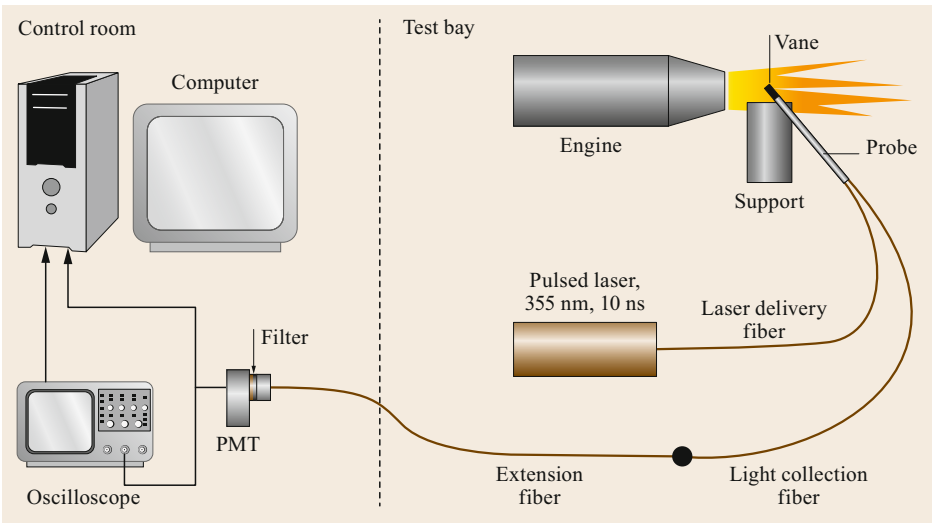


Fig. 36.11 Schematic of experimental luminescence lifetime thermometry for high-temperature engine components using optical fibers (after [36.89])

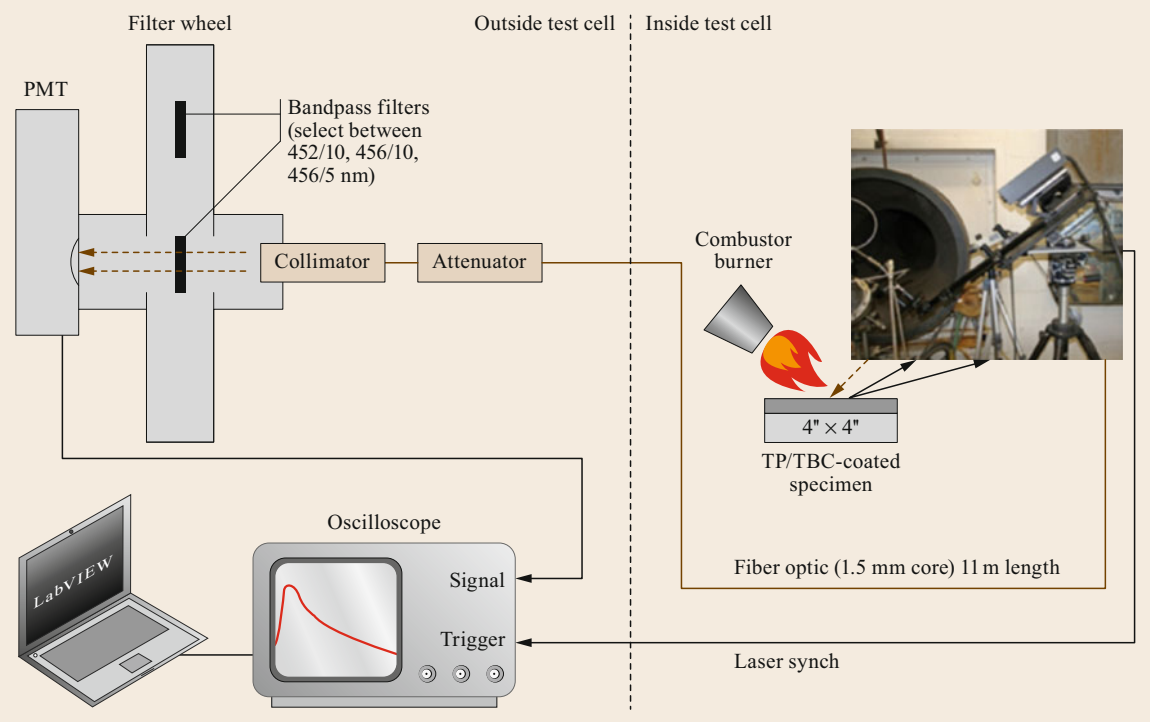


Fig. 36.12 Luminescence decay measurement setup for gas turbine engines using optical fibers (after [36.90])

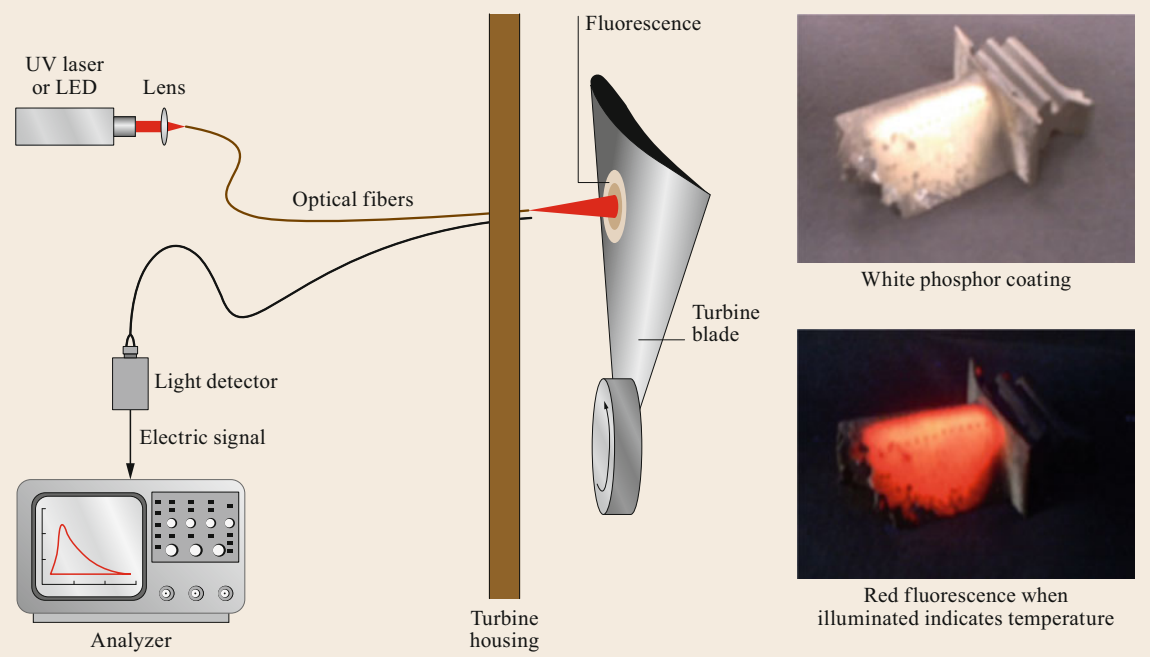


Fig. 36.13 Application of thermometry methodology for turbine engine blades using optical fibers (after [36.91])

decay time method of making temperature measurements has been successfully used in a variety of other structural measurement applications. Unfortunately, an attempt to make temperature measurements on first-stage turbine vanes yielded marginal results, indicating the need for additional development to enable this technique to be successfully demonstrated in the environment of a turbine engine.

A second, less commonly used method of temperature detection is based on measuring the intensity ratios of two separate lines of fluorescence, where the change in coating temperature is derived from the change of the phosphorescence spectrum. This method enables surface temperature distributions to be determined. In addition to making surface measurements, the intensity ratio method has the advantage that fibers that have become dirty or polluted have little effect on the measurement as it compares ratios between emission lines. The major disadvantage of this method is that it is a much more complicated test method to incorporate into gas turbine engines.

While the previously mentioned methods are focused on temperature detection, the inclusion of thermographic phosphor dopants into the physical structure of the TBC, especially if done in a layered fashion where different dopants are located at varied depths in the TBC, can essentially make the entire surface of the TBC a sensor. This *full-surface TBC sensor* would have the ability to detect a variety of parameters, such as heat flux, spallation, TBC erosion, or other damage or aging mechanisms the TBC may be experiencing.

36.3.4 Ultrahigh-Temperature Design

Here we review the ultrahigh-temperature (UHT) design consideration for optical fibers and FBGs.

Probes for high-temperature operation can be developed using silica fiber (for operation to around 1000 °C), pure silica core fiber (with glass transition temperature $\approx$ 1200 °C) for resistance to slightly higher temperature and sapphire fiber for operation to over 1700 °C (sapphire's melting point is approximately 2040 °C).

Single-crystal sapphire fibers were originally developed by *Fejer*, *Byer* and collaborators at Stanford in the 1980s [36.58, 59]. Multimode fibers are good choices for blade tip clearance measurement according to the results reported in [36.62]. Nevertheless, single-mode (or near single-mode) propagation in multimode sapphire fibers has been demonstrated with subpicosecond pulses [36.68], tapered fiber excitation [36.72], and hollow sapphire fibers. Using FBGs to detect subpicosecond pulses is still at the research level; more developments in photonics and electronics are needed to make it compatible for FBG reading at subpicosecond level.

Alternative fiber-like structures, such as those composed of sapphire, are considered attractive candidates for sensing applications at still higher temperatures (over 1500 °C) because of the high temperature survivability of sapphire. So far, such devices could only be employed as point sensors or, in the case of long period gratings, as broadband sensors operable only in transmission mode. Authors in [36.72, 73] have demonstrated FBGs written in multimode sapphire fibers with single-mode interrogation tests at 1288 °C (2350 °F). To achieve single-mode operation, by the coupling of the fundamental mode from a single-mode fiber into the fundamental mode in a multimode sapphire fiber, Mihailov used a mode-matching taper. Authors from Leibniz-Institute for Photonics Technologies (IPHT had earlier tested FBGs to 1745 °C (3173 °F) [36.53].

FBGs created with nonphotosensitive methods in fiber that is specifically designed for high-temperature, high-vibration environments have been proven to be capable sensors in environments such as gas turbine engines.

In successive rounds of PT6 testing sensors were installed both on the casing and in the exhaust [36.55, 56]. Appropriately packaged FBGs are robust to high shock, vibration, acoustic loading and temperature (to over 1000 °C) in a range of high-temperature applications. High-temperature FBGs have been shown to survive several months of use on gas turbines to 800 °C ($\approx$ 1470 °F) in [36.81]. NASA's GRC group has worked with the Armstrong Flight Research Center (AFRC) to perform testing of sensors calibrated to 800 °C in a turbine engine exhaust, with installation of the sensors in an emission sensing system (ESS) rig located behind a Pratt & Whitney F117 turbofan engine [36.82].

36.3.5 Rotating Component Vibration and Clearance Measurement Using Optical Probes

Two critical items for turbine engineering are blade tip timing (BTT) and tip clearance (TC), since the accurate measurement and optimization of these parameters leads to more reliable, effective, and secure engines [36.84]. BTT is a technique for measuring blade vibration that uses the differences between real and theoretical blade tip arrival times to calculate the vibration amplitude of the blade. Over the past few decades plentiful research has been developed and published on this technique (examples can be found in [36.62–66, 69, 75, 76, 78–80, 85]). Blade vibration measurement is crucial to optimize turbine operation and the prevention of fatigue-related blade failure [36.60]. Blade vibrations can occur due to several causes. For example, combustors or stators can produce synchronous responses in

turbine blades, and even irregularities in the casing or in the intake geometry can produce pressure variations that lead to synchronous dynamic responses in the rotor blades. In addition, rotating stall or adverse flow-blade interaction with negative aerodamping can induce nonsynchronous responses, such as flutter, which consists of a self-sustained aerodynamic instability. In an effort to enable us to predict the lifetimes of blades and to prevent damage that could lead to huge repair costs or even to engine destruction and loss of aircraft, a low-cost, lightweight, and effective blade vibration detection system was determined to be a necessity and has been developed. The BTT technique based on optical sensing discussed here fulfills these requirements by detecting the dynamic response of all blades on a stage.

Blade tip clearance (TC) is defined as the distance between the blade tip and the engine casing at that location, and its typical values vary depending on the size of the engine and the location of the blade within the engine, but are almost always less than 3 mm. This gap between the blade tip and the case is a significant factor on which engine efficiency depends, as an increase in TC decreases engine efficiency. A high TC allows an amount of air to flow back around the tip without generating useful work, whereas a lack of clearance can put engine integrity at risk, as blade tip rubs on the case can cause serious problems. TC reduction in gas turbine engines provides considerable benefits such as reductions in fuel consumption, aircraft noise, and emissions, while simultaneously providing additional environmental advantages [36.71, 83]. To get a quantitative understanding of these benefits, it is noteworthy to mention a 0.25 mm TC reduction typically provides a reduction of 1% in specific fuel consumption, and of 10 °C in exhaust gas temperature [36.83]. As a result, the engine can work at lower temperatures and the life of its components will be increased, thereby reducing engine lifecycle cost.

For the measurement of TC, traditional methods include the use of capacitive, eddy current and discharging probe sensors. Capacitive sensors have seen widespread usage, are simple and inexpensive, but they have poor frequency response and require iron blades. Eddy current probes provide noncontact measurements, but the magnetic disturbances of the turbine engine can interfere with their output. Additionally, they need to be calibrated in advance, because they are highly dependent on the tip shape and temperature [36.54]. Discharging probes, in much the same way as eddy current sensors, require electrically conducting blades and are only able to measure the shortest clearance. On the other hand, optical fiber-based sensors can provide lightweight, small size, noncontact measurements using simple instrumentation. Fiber-optic sensors also provide high sensitivity, resolution and bandwidth, an insensitivity to electromagnetic interference and the ability to make measurements of every blade.

36.3.6 Blade Tip Clearance (TC) Measurement Using Optical Fibers

The blade TC is difficult to measure in real time because:

1. The environmental temperature of the sensor probe is very high (more than 1100 °C).
2. There are limitations for the probe size.
3. There are limitations for the data acquisition process of a single measurement due to high velocity of the blade tip along the circumference.
4. A very high resolution (order of 0.025 mm) is required in a measurement range of several millimeters.

Optical sensing techniques are good candidates to overcome these issues [36.62].

The laser ranging (i.e., range finder) system based on the phase shifting method can be used to measure TC, as it is a noncontact measurement method and can withstand harsh operating environments [36.62, 63]. As previously reported in [36.77], this method has the potential to provide a TC measurement capability of about 10 μm, which is suitable for blade tip clearance measurement needs. A fiber-coupled sensor probe isolates the optical receiver from the harsh environment within the turbine engine. A multimode fiber can be used to collect the light reflected by the blade, since the signal-to-noise ratio (SNR) of the multimode fiber is acceptable (in comparison with single-mode fibers) for highly accurate tip clearance measurements.

A step-index multimode fiber can be utilized as the receiving fiber since the small diameter of the single-mode fiber makes it very difficult to couple the reflected light. The SNR is too low when a single-mode fiber is used. As different modes of the multimode fiber have different optical paths, when different modes are excited there will be an extra phase shifting of the modulated light, which will cause some measurement error.

36.3.7 Blade Tip Timing (BTT) Measurement Using Optical Fibers

BTT is a noncontact measurement technique and it is based on the measurement of the arrival times measured by sensors at several locations around the circumference of the stage of interest of all of the blades on that stage [36.64, 66]. A BTT measurement system consists of three main elements:

(i) The acquisition of raw blade arrival time data by a number of stationary probes placed in the casing,
(ii) The derivation of characteristic vibration parameters, such as blade displacement, velocity, and acceleration from the measured data
(iii) The processing of the characteristic parameters to describe the vibration properties of the bladed-disk assembly.

If the blades do not vibrate, their actual arrival times are identical to their theoretical arrival times (which can be obtained from a simple relationship involving the number of blades n, the rotation frequency ω, and the blade radius R). However, if the blades do experience vibration, their arrival times will precede or succeed the theoretical nonvibrating arrival times. The difference between the theoretical and the actual arrival time is related to the individual blade deflections and is used in the postprocessing section of the system to determine the frequency and amplitude of vibration, assuming an adequate number of probes are used around the circumference of the stage [36.64]. Furthermore, by placing probes at different circumferential positions, a sine wave model can be fitted to the blade deflections measured at each position. From these sinusoidal fits, the frequencies and amplitudes of the vibrations can be obtained.

As it is described in [36.64], it is possible to perform a *traveling wave analysis* in order to obtain the average amplitude of the blade tips at a particular nodal diameter (ND). The minimum value of ND is 0, which is obtained when all the blades vibrate with the same phase, whereas ND reaches its maximum value (half of the number of blades) when each blade is out of phase with its neighbor. This traveling wave analysis can be used for monitoring the integrity of the blades against flutter, crack propagation, domestic object damage and/or foreign object damage.

The traveling wave mode is the vibration condition of the blades. In a bladed disk system, the blades vibrate at the same amplitude but with a phase lag between them. This is known as the interblade phase angle (Ibpa). This phase lag between the blades is related to the nodal diameter (ND) of the bladed disk mode according to $\mathrm{ND} = 2\pi n/\mathrm{Ibpa}$, where n is the number of blades. Therefore, the frequency detected by the probe (f_{probe}) is the frequency of the blade (f_{blade}) plus the nodal diameter multiplied by the rotation frequency (ω). Here, the engine order (EO) is defined as $\mathrm{EO} = f_{\mathrm{blade}}/\omega$ [36.64].

An optical probe placed in the casing that measures the frequency of a rotating part senses both the frequency of the blade and the phase lag that a blade has with its neighbor, so that it is not possible to discriminate EO from ND, because the probe detects the arrival time of each blade as a superposition of both terms. The traveling wave analysis is mainly used for nonsynchronous responses, such as flutter, where EO plus ND has nonzero values; for synchronous responses it is zero, since the excitation and the response are in phase and, therefore, synchronous responses are more difficult to detect with the traveling wave analysis [36.64].

With a single-probe application, it is not possible to carry out an analysis of any synchronous dynamic response frequencies or amplitudes of vibration. Nevertheless, it is still possible to perform a traveling wave analysis in order to obtain the average nonsynchronous amplitude of the blades. This traveling wave analysis can be used to monitor the integrity of the blades against flutter, crack propagation, domestic object damage and/or foreign object damage. Synchronous responses are more difficult to detect and require more probes than nonsynchronous responses.

The arrival time of each blade can be estimated by calculating the second derivative of the raw signal from the optical probe (i.e., voltage signals from photodetectors). This derivative provides the change in concavity/convexity of the raw signal. By choosing a threshold value for the second derivative of the raw signal, the blade arrival time can be obtained for every blade. From these arrival times we can obtain the deflection or deviation of each blade from its theoretical nonvibrating equilibrium position. This deviation provides us with useful information for blade health monitoring to predict possible damage to the blades. By plotting the deviations in real time, flutter or crack propagations can be detected as the deviation of a given blade increases with time and gets close to predefined values of concern.

With the determination of the deflection levels of the blades a fast Fourier transform can be performed to obtain the traveling wave spectrum of the system, which provides the average tip amplitude of all the blades. The traveling wave spectrum gives an average value for all the blade vibration amplitudes at a certain nodal diameter and can also be used to monitor the system in real time, in order to assure that the average blade tip amplitude does not exceed any predefined maximum value. A recently developed technique is presented in [36.67], which is a nonuniformly undersampled BTT signal reconstruction method for blade vibration monitoring.

36.3.8 Exhaust Gas Temperature (EGT) Measurement

The exhaust gas temperature (EGT) is defined as the gas temperature at the exit of the turbine; the sensors used to measure this parameter are justifiably considered to

be among the most vulnerable elements of turbine engine instrumentation. EGT measurement is needed for optimizing fuel economy, and for the diagnosis and prognosis of potential structural issues. Turbine blade temperature is a good indicator for normal life consumption of turbine blades [36.92]. Currently, direct sensor measurements made on turbine modules are seriously limited due to the UHT environment. EGT sensors located downstream from the highest temperature sections provide a means to infer the approximate temperatures seen by the turbine blades/disks. Unfortunately, these sensors, which themselves are subject to frequent failure, provide a fairly inaccurate indication of the actual metal temperature profiles. Accurate EGT measurements with durable sensors are a significant area of need for high-performance military turbine engines where the margin between hot section operating temperature and turbine blade material capability is constantly shrinking [36.92].

EGT and other turbine temperature sensors are susceptible to degradation due to high-temperature oxidation, erosion and contaminant intrusion into probes and wiring harnesses. Thermocouples are easily affected by noise, electromagnetic interference, and/or other environmental factors. Military and commercial field experience indicates that gas path thermocouple removals affect aircraft availability and add maintenance time. In addition, the adaptive engines of the future are driving the control system to outperform legacy design and cause higher temperatures.

Existing technologies for measuring EGT typically implement high-temperature capability with thermocouples but extension to higher temperatures is not guaranteed. Other technologies that have been investigated include thin-film thermocouples, pyrometers, spectroscopy, and radioactive isotope-based sensors [36.92]. They are not mature, accurate or cost-effective for engine implementation. In addition to being not mature, multicolor pyrometers are also complex, emissivity dependent, and expensive [36.92]. With alternatives that use fiber-optic technologies [36.93–98] to measure exhaust gas effects, measurement of significantly higher temperatures is possible for EGT sensing.

Fiber-optic and laser-based sensing systems are envisioned [36.92–98] to be used for EGT measurement and to accurately assess the condition and life usage of the turbine engine hot section, on a blade-by-blade basis. To understand the extreme environmental conditions experienced in turbine engines, there is a need for robust sensors to obtain high-precision EGT measurements. Development of new technologies for reliable EGT measurement using optical fibers in future high-performance turbine engines is needed. The control and health management of modern turbine engines depends on sensing a wide variety of quantities throughout the engine, (e.g., temperatures, pressures, and vibration, etc.) with different redundancy, reliability, and accuracy requirements; reliable EGT measurement is one of the challenges for gas turbine engine control and health management improvement.

An optical sensor system extends gas temperature measurement capability in turbine engines beyond the present generation of EGT sensing technologies. The sensing element, which consists of Bragg gratings embedded inside a sapphire lightguide, is capable of operating in UHT conditions. The emissive insert generates an optical signal as a function of temperature. An optically averaged EGT measurement system can be developed by combining the optical signals from multiple individual sensing probes at a single detector assembly.

36.3.9 Practical Harsh Environment Considerations

Many of the fiber-optic applications involve integration into systems that operate in harsh environments, including extreme low to high temperatures, shock, vibration, radiation, corrosive conditions, chemically aggressive environments, high electromagnetic, and high radio-frequency interference and pressure. Fiber-optic technologies have demonstrated the ability to overcome the issues associated with the harsh environments and provide situational awareness to platforms operating in these environments.

Actual operating environments usually include a combination of several of the above issues. Optical fiber systems deployed in a harsh environment require appropriate design of both the optical fibers and cabling to guarantee long-term and reliable operation. Optical fiber development for harsh environment operation includes both fiber material (e.g., glass) and coating materials development. Glass development activities include changes to the core and cladding composition and refractive index profile design changes. Coating technology development includes for example material changes to improve the thermal characteristics.

The United States Department of Defense (US DoD) has invested in performance enhancements and life extension for various aircraft systems (e.g., F-15, F-16, F-22, F-35 and others), which affords an opportunity to introduce fiber-optic communication and sensing technology into future aircraft platforms via engine replacements that contribute to effectiveness of mission, safety and protection of the pilot. The aircraft and engine sensing architecture can provide the capability for distributed sensing with improved predictive maintenance. This accelerates the development and ability to deploy military capabilities for use by the Air Force.

One of the most challenging examples of harsh environment measurement needs is for gas turbine engines. The use of smart sensors for jet engine control is currently limited by the availability of mature high-temperature electronic components that can withstand the engine operating environment. One of the most challenging turbine-engine sensor requirements is measuring the gas temperature as it exits the combustor and enters the turbine. As engine temperatures have increased, the durability and performance limit of engine temperature sensors remain a challenge. Thermocouples are commonly used for engine temperature sensing, but their lifetimes are significantly decreased in this high-temperature environment. As a result, these sensors have been moved downstream to a cooler operating condition. The turbine inlet temperature is then estimated using an empirically derived relationship with resulting inaccuracies. Fiber-optic-based sensors have been demonstrated to be capable of functioning in harsh environments (e.g., in temperatures of 1000 °C and higher [36.99, 100] and in presence of vibration [36.101, 102]). These achievements potentially open up a myriad of new applications and opportunities for robust, highly capable sensors. As this technology advances, smart sensors and fiber-optic buses will increasingly appear in many different engine or avionics applications.

36.3.10 Sustainment

The sustainment problem with the existing systems is reliability, vulnerability, weight, and cost. The benefit of distributed fiber-optic sensor systems for avionics and gas turbine engines, from a sustainment perspective, is increased modularity, improved sensor data quality, and decreased wiring, which translates into reduced repair and maintenance costs and time, fewer engine inspections, and reduced operational risk. Distributed fiber-optic sensor system support and test equipment can be developed with less cost compared to the existing systems. Some of the sustainment characteristics of fiber-optic systems are as follows:

- Diagnostics and prognostics of fiber-optic system will be faster and easier. A repair process manual can be developed for fiber-optic sensor systems in the field.
- Optical fiber is made of glass, which is chemically inert and does not degrade with time, hence there is no need for signal reduction compensation due to exposure to environment—light attenuation is only 0.1 dB/km, which is negligible due to finite dimensions within the airborne systems.
- Sensor calibration can be done during the manufacturing process, and since there are no electronics within the sensor itself, there is no need for periodic calibration. Sensor interrogator electronics have self-calibration capabilities.
- Optical fiber can be ruggedized for example by using metal sleeves for harsh environments. There is still weight benefit from optical fibers after enclosure, since signals of multiple sensors are transmitted through one fiber.
- Specialized photonics equipment (including hardware and software) can be used for maintenance and calibration.
- Pig-tail connections in fiber-optic systems provide capability for localized optical fiber maintenance, and it is easier to repair optical fibers on the ground. Unlike hard-wire degradation due to harsh conditions, aging, and effects of vibration, shock, and radiation, optical fibers are less vulnerable to these environmental issues.
- Usage of fiber-optic systems does not require any high-temperature electronics, unless the data needs to be measured right at the site. Because of the small size and low transmission loss in optical fibers, sensors can be placed in high-temperature harsh environments at a distance away from the interrogator.
- Software maintenance for fiber-optic sensor systems is easy and a process for software maintenance can be provided.
- Condition-based maintenance (CBM) is improved using optical fibers since more data will be available from more sensors on a single fiber.

36.4 Conclusion

This chapter provided an overview of avionics optical fiber communication and sensing systems for use in vehicles. Optical fibers are the most common means of optical communication. In many applications they are more cost effective than communication via copper wiring. Fiber optics is an enabling technology for lots of advanced sensing techniques envisioned for transition to aerospace platforms in the near future. One important extension of fiber-optic communication technology for aerospace applications is the FBL approach for aircraft control systems. Inherent characteristics of FBL like light weight, compact size, large bandwidth, and

EMI-immunity make it an ideal technology for future flight control systems. FBL control systems offer inherent resistance to the new aircraft generation working within more hostile military environments.

Aircraft are loaded with sensors that collect a tremendous amount of data that they must move through their bus structure, to process it or move it to a recording device for later download. For military and commercial aviation applications, optical fiber buses can handle 40 Gb/s with less loss and longer distances in comparison to wire cables. Although fiber is easier to deal with for high speeds, and it has the overall benefit of weight-per-foot advantage, there are still complications including limited bend radius, and the trouble to deal with delicate interconnections in harsh environments with dust and/or foreign object debris.

In addition to their value in communication, optical fibers have the capability to provide information about the environment they are exposed to. There are various types of fiber-optic sensors that can provide detailed data on parameters such as temperature, strain, pressure, vibration, and acoustic emissions. Accurate knowledge of these parameters is vital to the system design and safe operation of various components, which must often operate in high temperature or mechanically hostile environments.

In summary, the expected benefits of fiber-optic communication and sensing for avionics applications are as follows:

- The fiber-optic network would provide a high-speed backbone that can support the addition or subtraction of distributed sensors and subsystems (in a modular plug-and-play design architecture this can be done without redesign of the FADEC).
- Fibers optics play an essential role in enabling new, more capable measurement techniques to be incorporated into future engine testing.
- Benefits include a system that is lightweight, highly reliable, modular, scalable, capable of high data rates and large capacity, cost effective and immune to radiation/interference.
- With distributed fiber-optic sensor systems a large number of sensors can be used, which generates a high volume of data for improving the diagnostics and prognostics capability for jet engine systems.
- Thanks to shorter downtime and increased platform availability due to fiber-optic systems, predictive maintenance also reduces the total acquisition development and sustainment cost of the engine.
- Since multiple sensors can exist on a single optical fiber, there is only one streamline of sensor maintenance for all of these fiber sensors, which decreases the cost and complexity in shop floor processes.
- Fiber-optic sensors can be used in high-temperature components of the engine (such as combustor), as opposed to the current sensors, which are not useful in very high-temperature engine areas.
- Technology will be applicable to various systems including propulsion systems for advanced tactical, rotary wing and commercial aircraft.
- Network ability to integrate legacy and smart sensors.
- Network ability to overcome and reroute around nodal failures.
- Cost reductions in instrumentation system complexity and installation labor.
- Improvements in engine sensor scalability and data quality.
- Improved engine test schedules and accelerated time to market.
- Increase in fan efficiency, since measuring fan inlet and exit temperatures is used in the calculation of fan efficiency.
- Improving aircraft and engine prognostics and health management (PHM) systems by providing a large volume of data using a large number of sensors all over the engine especially in high-temperature areas.
- Reduced maintenance costs with built-in diagnostics at each smart module, improved fault isolation (for mission readiness), fewer parts obsolescence issues.
- Reduced control system (FADEC and cable) weight.
- Standard components interface to promote competition in supply, more advanced technology options (e.g., optical) and cost control, affordable modular upgrade/reuse.
- Smaller core packaging.
- Enhanced hot section sensing and actuation.

References

36.1 R. Paschotta: *Encyclopedia of Laser Physics and Technology* (Wiley-VCH, Weinheim 2008), entry: Multi-mode fibers

36.2 New World Encyclopedia: Optical fiber – Principle of operation, multimode fiber, http://www.newworldencyclopedia.org/entry/Optical_fiber (2018)

36.3 S.E. Baldini, D.J. Tubbs, W.A. Stange: Embedding fiber optic sensors in titanium matrix composites, Proc. SPIE (1990), https://doi.org/10.1117/12.24856

36.4 D. Botez, G.J. Herskowitz: Components for optical communications systems: A review, Proc. IEEE **68**(6), 689–731 (1980)

36.5 M. Azadeh: Fiber optic communications: A review. In: *Fiber Optics Engineering* (Springer, Berlin, Heidelberg 2009)

36.6 D. Culley, A. Behbahani: Communication needs assessment for distributed turbine engine control. In: *44th AIAA/ASME/SAE/ASEE Jt. Propuls. Conf.* (2008), https://doi.org/10.2514/6.2008-5281

36.7 R.P.G. Collinson: *Introduction to Avionics Systems*, 3rd edn. (Springer, Berlin, Heidelberg 2011)

36.8 A. Behbahani, D. Culley, B.J. Smith, C. Darouse, R. Millar, B. Wood, J. Krodel, S. Carpenter, B. Mailander, T. Mahoney, R. Quinn, C. Bluish, B. Hegwood, G. Battestin, W. Roney, W. Rhoden, B. Storey: Status vision and challenges of an intelligent distributed engine control architecture. In: *Aerosp. Technol. Conf. Expo.* (2007), https://doi.org/10.4271/2007-01-3859

36.9 D. Culley, T.R. Saus: Concepts for distributed engine control. In: *43rd AIAA/ASME/SAE/ASEE Jt. Propuls. Conf.* (2007), https://doi.org/10.2514/6.2007-5709

36.10 M. Pakmehr, M. Dhingra, N. Fitzgerald, J. Paduano, M. Wolf, E. Feron, A. Behbahani: Distributed architectures integrated with high-temperature electronics for engine monitoring and control. In: *47th AIAA/ASME/SAE/ASEE Jt. Propuls. Conf.* (2011), https://doi.org/10.2514/6.2011-6148

36.11 M. Pakmehr, M. Mounier, N. Fitzgerald, G. Kiwada, J. Paduano, E. Feron, A. Behbahani: Distributed control of turbofan engines. In: *45th AIAA/ASME/SAE/ASEE Jt. Propuls. Conf.* (2009)

36.12 J. Rhea: Fly by light, http://www.airforcemag.com/MagazineArchive/Pages/1988/March%201988/0388light.aspx (1988)

36.13 G.T. Seng: Overview of NASA research in fiber optics for aircraft control, NASA-TM-100919 (1988)

36.14 J.R. Todd: *Fly-by-Light Technology Development Plan*, NASA Contractor Report 181954 (McDonnel Douglas, Long Beach 1990)

36.15 J.R. Todd, J.A. Hay, T. Dinh: Integrating fly-by-light/power-by-wire flight control systems on transport aircraft. In: *12th AIAA/IEEE Digit. Avion. Syst. Conf.* (1993) pp. 457–462

36.16 J.R. Todd: Fly-by-light flight control system development for transport aircraft. In: *15th AIAA/IEEE Digit. Avion. Syst. Conf.* (1996) pp. 153–158

36.17 D.W. Seal, T.L. Weaver, B.L. Kessler, C.A. Bedoya, R.E. Mattes: *Fiber Optic Control System Integration for Advanced Aircraft: Electro-Optic and Sensor Fabrication, Integration, and Environmental Testing for Flight Control System*, NASA Contractor Report 191194 (Southwest Res. Inst., San Antonio 1994)

36.18 D.J. Halski: Fly-by-light flight control system, Proc. SPIE (1995), https://doi.org/10.1117/12.210092

36.19 D. J. Halski: FLASH fly-by-light flight control demonstration results overview, SAE Technical Paper 952044 (1995)

36.20 B.W. Harris: *Fiber Optics for Flight Control Systems*, M.Sc. Thesis (University of Dayton, Dayton 2014)

36.21 International Air Transport Association: IATA technology roadmap technical annex, https://www.iata.org/whatwedo/environment/Documents/technology-roadmap-annex.pdf (2009)

36.22 Aviation Week Network: First tranche 3 Typhoon readied for flight, http://aviationweek.com/awin/first-tranche-3-typhoon-readied-flight (2013)

36.23 S. Pope: Gulfstream testing fly-by-light control, http://www.ainonline.com/aviationnews/aviation-international-news/2008-04-01/gulfstream-testing-fly-light-controls (2008)

36.24 P.G. Hamel: *In-Flight Simulators and Fly-by-Wire/Light Demonstrators: A Historical Account of International Aeronautical Research* (Springer, Berlin, Heidelberg 2017)

36.25 M. Gestwa: ACT/FHS flying helicopter simulator, http://www.dlr.de/dlr/en/desktopdefault.aspx/tabid-10203/339_read-269#/gallery/135 (2011)

36.26 J. Levis, B. Sutterfield, R. Stevens: Fiber optic communication within the F-35 mission system. In: *IEEE Conf. Avion. Fiber-Opt. Photonics* (2006), https://doi.org/10.1109/AVFOP.2006.1707476

36.27 B. Mukherjee: *Optical WDM Networks* (Springer, New York 2006)

36.28 S.F. Habiby, M.J. Hackert: Ronia results: WDM-based optical networks in aircraft application. In: *IEEE Conf. Avion. Fiber-Opt. Photonics* (2008), https://doi.org/10.1109/AVFOP.2008.4653181

36.29 S.F. Habiby, R. Vaidyanathan: WDM optical backbone networks in aircraft applications: Networking challenges and standards progres. In: *IEEE Mil. Commun. Conf.* (2009), https://doi.org/10.1109/MILCOM.2009.5379942

36.30 S.F. Habiby, R. Dutt: Optical network architecture, technology and component challenges in aircraft network application. In: *IEEE Conf. Avion. Fiber-Opt. Photonics* (2009), https://doi.org/10.1109/AVFOP.2009.5342634

36.31 S.F. Habiby: Advances in WDM LAN standards development for aerospace applications. In: *IEEE Conf. Avion. Fiber-Opt. Photonics* (2006), https://doi.org/10.1109/AVFOP.2006.1707480

36.32 A.S. Glista, M.W. Beranek: Wavelength division multiplexed (WDM) optical technology solutions for next generation aerospace networks. In: *Digit. Avion. Syst. Conf.* (2003), https://doi.org/10.1109/DASC.2003.1245942

36.33 W.P. Krug: WDM for platforms: From systems architectures to systems-on-chips. In: *IEEE Conf. Avion. Fiber-Opt. Photonics* (2005), https://doi.org/10.1109/AVFOP.2005.1514129

36.34 L. Green, P.B. Ruffin, J.C. Holt, J.S. Hamilton, R.J. Turner, M.K. Gallman: A case for embedded optical communication. In: *IEEE Mil. Commun. Conf.* (2008), https://doi.org/10.1109/MILCOM.2008.4753046

36.35 J.W. Bretthauer, C.H. Chalfant, F.J. Orlando, E. Rezek, M. Sawyer: *Spaceborne Fiber Optic Data Bus (SFODB)* (NASA, Greenbelt 2008)

36.36 D.K. Hunter, E.D. Lowe: Wavelength use in optical WDM bus networks without wavelength conversion. In: *Opt. Netw. Appl.* (1998), https://doi.org/10.1364/ONA.1998.RBS4

36.37 A.J. McLaughlin, J.S. Aitchison, J.H. Marsh, R.D. Gardner, I. Andonovic, D.K. Hunter: Design and development of a multi-gigabit WDM network for use in the aerospace environment. In: *IEEE Aerosp. Conf. Proc.* (2000), https://doi.org/10.1109/AERO.2000.879840

36.38 D.G. Wick: Realizing distributed engine control subsystems through application of high-temperature intelligent engine sensors and control electronics, SAE Technical Paper 2000-01-1363 (2000)

36.39 G.W. Hunter, J.D. Wrbanek, R.S. Okojie, P.G. Neudeck, G.C. Fralick, L. Chen, J. Xu, G.M. Beheim: Development and application of high temperature sensors and electronics for propulsion application, Proc. SPIE (2006), https://doi.org/10.1117/12.668458

36.40 J.P. Moreaux: Data transmission system for aircraft, US Patent 6925088 (2005)

36.41 B. Harris, B. Tran: Fiber optic AFDX for flight control system. In: *IEEE Conf. Avion. Fiber-Opt. Photonics* (2012), https://doi.org/10.1109/AVFOP.2012.6344056

36.42 AIM Avionics Databus Solutions: AFDX tutorial training, http://www.afdx.com/pdf/AFDX_Training_October_2010_Full.pdf (2010)

36.43 A. Al Sheikh, O. Brun, M. Cheramy, P.-E. Hladik: Optimal design of virtual links in AFDX network, Real-Time Syst. **49**(3), 308–336 (2013)

36.44 P. Heise, I. Gaillardet, H. Rahman, V. Mannur: Avionics full duplex Ethernet and the time sensitive networking standard. In: *IEEE 802.1 Interim Meet., Pittsburgh* (2015)

36.45 Alliance for Telecommunications Industry Solutions: *ATIS Telecom Glossary* (ATIS, Washington 2012), available from: http://www.atis.org/glossary/definition.aspx?id=3516

36.46 A. von Moll, A. Behbahani: Comparison of communication architectures and network topologies for distributed propulsion controls. In: *59th Int. Instrum. Symp.* (2013), available from: http://www.dtic.mil/dtic/tr/fulltext/u2/a586909.pdf

36.47 J.Y. Choi: Evaluation of ethernet based control network used in the distributed control system. In: *Proc. 15th CISL Winter Worksh., Kushu* (2002)

36.48 M.G. Larsen: Evolution of a hierarchical ring bus network. In: *Int. Worksh. Manag. Requir. Knowl., Chicago* (1987), https://doi.org/10.1109/AFIPS.1987.47

36.49 E. Gutiérrez Fernández: *Management System for Unmanned Aircraft System*, M.Sc. Thesis (Universitat Politècnica de Catalunya, Barcelona 2010)

36.50 M. Huang, B. Bode: A performance comparison of tree and ring topologies in distributed system. In: *19th IEEE Int. Parallel Distrib. Process. Symp.* (2005), https://doi.org/10.1109/IPDPS.2005.57

36.51 G.A. Cranch, G.M.H. Flockhart, C.K. Kirkendall: Efficient large-scale multiplexing of fiber Bragg grating and fiber Fabry-Perot sensors for structural health monitoring applications, Proc. SPIE (2006), https://doi.org/10.1117/12.657416

36.52 J.R. Dunphy, G. Meltz, R.M. Elkow: Distributed strain sensing with a twin-core fiber optic sensor. In: *32nd Int. Instrum. Symp.* (1986) pp. 145–149

36.53 M. Busch, W. Ecke, I. Latka, D. Fischer, R. Willsch, H. Bartelt: Inscription and characterization of Bragg gratings in single-crystal sapphire optical fibres for high-temperature sensor applications, Meas. Sci. Technol. **20**, 115301 (2009)

36.54 S.Z. Cao, F.J. Duan, Y.G. Zhang: Measurement of rotating blade tip clearance with fibre-optic probe, J. Phys. Conf. Ser. **48**, 873–877 (2006)

36.55 M. Pakmehr, J. Costa, R.J. Black, A. Behbahani: A review of fiber optic technology for turbine engine instrumentation channel: Control, PHM, and test cell applications. In: *50th AIAA/ASME/SAE/ASEE Jt. Propuls. Conf.* (2014), https://doi.org/10.2514/6.2014-3534

36.56 J. Costa, R. Black, B. Moslehi: Advances in high temperature fiber optic sensors for turbine engine application, Proc. ISA **491**, 341 (2012)

36.57 Department of Defense: MIL-HDBK-1783B handbook: Engine structural integrity program (ENSIP), http://everyspec.com/MIL-HDBK/MIL-HDBK-1500-1799/MIL-HDBK-1783B_CHG-2_26101 (2004)

36.58 M. Fejer, J. Nightingale, G. Magel, R. Byer: Laser-heated miniature pedestal growth apparatus for single-crystal optical fiber, Rev. Sci. Instrum. **55**, 1791–1796 (1984)

36.59 M.M. Fejer: *Single Crystal Fibers: Growth Dynamics and Nonlinear Optical Interaction*, Dissertation (Ginzton Labs, Stanford 1986)

36.60 V. Georgiev, M. Holík, V. Kraus, A. Krutina, Z. Kubín, J. Liška, M. Poupa: The blade flutter measurement based on the blade tip timing method. In: *Proc. 15th WSEAS Int. Conf. Syst.* (2011) pp. 270–275

36.61 A. Ghoshal, D. Le, H.S. Kim: Technological assessment of high temperature sensing systems under extreme environment, Sens. Rev. **32**(1), 66–71 (2012)

36.62 H. Guo, F. Duan, G. Wu, J. Zhang: Blade tip clearance measurement of the turbine engines based on a multi-mode fiber coupled laser ranging system, Rev. Sci. Instrum. **85**(11), 115105 (2014)

36.63 B. Journet, G. Bazin, F. Bras: Conception of an adaptive laser range finder based on phase shift measurement. In: *Proc. IEEE Ind. Electron. Control Instrum.* (1996) pp. 784–789

36.64 I. García, J. Beloki, J. Zubia, G. Aldabaldetreku, M.A. Illarramendi, F. Jiménez: An optical fiber bundle sensor for tip clearance and tip timing measurements in a turbine rig, Sensors **13**, 7385–7398 (2013)

36.65 A. Vakhtin, S.-J. Chen, S. Massick: Optical probe for monitoring blade tip clearance. In: *47th AIAA Aerosp. Sci. Meet.* (2009), https://doi.org/10.2514/6.2009-507

36.66 S. Heath, M. Imregun: A survey of blade tip-timing measurement techniques for turbomachinery

vibration, J. Eng. Gas Turbine. Power **120**, 784–791 (1998)
36.67 Z. Hu, J. Lin, Z.-S. Chen, Y.-M. Yang, X.-J. Li: A non-uniformly under-sampled blade tip-timing signal reconstruction method for blade vibration monitoring, Sensors **15**(2), 2419–2437 (2015)
36.68 S. Jamison, R. McGowan, D. Grischkowsky: Single-mode waveguide propagation and reshaping of sub-ps terahertz pulses in sapphire fiber, Appl. Phys. Lett. **76**, 1987–1989 (2000)
36.69 T. Kawashima, H. Iinuma, T. Wakatsuki, N. Minagawa: Turbine blade vibration monitoring system. In: *37th Int. Gas Turbine Aeroengine Congr. Expo.* (1992), https://doi.org/10.1115/92-GT-159
36.70 A. Kurtz, A. Ned, A. Epstein: Ultra high temperature, miniature, SOI sensors for extreme environment. In: *IMAPS Int. HiTEC Conf.* (2004)
36.71 S.B. Lattime, B.M. Steinetz: Turbine engine clearance control systems: Current practices and future direction. In: *38th AIAA/ASME/SAE/ASEE Jt. Propuls. Conf.* (2002), https://doi.org/10.2514/6.2002-3790
36.72 S.J. Mihailov, D. Grobnic, C.W. Smelser: High-temperature multiparameter sensor based on sapphire fiber Bragg grating, Opt. Lett. **35**(16), 2810–2812 (2010)
36.73 S.J. Mihailov: Fiber Bragg grating sensors for harsh environment, Sensors **12**, 1898–1918 (2012)
36.74 A. Mita, H. Kazama: Simple Mechanism for Vibration Sensor Based on Fiber Bragg Grating. In: *Proc. 4th Worksh. Struct. Health Monit.* (2003) pp. 1468–1475
36.75 P.E. McCarthy, J.W. Thompson Jr.: *Development of a noninterference technique for measurement of turbine engine compressor blade stress* (US Air Force, Arnold 1980), available from: https://apps.dtic.mil/dtic/tr/fulltext/u2/a086170.pdf
36.76 W.C. Nieberding, J.L. Pollack: Optical detection of blade flutter. In: *Proc. Gas Turbine Conf. Prod. Show, Philadelphia* (1977)
36.77 A. Pesatori, M. Norgia, C. Svelto, M. Zucco, M. Stupka, A.D. Marchi: High-resolution mode-locked laser rangefinder with harmonic down-conversion, IEEE Trans. Instrum. Meas. **61**(5), 1536–1542 (2012)
36.78 H. Roth: Vibration measurements on turbomachine rotor blades with optical probes. In: *Proc. Jt. Fluids Eng. Gas Turbine Conf. Prod. Show, New Orleans* (1980) pp. 215–224
36.79 W.B. Watkins, W.W. Robinson, R.M. Chi: Noncontact engine blade vibration measurements and analysis. In: *21st Jt. Propuls. Conf.* (1985), https://doi.org/10.2514/6.1985-1473
36.80 W.B. Watkins, R.M. Chi: Noninterference blade-vibration measurement system for gas turbine engines, J. Propuls. Power **5**, 727–730 (1989)
36.81 M. Willsch, T. Bosselmann, P. Flohr, R. Kull, W. Ecke, I. Latka, D. Fischer, T. Thiel: Design of fiber optical high temperature sensors for gas turbine monitoring, Proc. SPIE (2009), https://doi.org/10.1117/12.835875
36.82 G. Adamovsky, J.R. Mackey, B.M. Floyd, K. Elam, M. Martinez: Development and performance verification of fiber optic temperature sensors in high temperature engine environments. In: *50th AIAA/ASME/SAE/ASEE Jt. Propuls. Conf.* (2014), https://doi.org/10.2514/6.2014-3922
36.83 M.W. Wiseman, T.-H. Guo: An investigation of life extending control techniques for gas turbine engines. In: *Proc. Am. Control Conf.* (2001), https://doi.org/10.1109/ACC.2001.946211
36.84 D.C. Ye, F.J. Duan, H.T. Guo, Y. Li, K. Wang: Turbine blade tip clearance measurement using a skewed dual-beam fiber optic sensor, Opt. Eng. **51**(8), 081514 (2012)
36.85 I.E. Zablotskii, Y.A. Korostelev, L.B. Sviblov: *Contactless Measuring of Vibrations in the Rotor Blades of Turbines* (Defense Technical Information Center, Fort Belvoir 1974)
36.86 M. Zielinski, G. Ziller: Optical blade vibration measurement at MTU. In: *AGARD Conf. Proc.* (1989) p. 31
36.87 M. Zielinski, G. Ziller: Noncontact vibration measurements on compressor rotor blades, Meas. Sci. Technol. **11**, 847–859 (2000)
36.88 The Propulsion Instrumentation Working Group: http://www.piwg.org
36.89 T.P. Jenkins, J.I. Eldridge, S.W. Allison, R.P. Howard, E.H. Jordan, D.E. Wolfe: An experimental investigation of luminescence lifetime thermometry for high temperature engine components using coatings of YAG:Dy and YAG:Tm. In: *59th Int. Instrum. Symp.* (2013)
36.90 J.I. Eldridge, T.P. Jenkings, S.W. Allison, G.S. Cruzen, J.J. Condevaux, J.R. Senk, A.D. Paul: Real-time thermographic-phosphor-based temperature measurements of thermal barrier coating surfaces subjected to a high-velocity combustor burner environment. In: *57th Int. Instrum. Symp.* (2011)
36.91 S.W. Allison: ISA 107.2 temperature measurement using thermographic phosphor. In: *Turbine Engine Technol. Symp.* (2012)
36.92 A. von Moll, A.R. Behbahani, G.C. Fralick, J.D. Wrbanek, G.W. Hunter: A review of exhaust gas temperature sensing techniques for modern turbine engine control. In: *50th AIAA/ASME/SAE/ASEE Jt. Propuls. Conf.* (2014), https://doi.org/10.2514/6.2014-3977
36.93 G. Tregay, P. Calabrese, M. Finney, K. Stukey: Durable fiber-optic sensor for gas temperature measurement in the hot section of turbine engine, Proc. SPIE (1994), https://doi.org/10.1117/12.188847
36.94 M. Willsch, T. Bosselmann, P. Flohr, R. Kull, W. Ecke, I. Latka, D. Fischer, T. Thiel: Design of fiber optical high temperature sensors for gas turbine monitoring, Proc. SPIE (2009), https://doi.org/10.1117/12.835875
36.95 H. Xia, D. Byrd, S. Dekate, B. Lee: High-density fiber optical sensor and instrumentation for gas turbine operation condition monitoring, J. Sens. (2013), https://doi.org/10.1155/2013/206738
36.96 D. Grobnic, S.J. Mihailov, C.W. Smelser, H. Ding: Sapphire fiber Bragg grating sensor made using

femosecond laser radiation for ultrahigh temperature application, IEEE Photonics Technol. Lett. **16**(11), 2505–2507 (2004)

36.97 A. Wang, H. Xiao, J. Wang, Z. Wang, W. Zhao, R.G. May: Self-calibrated Interferometric-intensity-based optical fiber sensor, J. Lightwave Technol. **19**(10), 1495–2001 (1501)

36.98 M. Gottlieb, G.B. Brandt: Fiber-optic temperature sensor based on internally generated thermal radiation, Appl. Opt. **20**(19), 3408–3414 (1981)

36.99 M. Pakmehr, J. Costa, G. Lu, A. Behbahani, K. Semega: Extreme environment sensor system development: Optical turbine inlet temperature (TIT) sensing for gas turbine engines. In: Proceedings of the AIAA Propulsion and Energy Forum and Exposition, Indianapolis, IN, Aug 2019. https://doi.org/10.2514/6.2019-4388

36.100 M. Pakmehr, J. Costa, G. Lu, A. Behbahani: Optical exhaust gas temperature (EGT) sensor and instrumentation for gas turbine engines. In: Proceedings of the NATO STO Meeting: Transitioning Gas Turbine Instrumentation from Test Cells to On-Vehicle Applications STO-MP-AVT-306, Athens, Greece, Feb 2019. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-AVT-306/MP-AVT-306-18.pdf

36.101 Y. Rodríguez García, J.M. Corres, J. Goicoechea: Vibration detection using optical fiber sensors. J. Sensors **2010**, 936487 (2010). https://doi.org/10.1155/2010/936487

36.102 X. Liu, B. Jin, Qing Bai, Y. Wang, D. Wang, Y. Wang: Distributed fiber-optic sensors for vibration detection. Sensors **16**, 1164 (2016). https://doi.org/10.3390/s16081164

Alireza Behbahani

Turbine Engine Division, Advanced Integrated Controls and Prognostic / Diagnostic Systems Group
Air Force Research Laboratory (AFRL)
Wright-Patterson Air Force Base, OH, USA
alireza.behbahani@us.af.mil

Dr Alireza (Al) Behbahani's careers spans over 39 years and includes highly-citedresearch contributions and significant professional service, industrial, teaching, andprogram management experience in control systems. He is Senior Aerospace Engineerand Group Leader for the Advanced Integrated Controls and Prognostic/-DiagnosticSystems Group at the Turbine Engine Division at the Air Force Research Laboratory.He has a BSEE from Utah State, and an MS and a PhD.

Mehrdad Pakmehr

optoXense, Inc.
San Ramon, CA, USA
mp@optoxense.com

Dr Mehrdad Pakmehr received his PhD from Georgia Institute of Technology at Atlanta in 2013. His main expertise is in flight dynamics and control. He is a cofounder of optoXense Inc. and as VP Research he works on control and instrumentation of turbine engines, development of optical sensing systems, and expanding fiber optic sensing for health monitoring and aerospace applications.

William A. Stange

Universal Technologies Corp.
Dayton, OH, USA
bstange@utcdayton.com

William Stange received his MS in Mechanical Engineering from Ohio State University. His career with the Air Force Research Laboratories spanned 40 years, where he worked primarily on turbine engine structural design and durability, and developing new instrumentation for engine test cells. Retiring from the Air Force Bill joined Universal Technologies Corporation continuing work on turbine engine instrumentation and controls.

Subject Index

A

B. Mukherjee et al. (Eds.), *Springer Handbook of Optical Networks*, Springer Handbooks, https://doi.org/10.1007/978-3-030-16250-4

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z