Chapter 11
MPEG Video Coding I — MPEG-1 and 2

11.1 Overview

- **MPEG**: *Moving Pictures Experts Group*, established in 1988 for the development of digital video.
- It is appropriately recognized that proprietary interests need to be maintained within the family of MPEG standards:
  - Accomplished by defining only a compressed bitstream that implicitly defines the decoder.
  - The compression algorithms, and thus the encoders, are completely up to the manufacturers.

11.2 MPEG-1

- MPEG-1 adopts the CCIR601 digital TV format also known as SIF (*Source Input Format*).
- MPEG-1 supports only non-interlaced video. Normally, its picture resolution is:
  - 352×240 for NTSC video at 30 fps
  - 352×288 for PAL video at 25 fps
  - It uses 4:2:0 chroma subsampling
- The MPEG-1 standard is also referred to as ISO/IEC 11172. It has five parts: 11172-1 Systems, 11172-2 Video, 11172-3 Audio, 11172-4 Conformance, and 11172-5 Software.

Motion Compensation in MPEG-1

- Motion Compensation (MC) based video encoding in H.261 works as follows:
  - In Motion Estimation (ME), each macroblock (MB) of the Target P-frame is assigned a best matching MB from the previously coded I or P frame - **prediction**.
  - **prediction error:** The difference between the MB and its matching MB, sent to DCT and its subsequent encoding steps.
  - The prediction is from a previous frame — **forward prediction**.

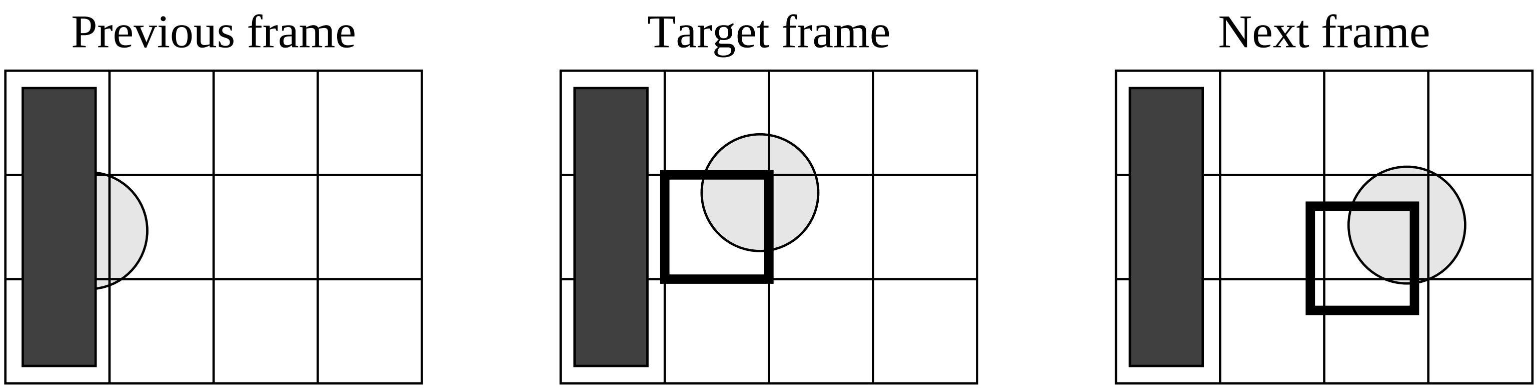


Fig 11.1: The Need for Bidirectional Search.

The MB containing part of a ball in the Target frame cannot find a good matching MB in the previous frame because half of the ball was occluded by another object. A match however can readily be obtained from the next frame.

Motion Compensation in MPEG-1 (Cont'd)

- MPEG introduces a third frame type — *B-frames*, and its accompanying bi-directional motion compensation.
- The MC-based B-frame coding idea is illustrated in Fig. 11.2:
  - Each MB from a B-frame will have up to *two* motion vectors (MVs) (one from the forward and one from the backward prediction).
  - If matching in both directions is successful, then two MVs will be sent and the two corresponding matching MBs are averaged (indicated by '%' in the figure) before comparing to the Target MB for generating the prediction error.
  - If an acceptable match can be found in only one of the reference frames, then only one MV and its corresponding MB will be used from either the forward or backward prediction.

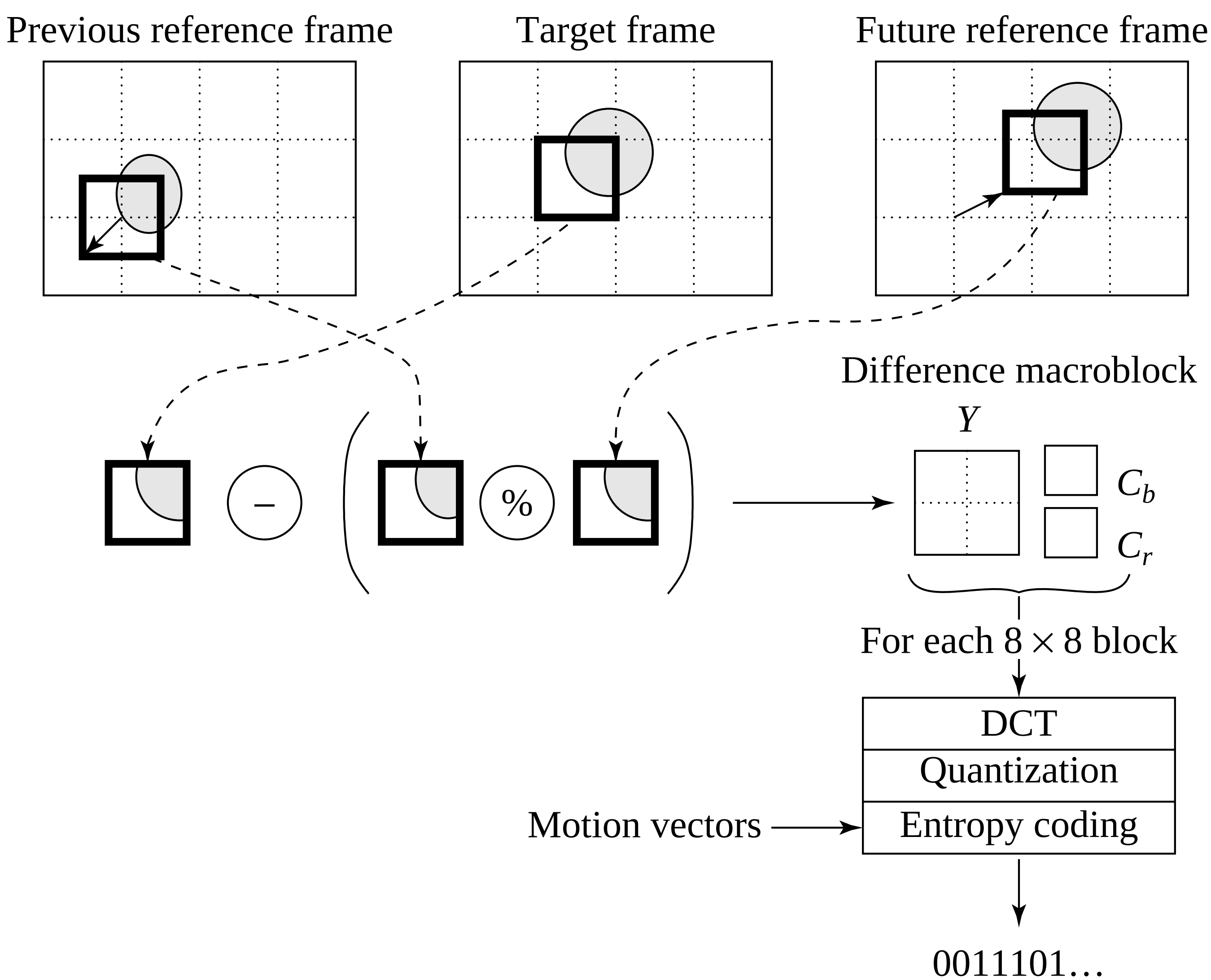


Fig 11.2: B-frame Coding Based on Bidirectional Motion Compensation.

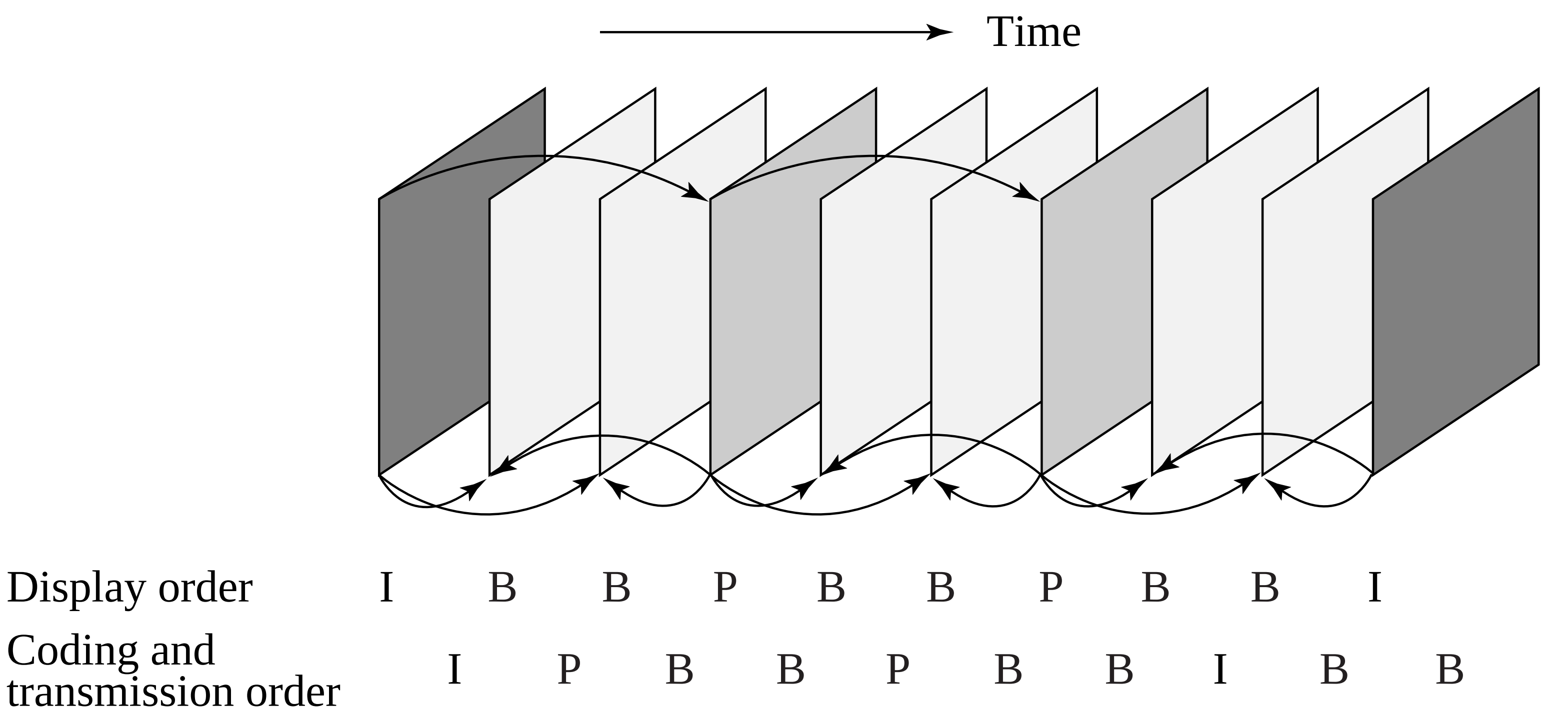


Fig 11.3: MPEG Frame Sequence.

Other Major Differences from H.261

- Source formats supported:
  - H.261 only supports CIF (352×288) and QCIF (176×144) source formats, MPEG-1 supports SIF (352×240 for NTSC, 352×288 for PAL).
  - MPEG-1 also allows specification of other formats as long as the Constrained Parameter Set (CPS) as shown in Table 11.1 is satisfied:

Table 11.1: The MPEG-1 Constrained Parameter Set

Parameter	Value
Horizontal size of picture	≤ 768
Vertical size of picture	≤ 576
No. of MBs / picture	≤ 396
No. of MBs / second	$\leq 9,900$
Frame rate	≤ 30 fps
Bit-rate	$\leq 1,856$ kbps

Other Major Differences from H.261 (Cont'd)

- Instead of GOBs as in H.261, an MPEG-1 picture can be divided into one or more **slices** (Fig. 11.4):
  - May contain variable numbers of macroblocks in a single picture.
  - May also start and end anywhere as long as they fill the whole picture.
  - Each slice is coded independently — additional flexibility in bit-rate control.
  - Slice concept is important for error recovery.

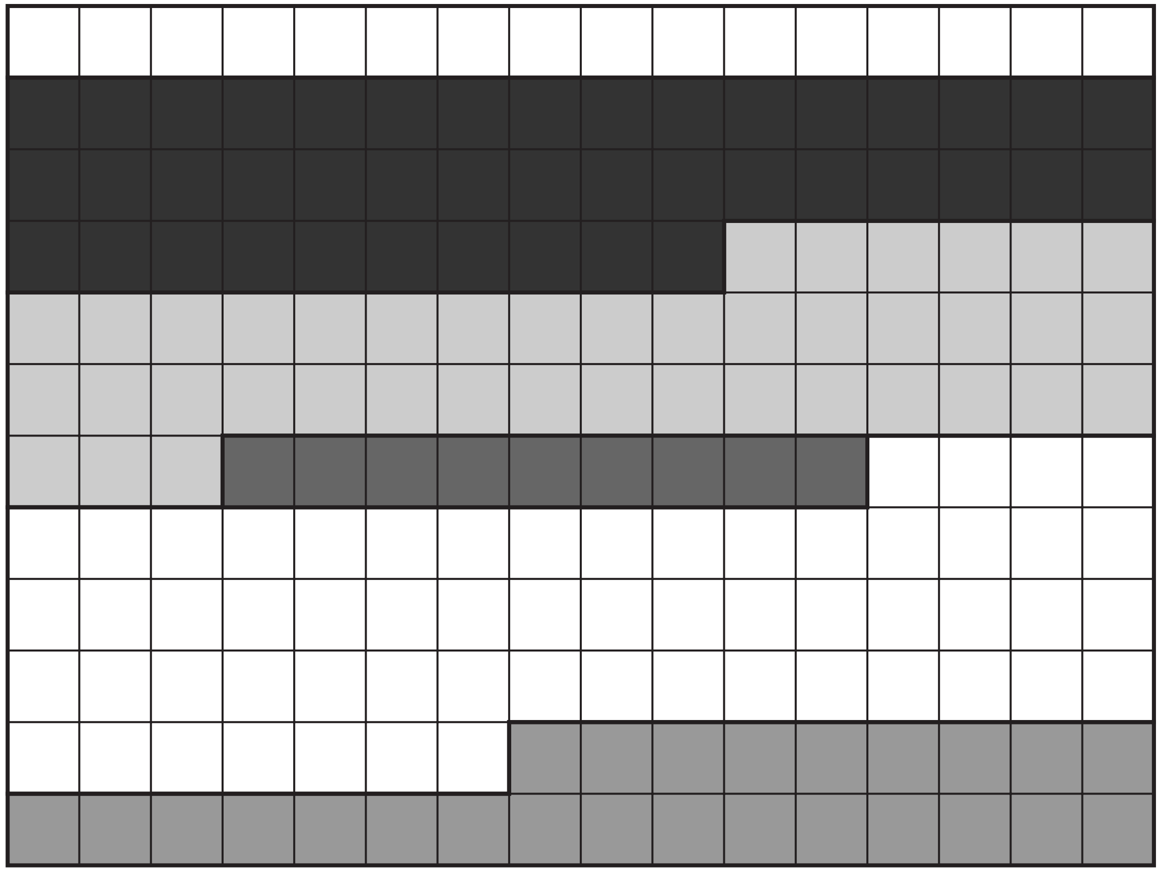


Fig 11.4: Slices in an MPEG-1 Picture.

Other Major Differences from H.261 (Cont'd)

- Quantization:
  - MPEG-1 quantization uses different quantization tables for its Intra and Inter coding (Table 11.2 and 11.3).

For DCT coefficients in Intra mode:

$$QDCT[i,j] = round\left(\frac{8 \times DCT[i,j]}{step\_size[i,j]}\right) = round\left(\frac{8 \times DCT[i,j]}{Q_1[i,j] * scale}\right) \quad (11.1)$$

For DCT coefficients in Inter mode,

$$QDCT[i,j] = \left\lfloor\frac{8 \times DCT[i,j]}{step\_size[i,j]}\right\rfloor = \left\lfloor\frac{8 \times DCT[i,j]}{Q_2[i,j] * scale}\right\rfloor \quad (11.2)$$

Table 11.2: Default Quantization Table (Q_1) for Intra-Coding

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Table 11.3: Default Quantization Table (Q_2) for Inter-Coding

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Other Major Differences from H.261 (Cont'd)

- MPEG-1 allows motion vectors to be of sub-pixel precision (1/2 pixel). The technique of "bilinear interpolation" for H.263 can be used to generate the needed values at half-pixel locations.

- Compared to the maximum range of ± 15 pixels for motion vectors in H.261, MPEG-1 supports a range of $[-512, 511.5]$ for half-pixel precision and $[-1,024, 1,023]$ for full-pixel precision motion vectors.

- The MPEG-1 bitstream allows random access — accomplished by GOP layer in which each GOP is time coded.

Typical Sizes of MPEG-1 Frames

- The typical size of compressed P-frames is significantly smaller than that of I-frames — because temporal redundancy is exploited in inter-frame compression.

- B-frames are even smaller than P-frames — because of (a) the advantage of bi-directional prediction and (b) the lowest priority given to B-frames.

Table 11.4: Typical Compression Performance of MPEG-1 Frames

Type	Size	Compression
I	18 kB	7:1
P	6 kB	20:1
B	2.5 kB	50:1
Avg	4.8 kB	27:1

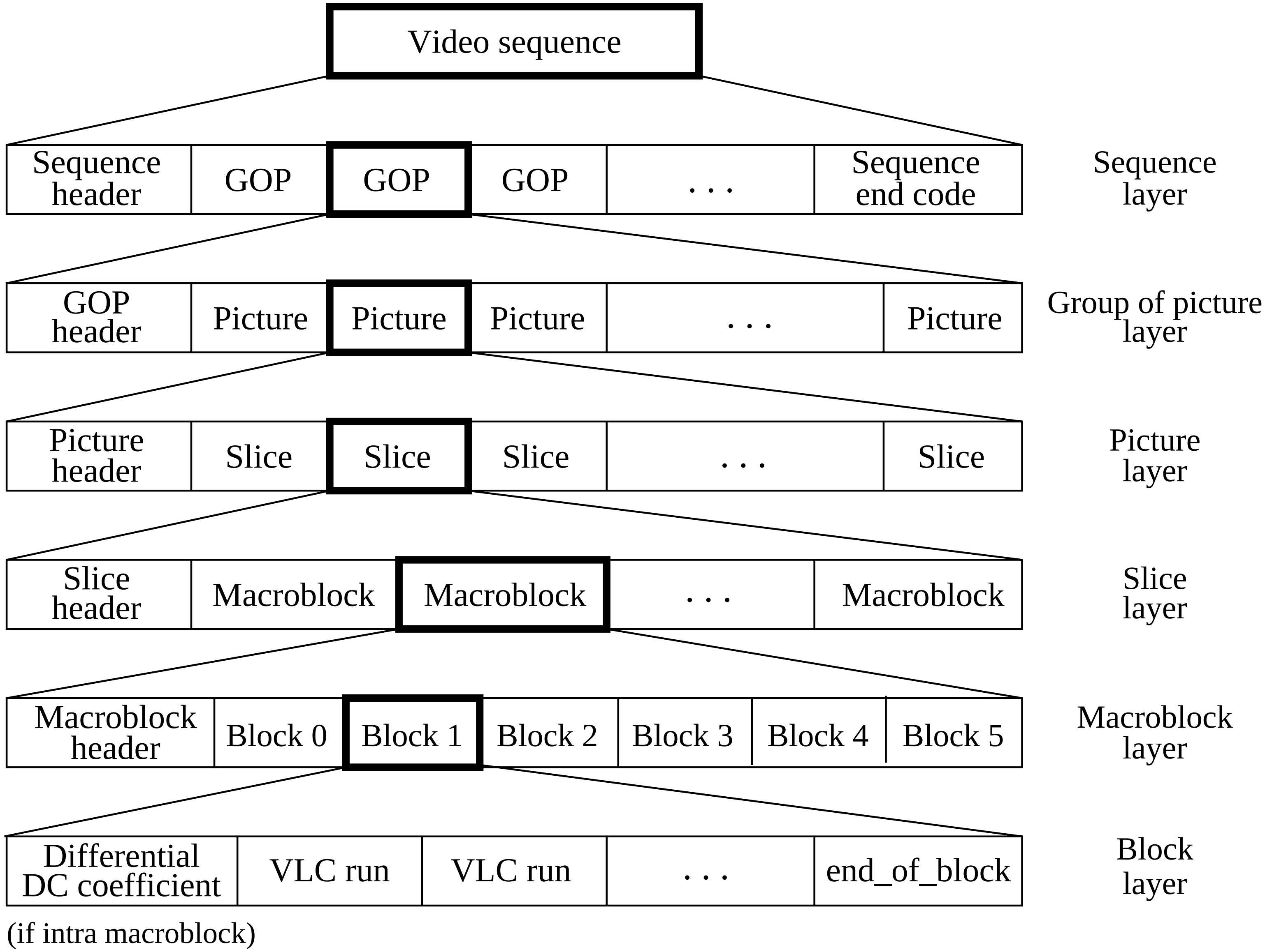


Fig 11.5: Layers of MPEG-1 Video Bitstream.

11.3 MPEG-2

- **MPEG-2**: For higher quality video at a bit-rate of more than 4 Mbps.
- Defined seven **profiles** aimed at different applications:
  - **Simple, Main, SNR scalable, Spatially scalable, High, 4:2:2, Multiview**.
  - Within each profile, up to four *levels* are defined (Table 11.5).
  - The DVD video specification allows only four display resolutions: 720×480, 704×480, 352×480, and 352×240 — a restricted form of the MPEG-2 Main profile at the Main and Low levels.

Table 11.5: Profiles and Levels in MPEG-2

Level	Simple Profile	Main Profile	SNR Scalable Profile	Spatially Scalable Profile	High Profile	4:2:2 Profile	Multiview Profile
High		*			*		
High 1440		*		*	*		
Main	*	*	*		*	*	*
Low		*	*				

Table 11.6: Four Levels in the Main Profile of MPEG-2

Level	Max Resolution	Max fps	Max Pixels/sec	Max coded Data Rate (Mbps)	Application
High	$1,920 \times 1,152$	60	62.7×10^6	80	film production
High 1440	$1,440 \times 1,152$	60	47.0×10^6	60	consumer HDTV
Main	720×576	30	10.4×10^6	15	studio TV
Low	352×288	30	3.0×10^6	4	consumer tape equiv.

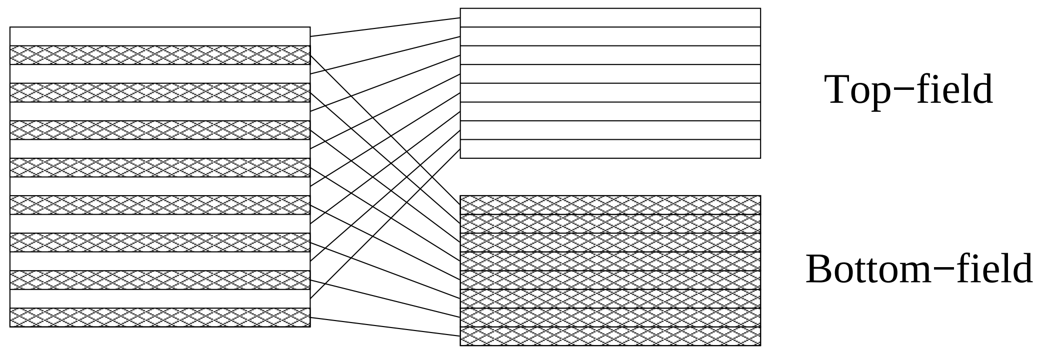
Supporting Interlaced Video

- MPEG-2 must support interlaced video as well since this is one of the options for digital broadcast TV and HDTV.

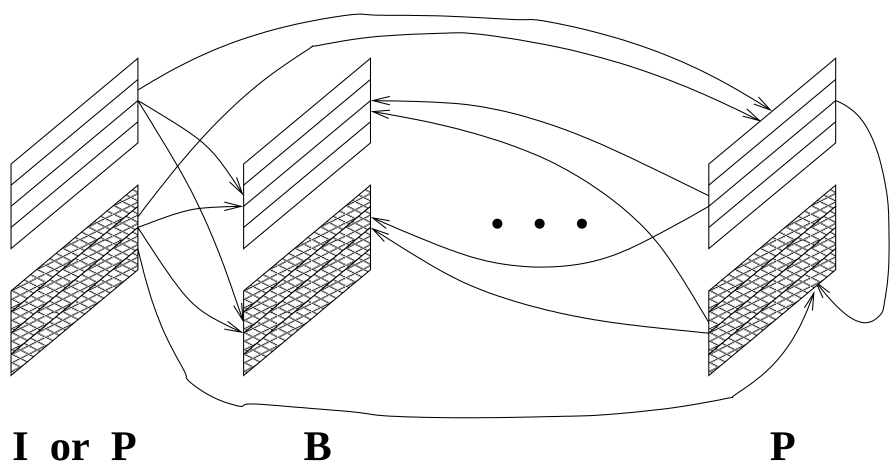
- In interlaced video each frame consists of two fields, referred to as the *top-field* and the *bottom-field*.

  – In a *Frame-picture*, all scanlines from both fields are interleaved to form a single frame, then divided into 16×16 macroblocks and coded using MC.

  – If each field is treated as a separate picture, then it is called *Field-picture*.



(a) Frame−picture vs. Field−pictures



(b) Field Prediction for Field−pictures

Fig. 11.6: Field pictures and Field-prediction for Field-pictures in MPEG-2.

Five Modes of Predictions

- MPEG-2 defines **Frame Prediction** and **Field Prediction** as well as five prediction modes:

  1. **Frame Prediction for Frame-pictures:** Identical to MPEG-1 MC-based prediction methods in both P-frames and B-frames.

  2. **Field Prediction for Field-pictures:** A macroblock size of 16×16 from Field-pictures is used. For details, see Fig. 11.6(b).

3. **Field Prediction for Frame-pictures:** The top-field and bottom-field of a Frame-picture are treated separately. Each 16×16 macroblock (MB) from the target Frame-picture is split into two 16×8 parts, each coming from one field. Field prediction is carried out for these 16×8 parts in a manner similar to that shown in Fig. 11.6(b).

4. 16×8 **MC for Field-pictures:** Each 16×16 macroblock (MB) from the target Field-picture is split into top and bottom 16×8 halves. Field prediction is performed on each half. This generates two motion vectors for each 16×16 MB in the P-Field-picture, and up to four motion vectors for each MB in the B-Field-picture.

   This mode is good for a finer MC when motion is rapid and irregular.

5. **Dual-Prime for P-pictures:** First, Field prediction from each previous field with the same parity (top or bottom) is made. Each motion vector $\mathbf{mv}$ is then used to derive a calculated motion vector $\mathbf{cv}$ in the field with the opposite parity taking into account the temporal scaling and vertical shift between lines in the top and bottom fields. For each MB the pair $\mathbf{mv}$ and $\mathbf{cv}$ yields two preliminary predictions. Their prediction errors are averaged and used as the final prediction error.

   This mode mimics B-picture prediction for P-pictures without adopting backward prediction (and hence with less encoding delay).

   This is the only mode that can be used for either Frame-pictures or Field-pictures.

Alternate Scan and Field_DCT

- Techniques aimed at improving the effectiveness of DCT on prediction errors, only applicable to Frame-pictures in interlaced videos:
  - Due to the nature of interlaced video the consecutive rows in the 8×8 blocks are from different fields, there exists less correlation between them than between the alternate rows.
  - Alternate scan recognizes the fact that in interlaced video the vertically higher spatial frequency components may have larger magnitudes and thus allows them to be scanned earlier in the sequence.
- In MPEG-2, **Field_DCT** can also be used to address the same issue.

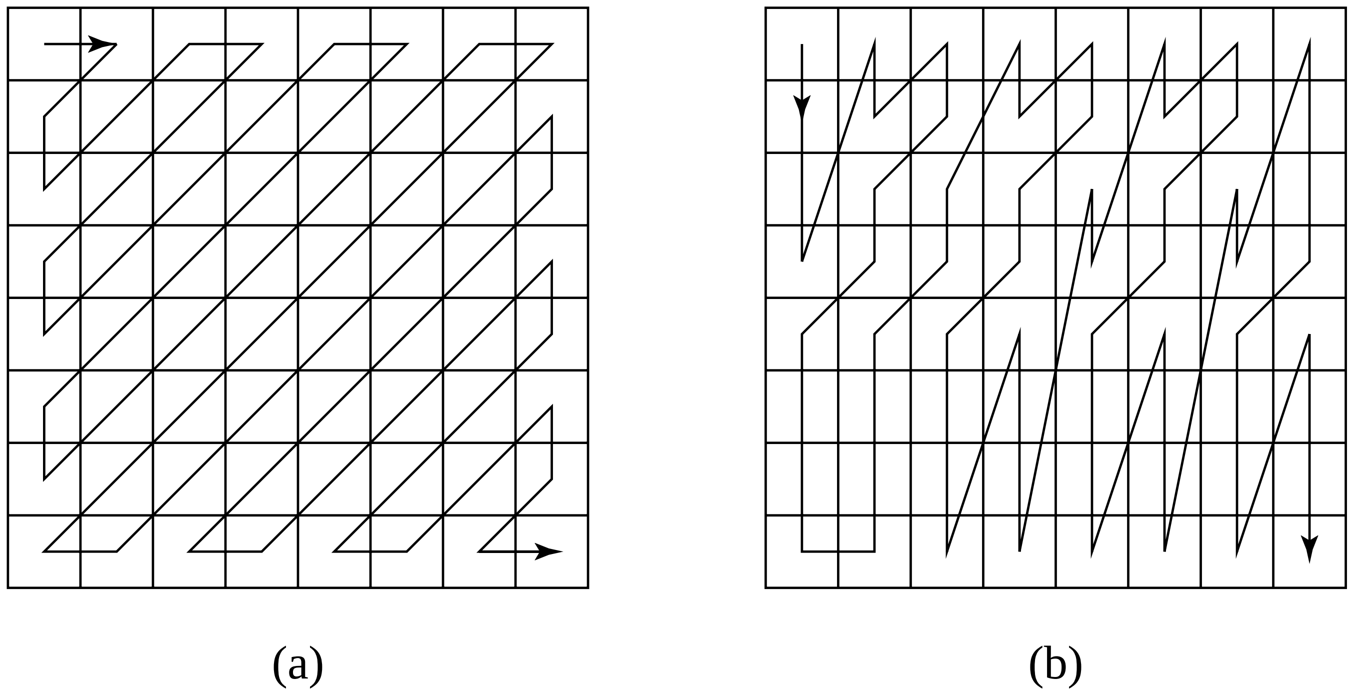


Fig 11.7: Zigzag and Alternate Scans of DCT Coefficients for Progressive and Interlaced Videos in MPEG-2.

MPEG-2 Scalabilities

- The MPEG-2 **scalable coding**: A base layer and one or more enhancement layers can be defined — also known as **layered coding**.
  - The base layer can be independently encoded, transmitted and decoded to obtain basic video quality.
  - The encoding and decoding of the enhancement layer is dependent on the base layer or the previous enhancement layer.
- Scalable coding is especially useful for MPEG-2 video transmitted over networks with following characteristics:
  - Networks with very different bit-rates.
  - Networks with variable bit rate (VBR) channels.
  - Networks with noisy connections.

MPEG-2 Scalabilities (Cont'd)

- MPEG-2 supports the following scalabilities:

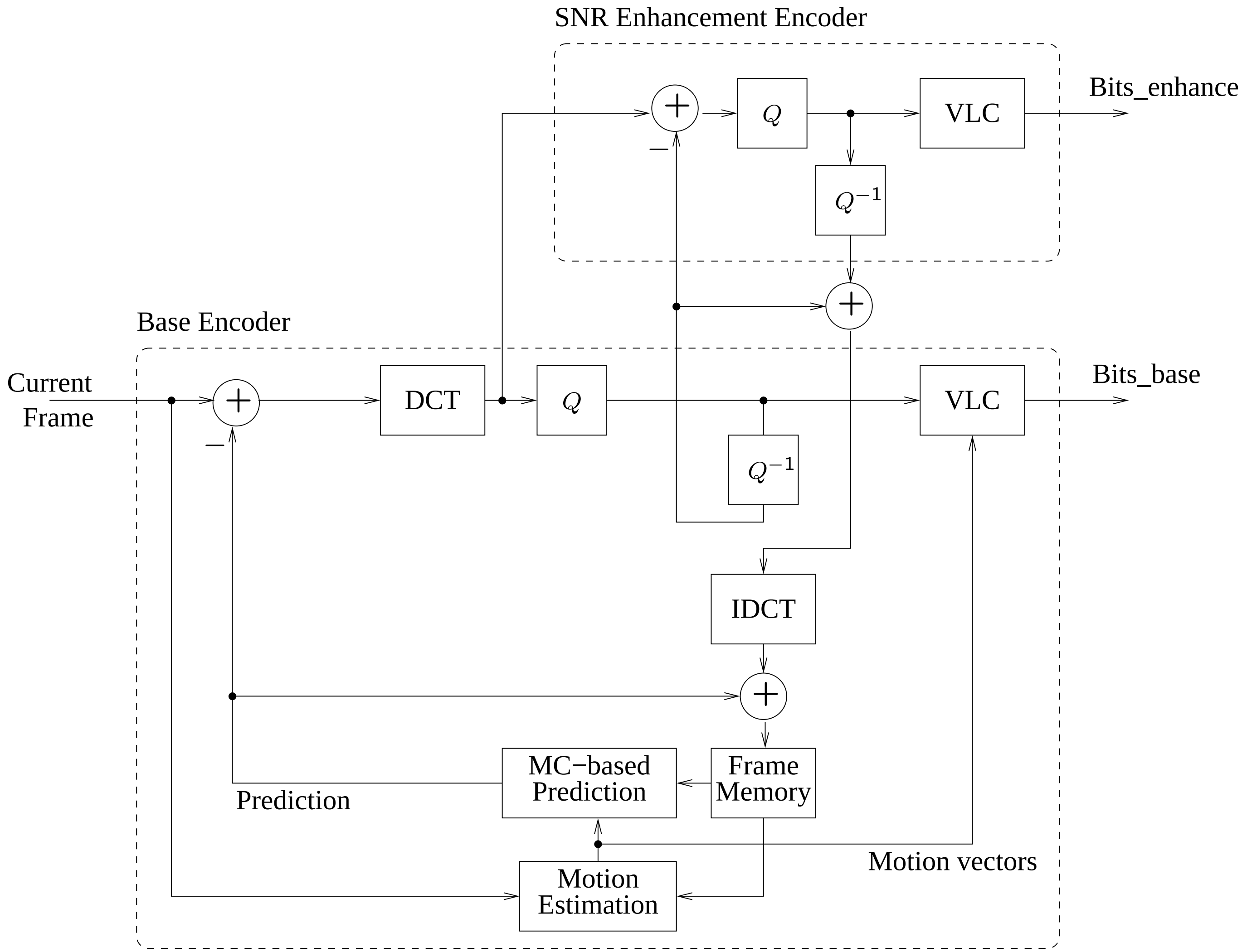
  1. SNR Scalability — enhancement layer provides higher SNR.
  2. Spatial Scalability — enhancement layer provides higher spatial resolution.
  3. Temporal Scalability — enhancement layer facilitates higher frame rate.
  4. Hybrid Scalability — combination of any two of the above three scalabilities.
  5. Data Partitioning — quantized DCT coefficients are split into partitions.

SNR Scalability

- **SNR scalability**: Refers to the enhencement/refinement over the base layer to improve the Signal-Noise-Ratio (SNR).

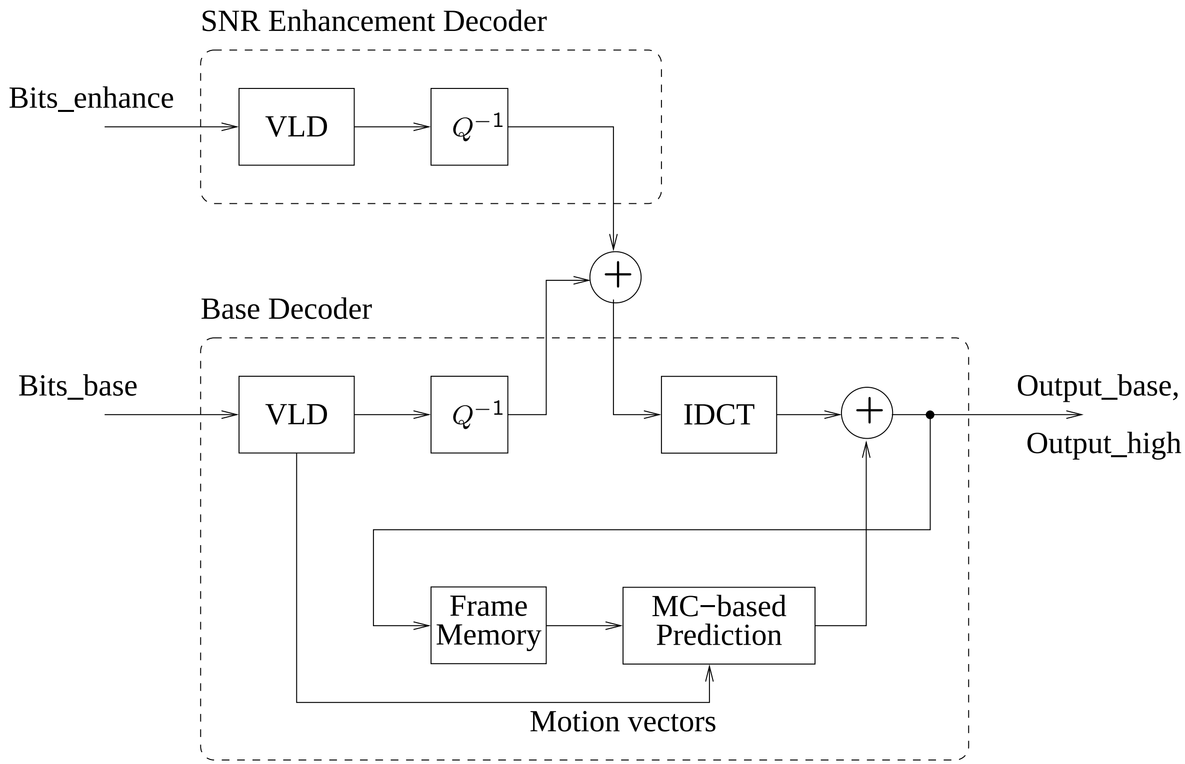
- The MPEG-2 SNR scalable encoder will generate output bitstreams *Bits_base* and *Bits_enhance* at two layers:

  1. At the Base Layer, a coarse quantization of the DCT coefficients is employed which results in fewer bits and a relatively low quality video.
  2. The coarsely quantized DCT coefficients are then inversely quantized (Q^{-1}) and fed to the Enhancement Layer to be compared with the original DCT coefficient.
  3. Their difference is finely quantized to generate a **DCT coefficient refinement**, which, after VLC, becomes the bitstream called Bits_enhance.



(a) Encoder

Fig 11.8 (a): MPEG-2 SNR Scalability (Encoder).



(b) Decoder

Fig 11.8 (b): MPEG-2 SNR Scalability (Decoder).

Spatial Scalability

- The base layer is designed to generate bitstream of reduced-resolution pictures. When combined with the enhancement layer, pictures at the original resolution are produced.

- The Base and Enhancement layers for MPEG-2 spatial scalability are not as tightly coupled as in SNR scalability.

- Fig. 11.9(a) shows a typical block diagram. Fig. 11.9(b) shows a case where temporal and spatial predictions are combined.

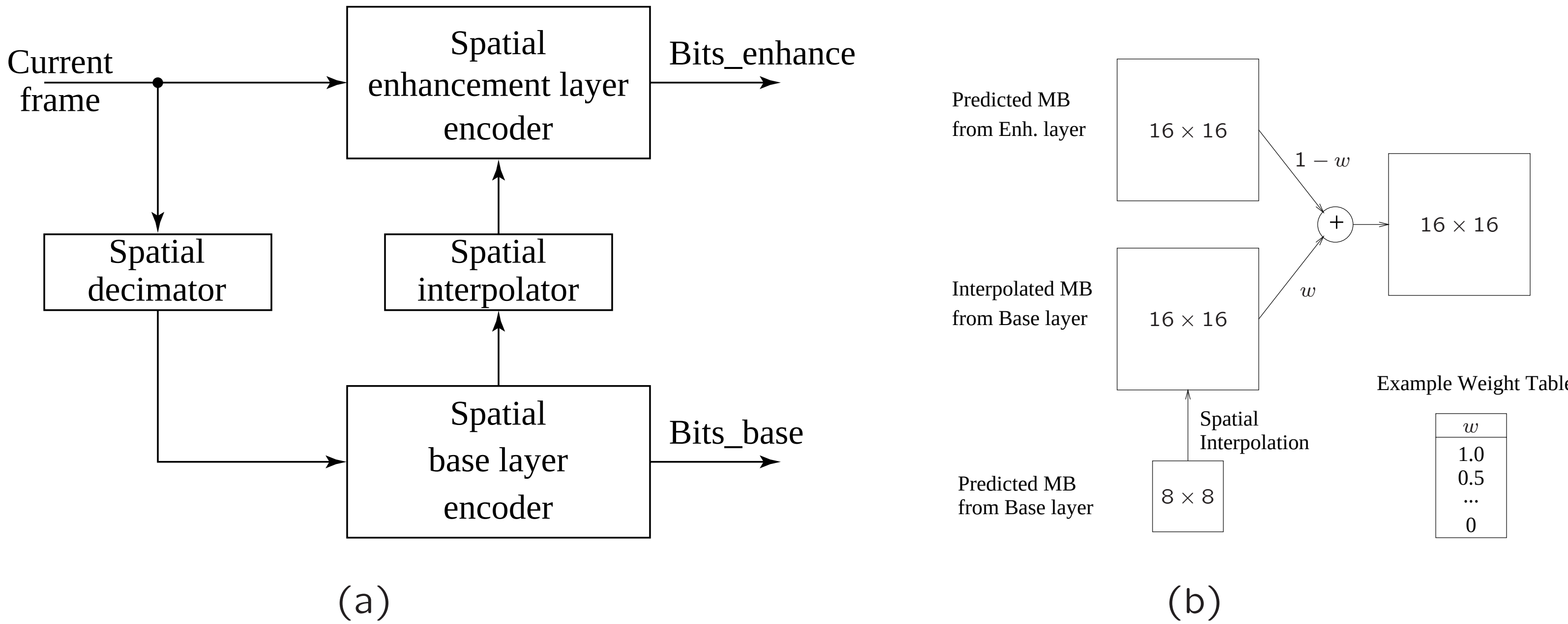
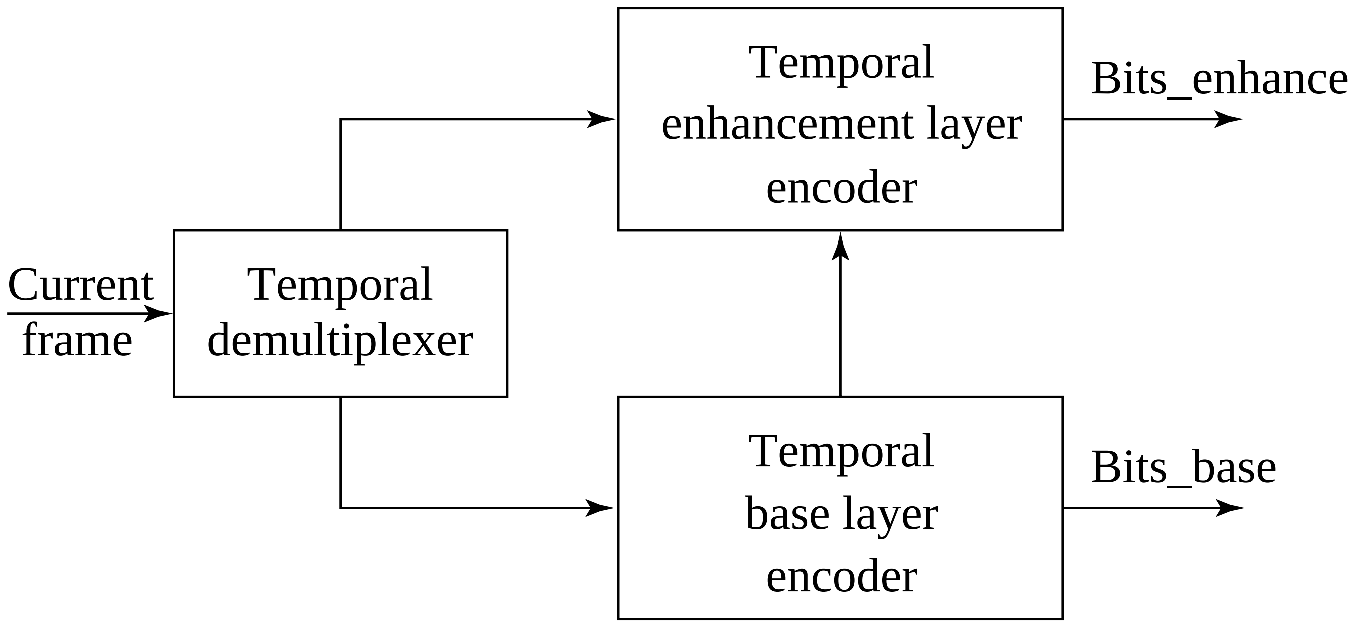


Fig. 11.9: Encoder for MPEG-2 Spatial Scalability. (a) Block Diagram. (b) Combining Temporal and Spatial Predictions for Encoding at Enhancement Layer.

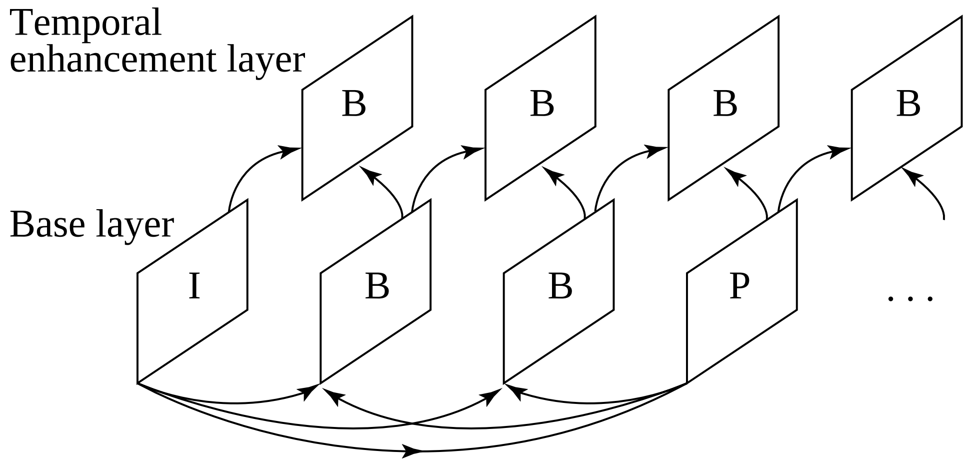
Temporal Scalability

- The input video is temporally demultiplexed into two pieces, each carrying half of the original frame rate.
- Base Layer Encoder carries out the normal single-layer coding procedures for its own input video and yields the output bitstream Bits_base.
- The prediction of matching MBs at the Enhancement Layer can be obtained in two ways:
  - Interlayer MC (Motion-Compensated) Prediction (Fig. 11.10(b))
  - Combined MC Prediction and Interlayer MC Prediction (Fig. 11.10(c))

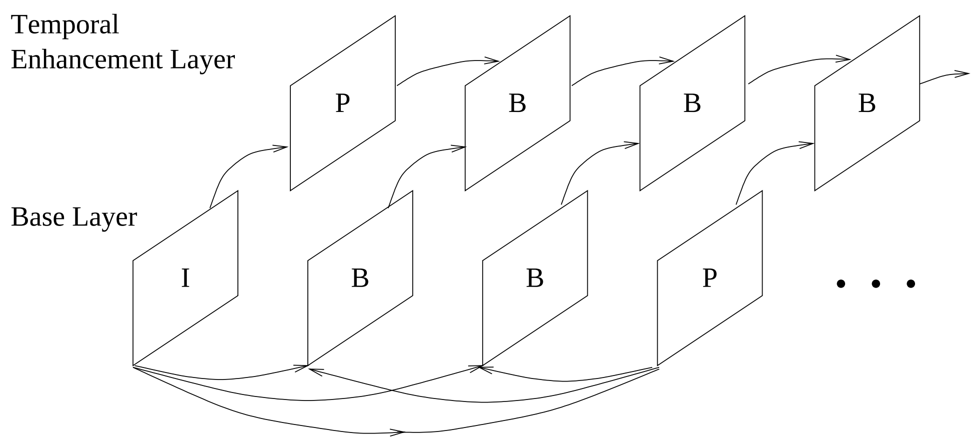


(a) Block Diagram

Fig 11.10: Encoder for MPEG-2 Temporal Scalability.



(b) Interlayer Motion-Compensated (MC) Prediction.



(c) Combined MC Prediction and Interlayer MC Prediction

Fig 11.10 (Cont'd): Encoder for MPEG-2 Temporal Scalability

Hybrid Scalability

- Any two of the above three scalabilities can be combined to form hybrid scalability:

  1. Spatial and Temporal Hybrid Scalability.
  2. SNR and Spatial Hybrid Scalability.
  3. SNR and Temporal Hybrid Scalability.

- Usually, a three-layer hybrid coder will be adopted which consists of Base Layer, Enhancement Layer 1, and Enhancement Layer 2.

Data Partitioning

- *Base partition* contains lower-frequency DCT coefficients, *enhancement partition* contains high-frequency DCT coefficients.

- Strictly speaking, data partitioning is not layered coding, since a single stream of video data is simply divided up and there is no further dependence on the base partition in generating the enhancement partition.

- Useful for transmission over noisy channels and for progressive transmission.

Other Major Differences from MPEG-1

- **Better resilience to bit-errors:** In addition to *Program Stream*, a *Transport Stream* is added to MPEG-2 bit streams.
- **Support of 4:2:2 and 4:4:4 chroma subsampling**.
- **More restricted slice structure:** MPEG-2 slices must start and end in the same macroblock row. In other words, the left edge of a picture always starts a new slice and the longest slice in MPEG-2 can have only one row of macroblocks.
- **More flexible video formats:** It supports various picture resolutions as defined by DVD, ATV and HDTV.

Other Major Differences from MPEG-1 (Cont'd)

- **Nonlinear quantization** — two types of scales are allowed:

  1. For the first type, $scale$ is the same as in MPEG-1 in which it is an integer in the range of [1, 31] and $scale_i = i$.
  2. For the second type, a nonlinear relationship exists, i.e., $scale_i \neq i$. The ith scale value can be looked up from Table 11.7.

Table 11.7: Possible Nonlinear Scale in MPEG-2

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$scale_i$	1	2	3	4	5	6	7	8	10	12	14	16	18	20	22	24
i	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
$scale_i$	28	32	36	40	44	48	52	56	64	72	80	88	96	104	112	

11.4 Further Exploration

- **Text books:**
  - *Video Compression Standard* by J.L. Mitchell et al
  - *Digital Video: An Introduction to MPEG-2* by B.G. Haskell et al

- **Web sites:** ⟶ Link to Further Exploration for Chapter 11.. including:
  - The MPEG home page.
  - MPEG FAQ page.
  - Overviews and working documents of the MPEG-1 and MPEG-2 standards.