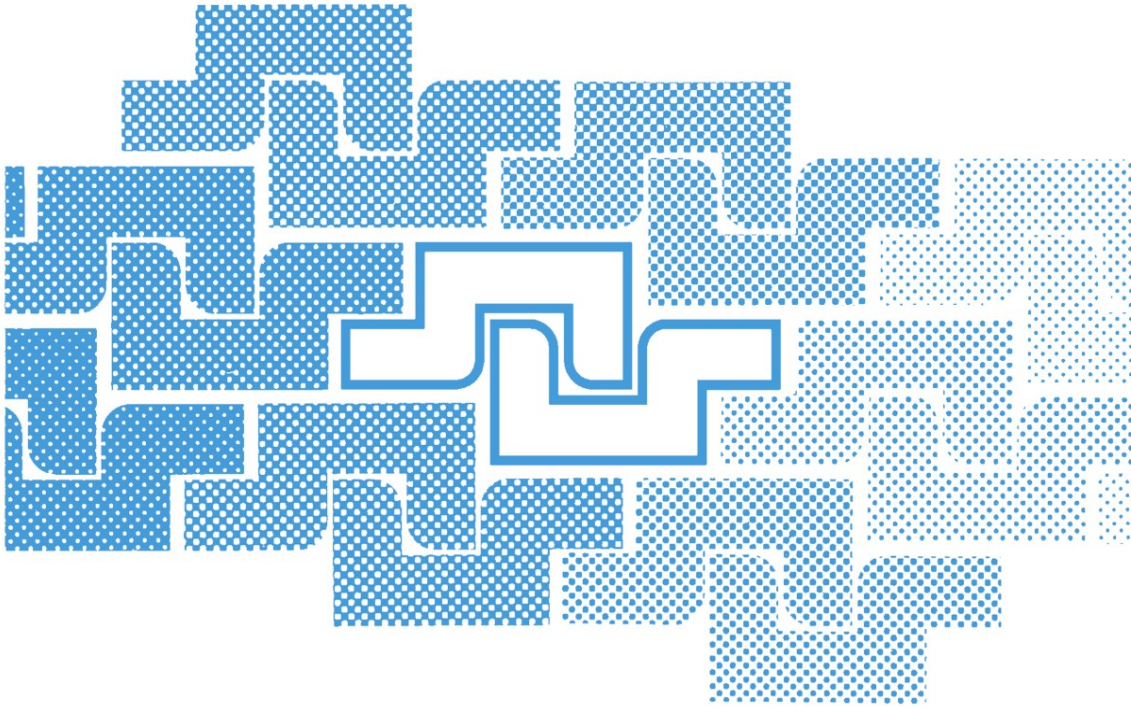

Effects of Space Weather on Technology Infrastructure

Edited by

Ioannis A. Daglis

NATO Science Series

II. Mathematics, Physics and Chemistry – Vol. 176

Effects of Space Weather on Technology Infrastructure

NATO Science Series

A Series presenting the results of scientific meetings supported under the NATO Science Programme.

The Series is published by IOS Press, Amsterdam, and Kluwer Academic Publishers in conjunction with the NATO Scientific Affairs Division

Sub-Series

I. Life and Behavioural Sciences	IOS Press
II. Mathematics, Physics and Chemistry	Kluwer Academic Publishers
III. Computer and Systems Science	IOS Press
IV. Earth and Environmental Sciences	Kluwer Academic Publishers
V. Science and Technology Policy	IOS Press

The NATO Science Series continues the series of books published formerly as the NATO ASI Series.

The NATO Science Programme offers support for collaboration in civil science between scientists of countries of the Euro-Atlantic Partnership Council. The types of scientific meeting generally supported are "Advanced Study Institutes" and "Advanced Research Workshops", although other types of meeting are supported from time to time. The NATO Science Series collects together the results of these meetings. The meetings are co-organized bij scientists from NATO countries and scientists from NATO's Partner countries – countries of the CIS and Central and Eastern Europe.

Advanced Study Institutes are high-level tutorial courses offering in-depth study of latest advances in a field.
Advanced Research Workshops are expert meetings aimed at critical assessment of a field, and identification of directions for future action.

As a consequence of the restructuring of the NATO Science Programme in 1999, the NATO Science Series has been re-organised and there are currently Five Sub-series as noted above. Please consult the following web sites for information on previous volumes published in the Series, as well as details of earlier Sub-series.

http://www.nato.int/science
http://www.wkap.nl
http://www.iospress.nl
http://www.wtv-books.de/nato-pco.htm

Series II: Mathematics, Physics and Chemistry – Vol. 176

Effects of Space Weather on Technology Infrastructure

edited by

Ioannis A. Daglis
Institute for Space Applications and Remote Sensing,
National Observatory of Athens, Greece

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

CD-ROM available only in print edition
eBook ISBN: 1-4020-2754-0
Print ISBN: 1-4020-2748-6

Created in the United States of America

Visit Springer's eBookstore at: http://ebooks.springerlink.com
and the Springer Global Website Online at: http://www.springeronline.com

TABLE OF CONTENTS

PREFACE

The 17 chapters of this book grew out of the tutorial lectures given by leading world-class experts at the NATO Advanced Research Workshop "Effects of Space Weather on Technology Infrastructure" - ESPRIT, which was held in Rhodes on March 25-29, 2004. All manuscripts were refereed and subsequently meticulously edited by the editor to ensure the highest quality for this monograph. I owe particular thanks to the lecturers of the ESPRIT Advanced Research Workshop for producing these excellent tutorial reviews, which convey the essential knowledge and the latest advances in our field. Due to the breadth, extensive literature citations and quality of the reviews we expect this publication to serve extremely well as a reference book. Multimedia material referring to individual chapters of the book is accessible on the accompanying CD.

The aim of ESPRIT was to assess existing knowledge and identify future actions regarding monitoring, forecasting and mitigation of space weather induced malfunction and damage of vital technological systems operating in space and on the ground.

The objectives of this Advanced Research Workshop were:

(a) to review, assess and organize the existing knowledge on dynamic physical processes in the near-Earth space affecting the functionality and integrity of vital technological infrastructure;

(b) to define the major mechanisms in which space weather disturbances influence technological systems;

(c) to plan the expansion of existing and the development of new reliable methods of monitoring space weather disturbances;

(d) to identify accurate methods of forecasting space weather disturbances and efficient methods of mitigating their effects on technological systems;

(e) to propose directions for future research, and promote co-operation and networking between scientists from different research disciplines and/or different countries.

The particular nature of the scientific and technological issues of ESPRIT, which are at the forefront of current research, defined the multi-disciplinary character of the Workshop. The greatest advantage of ESPRIT was the fact that many distinguished and internationally recognized experts in a large variety of fields (space physics, geomagnetism, space technology, wireless communications, electrical engineering) participated in the meeting and led to its success.

I am grateful to the NATO Scientific and Environmental Affairs Division for making this Advanced Research Workshop possible. Furthermore, I wish to thank the following sponsors for their contribution to the success of this conference:

Committee on Space Research (COSPAR)
European Space Agency (ESA)
National Observatory of Athens (NOA)
US Navy Office of Naval Research International Field Office (ONRIFO)
USAF European Office of Aerospace Research and Development (EOARD)
AMPTEK

I am indebted to the many individuals who were instrumental played a key role in the Workshop organization and realization. In particular, I would like to thank my co-director for this ARW Mikhail Panasyuk; the members of the Scientific Organizing Committee Daniel Baker, Eamonn Daly, Gregory Ginet, and John Kappenman; the session-chairmen Kanaris Tsinganos, Richard Horne, Robert McPherron, Maurizio Candidi, Murray Dryer, Rainer Schwenn, Dennis Papadopoulos, Isidoros Doxas, Bertram Arbesser-Rastburg, Risto Pirjola, Paul Cannon, and Volker Bothmer. Finally, I am grateful to Kostas Koutroumbas, Paul Bellaire, David Southwood, Chris Butler, David Burns, Roger Bonnet, and Maria Papadaki, who contributed in many different ways to the success of ESPRIT.

I would also like to spend a few words on the venue site, and will therefore indulge to a brief historical account. It appears rather appropriate to hold a conference devoted to the influence of the Sun on the achievements of Mankind in Rhodes, the island of Helios - the Sun God. Although originally distinct deities, Helios was confused, as early as the fifth century BC, with Apollo (originally the god of music, the arts, archery, healing and prophecy - and later of light), so that Apollo frequently took on the function of the Sun God himself. The epithets Phoebus “the brilliant” and Xanthos “the fair”, used to describe Apollo, point to this solar connection.

The liveliest cult of Helios in the ancient Greek world existed on the island of Rhodes. In honour of what was effectively their national deity and to commemorate their heroic defence against Demetrius Poliorcetes, the people of Rhodes commissioned the celebrated sculptor Chares of Lindos to create a huge statue of Helios.

This statue, which is known to us as the “Colossus of Rhodes”, was one of the Wonders of the ancient world. It was completed in 292 BC, twelve years after work began on it. Recent studies suggest that it was erected either on the eastern promontory of the Rhodes’ harbour, or even further inland, and was over 35 meters tall. Helios was represented with a crown of sun-rays, a spear in his left hand and a flaming torch held aloft in his right. Descriptions of this ancient statue inspired the design of France’s gift to the

people of the USA in 1884 - the Statue of Liberty as the inscription at the base of this New York landmark acknowledges.

The story behind Colossus is the following. In 408 BC, the ancient cities of the island, Ialysos, Kamiros and Lindos, joined a unified state, with a common capital, Rhodes. Rhodes thrived commercially and had strong economic ties with their main ally, the Greek king of Egypt, Ptolemy I Soter ("The Saviour"), one of the Diadochi (successors) of Alexander the Great, and founder of the Ptolemaic dynasty – the 32nd Dynasty which turned out to be the last of Egypt's great dynasties, with famous Cleopatra marking its end.

In 307 BC, the Antigonids of Macedonia who were rivals of the Ptolemies, besieged Rhodes in an attempt to break the Rhodo-Egyptian alliance. Demetrius Poliorcetes ("The Besieger") was the son of Antigonus I, also one of the Diadochi and founder of the Antigonid dynasty. On his father's orders, Demetrius sailed with a large fleet to Rhodes, and besieged the city by land and by sea. This siege has remained famous in military history for the great number of siege engines used, all designed and built by Demetrius and his engineers. The people of Rhodes defended their city with great courage and determination, and in 304 BC Demetrius raised the siege and presented them with the siege engines. To celebrate their victory, the Rhodians sold the equipment and used the money to erect the statue of Helios.

Athens, April 2004

Ioannis A. Daglis

Chapter 1

Specifying and Forecasting Space Weather Threats to Human Technology

Daniel N. Baker
Laboratory for Atmospheric and Space Physics, University of Colorado
1234 Innovation Drive, Boulder, CO 80303-7814, USA

Abstract Adverse space weather is one of the principal threats to modern human technology. Solar coronal mass ejections, large solar flares, and high-speed solar wind streams often lead to sequences of damaging disturbances within the Earth's magnetosphere, in the atmosphere, and even on the Earth's surface. Powerful geomagnetic storms can develop following solar disturbances and enhancements of energetic particle populations throughout the outer terrestrial radiation zone can subsequently result. High-energy particles can damage satellite solar power panels, confuse optical trackers, and deposit harmful charges into sensitive electronic components. Extreme solar, geomagnetic and solar wind conditions are often observed by a large array of international satellites and ground-based sensors. We discuss several of the types of space weather-related problems that have been identified in recent years. We present examples of space weather-induced spacecraft anomalies and failures. Recently developed models for specifying and forecasting high-energy electron enhancements throughout the Earth's radiation belts are discussed. Present models can be used to specify accurately the electron population anywhere in the outer radiation zone and can also provide several-day prediction of this dangerous environmental hazard. An important component of the present community effort has been to establish an NSF-sponsored Science and Technology Center called CISM (Center for Integrated Space Weather Modeling). This Center will develop coupled models from the sun to the Earth's atmosphere.

Keywords Space weather, spacecraft charging, geomagnetic storms, energetic particles, radiation belts

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 1-25.

1. INTRODUCTION

As shown by Figure 1 (taken from recent NASA Roadmap documents), the Sun and its interaction with the Earth is in many ways the archetype for present understanding of much of cosmic plasma physics. The upper group of insets in Figure 1 portray fundamental aspects of magnetospheres, topics pertaining to comparative planetary environments, and broad issues concerning the plasma universe. Understanding in all these domains springs from – and is built upon – our studies of Sun-Earth connections.

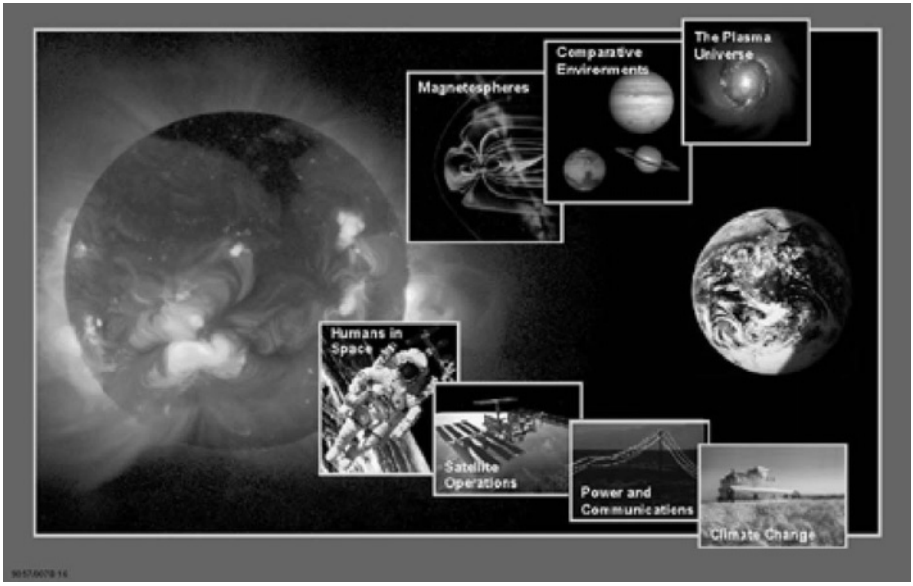

Figure 1. Scientific (upper insets) and applications-related (lower insets) aspects of Sun-Earth connections research (courtesy of NASA).

However, the lower group of insets in Figure 1 makes another very important point. The Sun-Earth environment that is studied for its basic science value is also a system that has crucial practical importance. The effects of the space environment on human beings in space, on satellite operations, on the electrical power grid, upon communication links, and probably even on climate make understanding of Sun-Earth connections critically important. Thus, space weather is vitally important from a pragmatic point of view, just as space plasma physics is important from a basic science point of view.

This chapter addresses several aspects of space weather. It describes effects on human technological systems due to solar and magnetospheric particles. It also describes present attempts to specify and forecast such

damaging particle populations. The paper concludes with a brief discussion of a newly established Center for Space Weather Modeling (CISM). It should be noted that there are many facets to space weather besides energetic particles and magnetospheric effects. For example, there is immense concern about ionospheric disturbances that can affect radio communication and low-altitude spacecraft. This chapter does not deal with these aspects of space weather. Other chapters in this book cover many of these complementary space weather problems.

2. ENERGETIC ION EFFECTS

Very energetic (E≥10 MeV) ions in near-Earth space can originate in several different ways. The most energetic particles are galactic cosmic rays (GCRs) which reach the Earth's vicinity from beyond our own solar system. GCRs consist mostly of energetic protons, but have a heavier ion component extending up to iron (Fe) nuclei. The GCR population is largely accelerated in supernova events and has individual particles extending up to extremely high energies; the most significant portion of the GCR population from a space weather standpoint is in the peak flux part of the energy spectrum, i.e., in the tens of MeV to ~1 GeV range. Polar ground-based technologies, or low-altitude polar-orbiting spacecraft, may be strongly affected by galactic cosmic rays. Humans in polar-transiting airplanes or on long-duration spaceflights outside of Earth's protective magnetic umbrella can also be subjected to strong GCR influence. The most problematic part of the cosmic rays is the highly ionizing and relatively abundant Fe nuclei. These heavy ions can cause severe tissue damage in humans and cause major single event upsets (SEUs) in space electronics systems.

Another energetic ion population of great concern is represented by solar energetic particle (SEP) events. SEPs are produced on an irregular basis during solar flares and in association with coronal mass ejections (CMEs). SEP events tend to be most frequent and most intense in the years around the sunspot activity maximum, but solar events can occur at almost any time in the solar activity cycle. SEP ions can penetrate into Earth's magnetosphere and can become trapped on magnetic field lines.

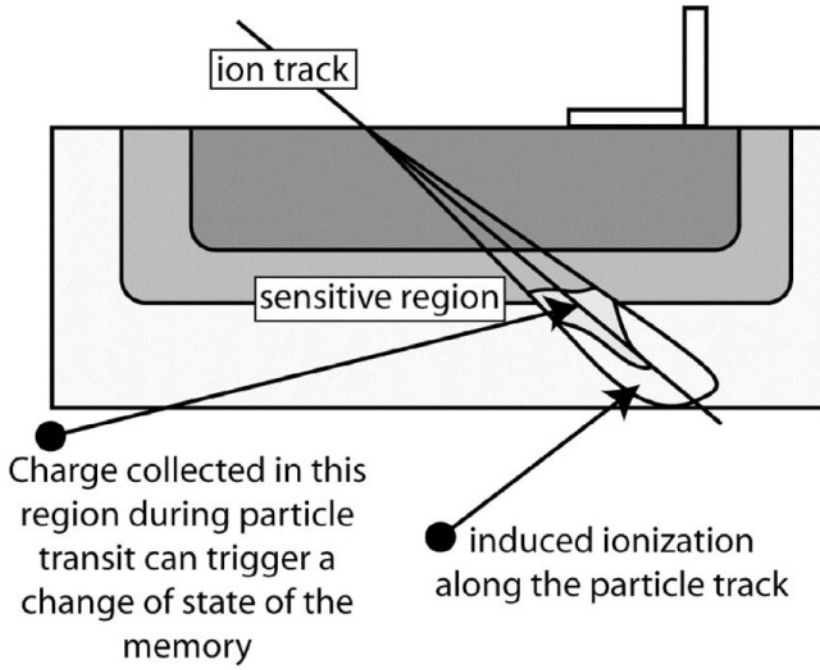

Figure 2. Ion interactions causing single-event upsets (SEUs) (adapted from Robinson, 1989).

As shown in Figure 2, single event upsets occur in microelectronics when an individual charged particle, usually a heavy ion, deposits enough charge at a sensitive portion of the circuit to cause that circuit to change state (Robinson, 1989). The size of the electronics element helps determine the sensitivity as well as the probability that an SEU will occur. It a particle deposits sufficient charge along a sensitive path in a device, then a number of things can happen. This is illustrated in Figure 2: A single heavy ion loses energy by ionizing the constituents of the material. If the energetic particle passes through a depletion region of a transistor in a flip-flop circuit, this can cause the device to change state. It is even possible for a proton to cause the same effect in the circuit by causing a nuclear reaction in or very close to the sensitive region. In very sensitive electronics, it is possible for a proton to cause an SEU directly.

Not only are galactic cosmic rays (specifically the heavy ion component) important in space weather, but protons and heavy ions from solar particle events or in the trapped radiation belts (especially in the South Atlantic Anomaly region) can contribute significant problems during critical phases of space missions. Therefore, it is quite important to recognize the spatial dependence of a particular phenomenon when considering system engineering options. The heavy ions in the Earth's radiation belts can often be handled by mass shielding. On the other hand, many times it is nearly

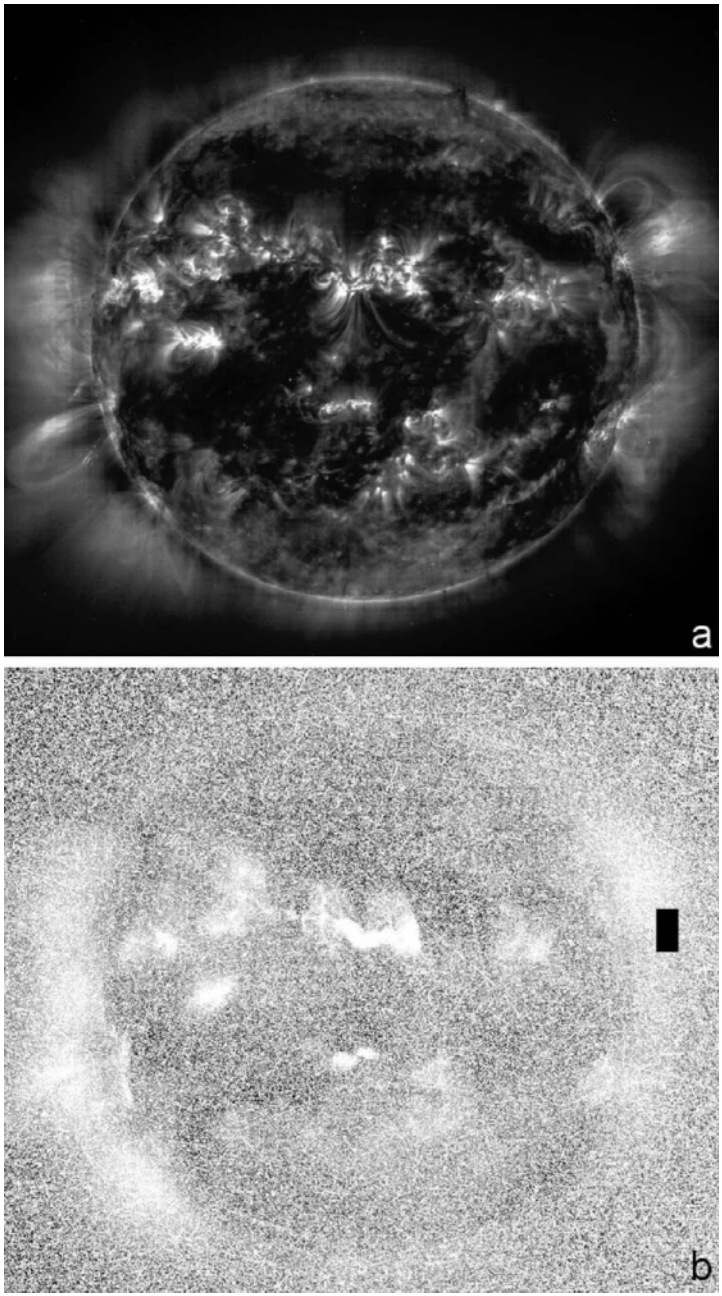

Figure 3. (a) An image of the solar corona taken by the SOHO/EIT experiment at ~0700 UT on 14 July 2000; (b) A similar image to (a) but taken at 1328 UT. The CCD of the EIT camera is virtually "blinded" by solar energetic particle background in the lower image (courtesy of the EIT Consortium of NASA/SOHO mission).

impossible to shield against very energetic cosmic heavy ions. The complexity of the space environment makes solutions to some SEU problems difficult. However, there often are workable and effective solutions.

Figure 3 shows another ion-related problem that has manifested itself increasingly with the use of CCD (charge-coupled device) cameras. Figure 3a shows a spectacular SOHO/EIT picture of the Sun taken in extreme ultraviolet wavelengths at ~0700 UT on 14 July 2000. A powerful solar flare and CME release took place, producing a large SEP event. As shown in Figure 3b, these solar particles rapidly propagated toward the Earth and reached the SOHO spacecraft. The image at 1328 UT on 14 July is virtually obscured by the "snow" created in the EIT CCD camera. This background is produced by energetic ions from the solar disturbance getting into the CCD.

The occurrence of large disturbances on the Sun's surface has long been known to be followed after a few hours to a few days by significant geomagnetic disturbances. The nature of the terrestrial effects can include large magnetic storms, ionospheric disruptions, and intense terrestrial surface events. Solar coronal disturbances often accelerate very energetic particles and also often give rise to strong traveling shock waves in the interplanetary medium. Given a proper interplanetary magnetic field (IMF) connection between the disturbance site on the Sun and the Earth, very energetic solar protons can begin reaching the terrestrial environs within tens of minutes and peak in a matter of hours. At Earth, these very energetic protons have ready access to the polar cap regions and outer magnetosphere. Thus, there can be very prompt effects from solar outbursts.

A more delayed effect results from the shock waves often produced in the solar wind by coronal mass ejections. Since radial propagation speeds are normally $\leq$1000 km/s for these disturbances, it takes of order 1.5-2.0 days for a shock wave to reach the Earth. The high solar wind speed accompanying the shock, plus high plasma number densities and strong southward IMF, normally compress the magnetosphere greatly and initiate sudden storm commencements (SSCs). The geomagnetic storms typically produce large magnetic disturbances due to ring current enhancements (see Gosling et al., 1991) and also result in intense substorm activity (Akasofu, 1981).

Of major significance is the fact that solar particle events, like galactic cosmic rays, can have important effects on passengers and crews of polar-crossing aircraft and manned spacecraft. Having warnings and alerts of possible solar events in certain extreme cases can be very important for the individuals involved. The damaging aspects of solar energetic particles on

spacecraft come in significant measure from fluence effects. Protons and other ions in the ten to several hundred MeV range are very penetrating, and one large event can be as damaging as years of on-orbit operation in the normal, ambient environment.

3. VERY HIGH ENERGY ELECTRONS

It has been demonstrated (e.g., Reagan et al., 1983; Vampola, 1987, Baker et al., 1987) that irradiance of space systems by very energetic electrons can cause deep dielectric charging. In this process, very high-energy (i.e., very penetrating) electrons bury themselves in dielectric materials (e.g., coaxial cables). These electrons then give rise to high electric fields (potential differences of several kilovolts) in their vicinity until eventually an intense breakdown occurs (see Figure 4). In many cases an irrefutable correlation of spacecraft anomalies with the high-energy electron environment exists and the plausible physical charging relationship is well established.

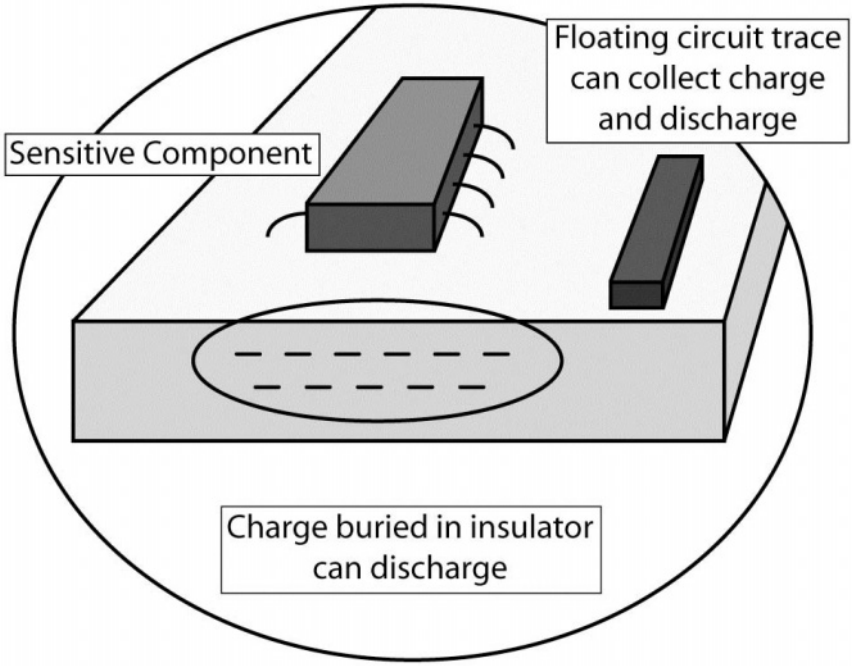

Figure 4. Deep dielectric charging scenario for space-borne electronics (adapted from Robinson, 1989).

Many examples of such dielectric charging have been presented by various authors (e.g., Vampola, 1987). One example of the repeated

occurrence was presented by Baker et al. (1987) and is shown here in Figure 5. Smoothed daily averages of E=1.4-2.0 MeV electron fluxes at geostationary orbit are plotted versus time (for late 1980 through early 1982). Also shown by bold vertical arrows are the main occurrences of star tracker anomalies onboard this geostationary operational spacecraft. The star tracker upsets were clearly associated with high intensities of relativistic electrons. However, some high intensity electron events did not produce star tracker anomalies (see Baker et al., 1987) so there were more subtle controlling factors as well. Evidently, electrons must build up in dielectric materials for quite some time before a harmful discharge can occur. During some intense events in late 1981, the star trackers were actually turned off and so no operational "anomalies" could be recorded. The actual problems tended quite clearly to occur only during relatively long-duration events and it was not only the peak intensity of electrons, but also the duration of exposure that proved to be important.

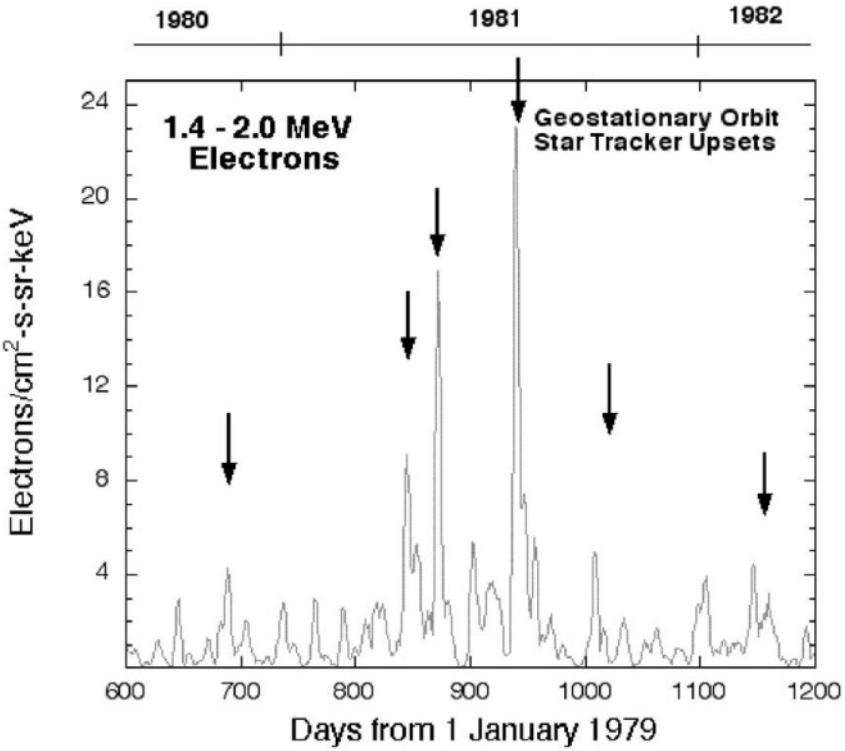

Figure 5. Fluxes of 1.4-2.0 meV electrons at geostationary orbit from late 1980 through early 1982. High electron flux events tended to be associated with star tracker anomalies (vertical arrows) on the spacecraft (adapted from Baker, 2001).

Numerous previous studies (e.g., Reagan et al, 1983; Robinson, 1989; Wrenn, 1995) have shown the clear role played by high-energy electrons in

many classes of spacecraft operational problems. Moreover, the quantitative level of radiation needed to produce deep-dielectric discharges has been rather clearly established in laboratory and spacecraft studies (e.g., Vampola, 1987). Figure 6, for example, adapted from Vampola's work shows results from the SCATHA mission which operated near geostationary orbit during the late 1970s and early 1980s. Deep dielectric discharges were monitored onboard the spacecraft and the daily fluences of E>300 keV electrons were concurrently measured. The probabilities of discharges went up dramatically when daily fluences exceeded 1011 electrons/cm^2. Above 1012 e/cm^2, the probability of discharges approached unity.

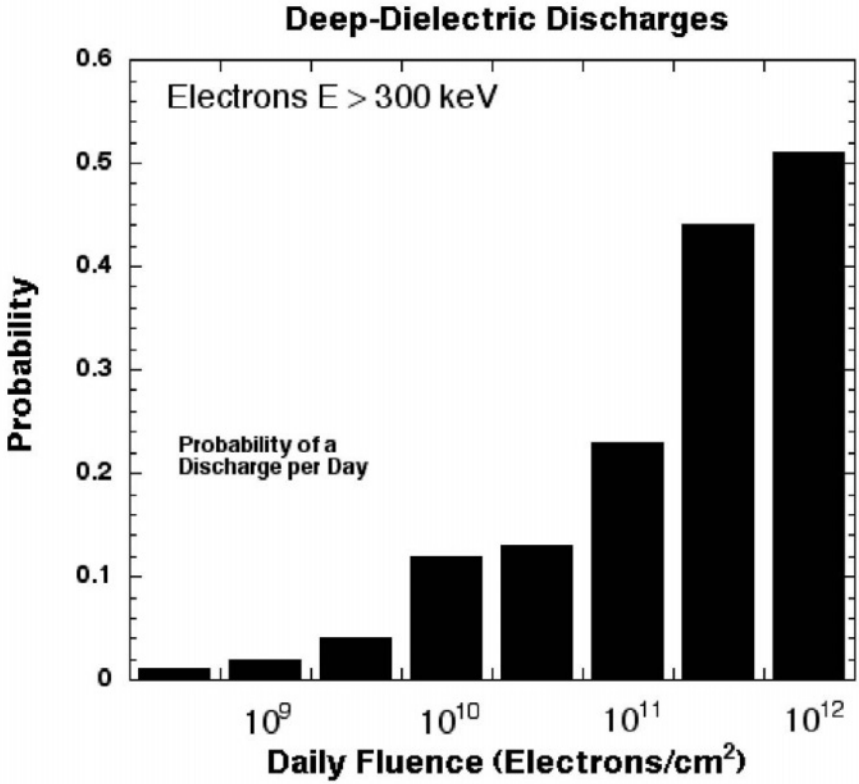

Figure 6. Experimental results from Vampola (1987) showing the probability of observing a dielectric discharge event as a function of the daily-integrated flux (fluence) of electrons with energy E>300 keV (from Baker, 2000).

4. A RADIATION BELT CONTENT INDEX

In recent papers (e.g., Baker et al., 1999; 2001), we have analyzed quite extensively the outer radiation belt electron population and its temporal-spatial variability. We have found that on ~daily time scales, the outer radiation belt exhibits a high degree of coherent behavior in the sense that the entire outer belt increases and decreases in intensity throughout a volume extending from L~2.5 at the inner edge to L~6.5 at its outer edge. This spatial extent of the radiation belt "torus" (2.5≤L≤6.5) can be examined – to first order – as a coherent entity (Baker et al., 2001 and references therein).

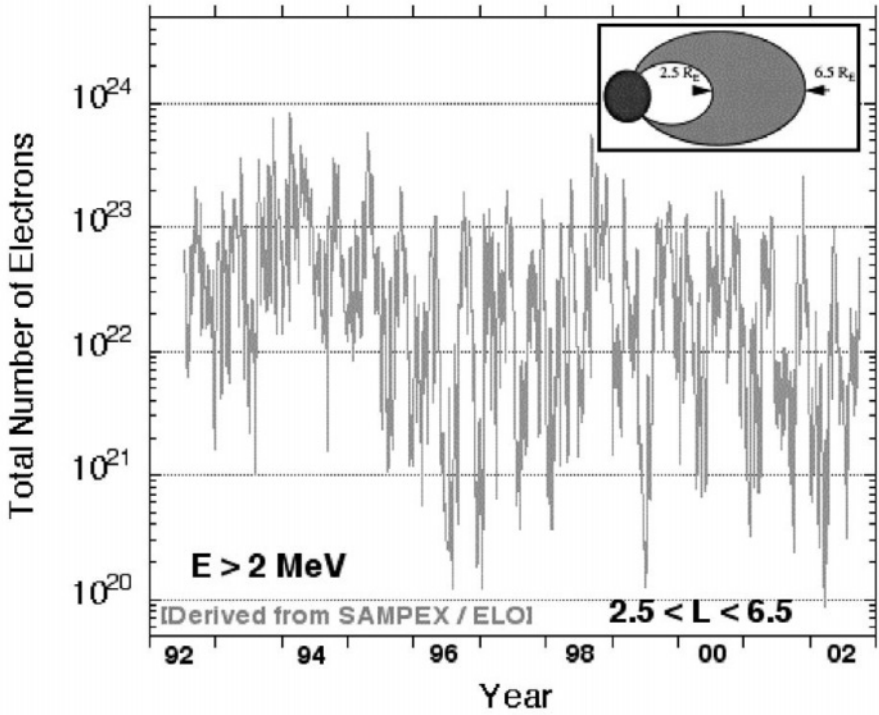

Figure 7. Daily values for the Radiation Belt Content Index as discussed in the text (adapted from Baker et al., 2003).

Using the low-altitude, polar-orbiting spacecraft SAMPEX, Baker et al. (1999) compared radiation belt electron properties with a highly-elliptical orbit spacecraft, POLAR. For a given L-shell, the POLAR and SAMPEX data compared very favorably as long as pitch angle and magnetic flux tube characteristics for the trapped electrons were taken into account. In general, the POLAR measurements showed much higher electron directional intensities since POLAR was measuring relatively close to the magnetic

equator, while SAMPEX was measuring close to the atmosphere (~600 km altitude). Methods were developed (Baker et al., 2001) to estimate the entire outer "radiation belt content" from the SAMPEX measurements.

SAMPEX was launched in July 1992 and has made measurements continuously since then to the present time. We have used the E>2MeV channel onboard SAMPEX to calculate a Radiation Belt Content (RBC) Index on a daily basis (see Baker et al., 2003). This RBC Index is plotted versus time in Figure 7. It shows the estimated total number of electrons (with E>2 MeV) in the outer belt at any particular time.

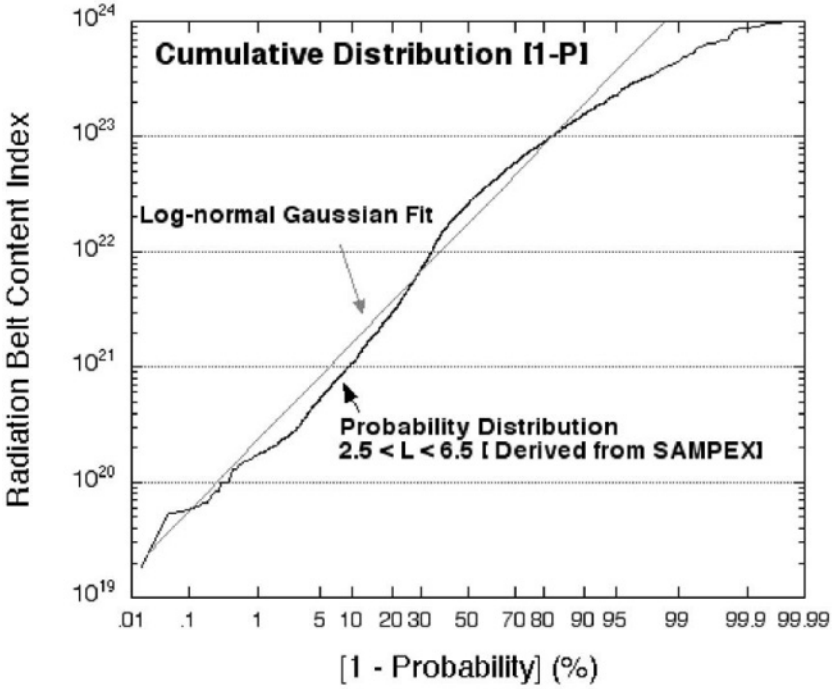

Figure 8. The cumulative probability curve for the Radiation Belt Content Index as discussed in the text (taken from Baker et al., 2003).

Figure 7 shows that the total RBC ranges over some four orders of magnitude during the decade-long record from SAMPEX. The highest content was nearly 10^{24} electrons in late 1993 and early 1994. The lowest RBC Index value was ~10^{20} electrons reached several times (~1996, mid-1999, and early in 2002). Note that the overall daily variance for the RBC Index was small in 1994, but was very large at the time of sunspot minimum (1996-97).

Although the processes occurring in the outer radiation belt are physically complex and spatially varied (see Vassiliadis et al., 2002), we suggest that from a space weather standpoint it is useful to have a simple,

robust index of radiation belt properties such as the RBC. At a glance, this index gives a global assessment of the radiation belt electron properties. The RBC Index allows a long-term evaluation of interesting radiation belt properties.

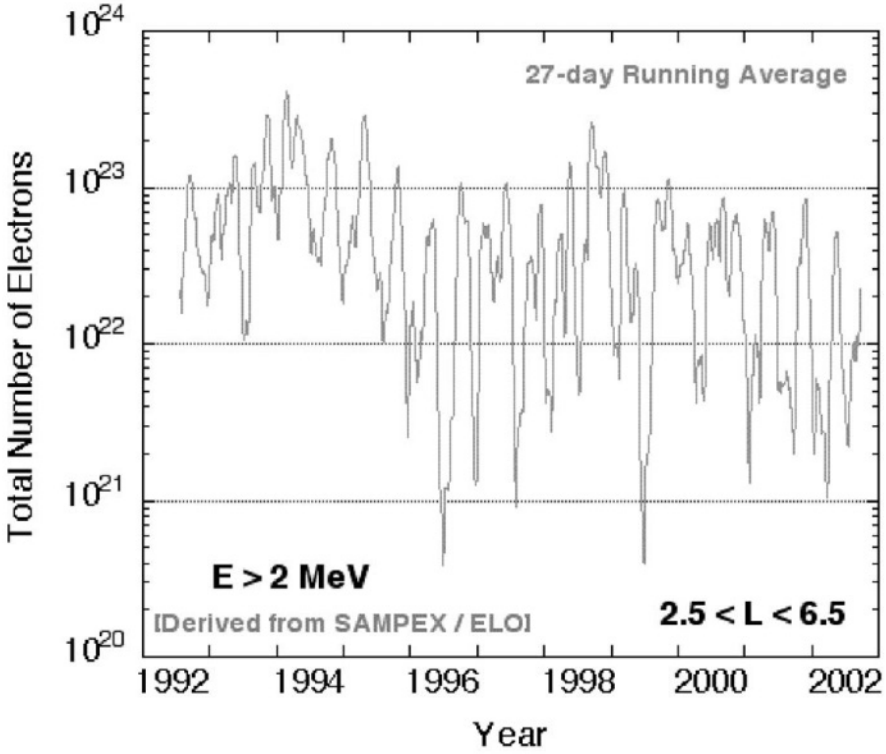

Figure 9. The Radiation Belt Content Index smoothed with a 27-day running average (from Baker et al., 2003).

Figure 8 (taken from Baker et al., 2003) shows one such long-term assessment. Using the daily RBC Index, we have calculated the "cumulative" probability distribution. In other words, we have computed the percentage of the time that the RBC is greater than a given value (we actually plot (1-probability) in Figure 8). It is seen that the probability is extremely high (p>99.99%) that the RBC will exceed 10^{19} electrons and it is very rare (p<0.01%) that the RBC will exceed 10^{24} electrons. We have plotted a straight-line log-normal Gaussian fit to the data in Figure 8. This gives a reasonably good fit to the data except at the highest RBC Index values. The "saturation" at the highest electron content may be a nonlinearity in the SAMPEX sensor capabilities, or it may be a natural limit to the radiation belt trapping. This is being examined further.

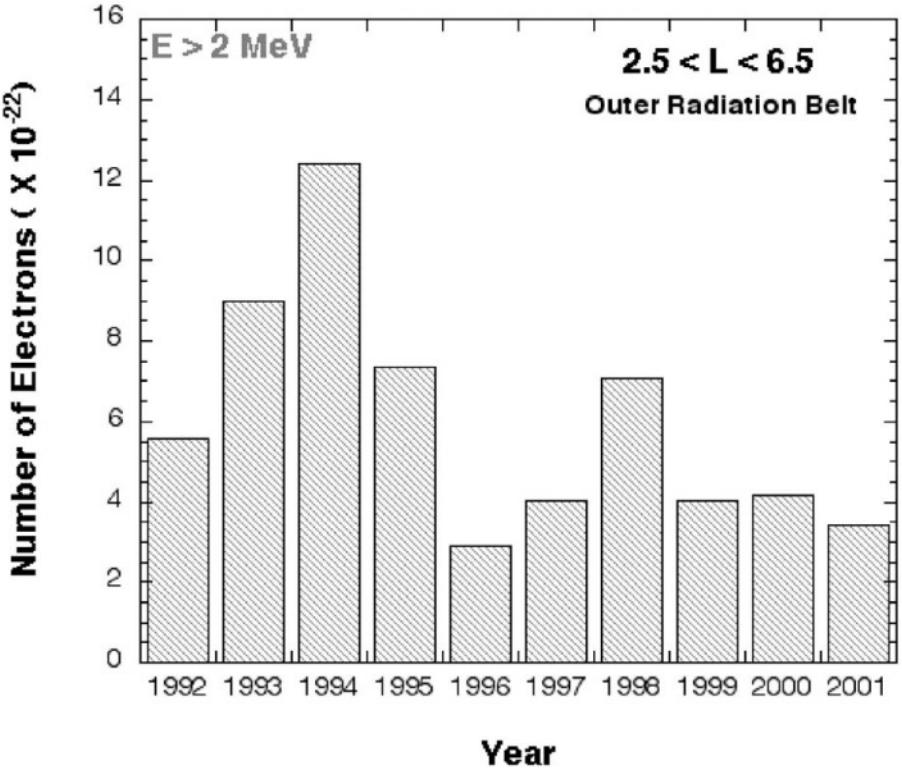

Figure 10a. The calculated number of electrons with energies E>2 MeV in Earth's outer radiation belt (2.5<L<6.5) for each year from 1992 to 2001 (adapted from Baker et al., 2001).

Another useful procedure for the RBC Index is to smooth the data to see long-term trends more clearly. In Figure 9, for example, we have computed 27-day running averages of the RBC Index (see Baker et al., 2003). This analysis emphasizes the point that 1993-94 was clearly the time of highest average radiation belt electron intensities. This degree of smoothing also points at important annual and seasonal trends.

Figure 10a shows the computed annual averages of electrons (E>2MeV) from 1992 to 2001 (an extension of results shown in Baker et al. (2001)). There is a very prominent maximum in 1994 (which extends broadly from 1993 to 1995). This is the time of the approach to sunspot minimum when solar coronal holes were prominent. Coronal holes give rise to high-speed solar wind streams and these obviously drive high-energy electron acceleration within the Earth's magnetosphere (Baker et al., 2001).

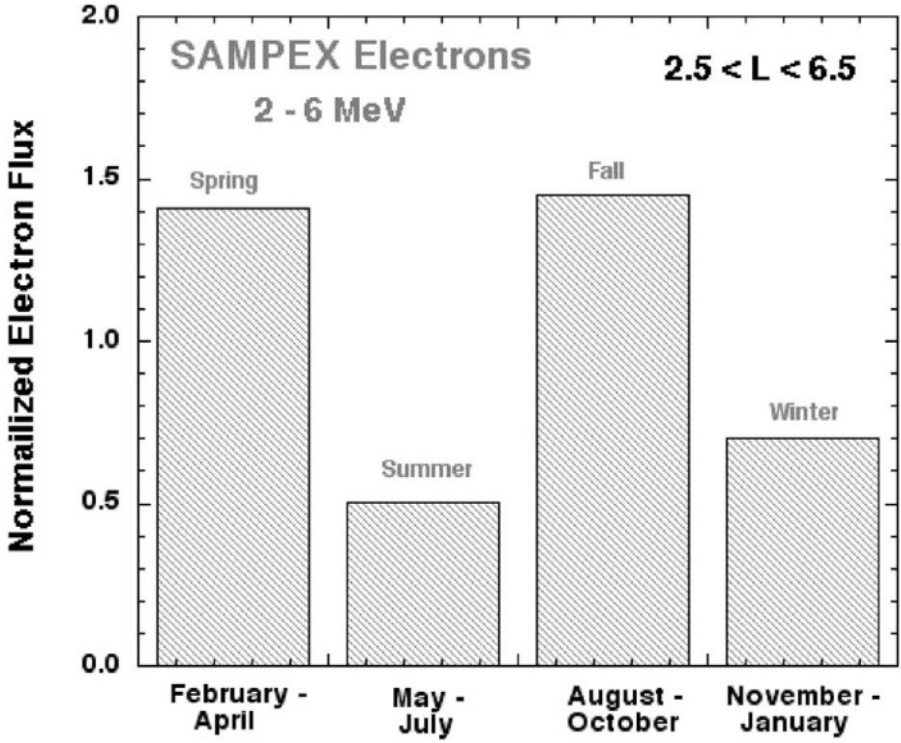

Figure 10b. A superposed analysis of seasonal fluxes of E>2MeV electrons throughout the outer radiation belt (2.5≤L≤6.5) for 1992-1999. A strong seasonal variation of electron fluxes is seen (from Baker et al., 1999).

Another feature in Figure 9 that is quite evident is the occurrence of two peaks in the smoothed RBC Index each year. These occur around each spring and fall equinox. Baker et al., (1999) analyzed this "semiannual" effect in the outer radiation belt electron content. Figure 10b shows the strong tendency for electron fluxes during the spring and fall seasons (Feb-Apr and Aug-Oct, respectively) to be about a factor of three higher than the fluxes near the summer and winter solstices (see, also, Li and Temerin, 2001). The superposed (normalized) results of Figure 10b are also quite obvious by visual inspection of Figure 9.

From a space weather point of view, the results of Figures 7, 9, and 10 are quite important. The RBC Index provides a simple, robust estimator of the entire outer radiation belt properties. From it one can assess radiation belt features on daily (Figure 7), solar-rotation (Figure 9), seasonal Figure 10b), and annual (Figure 10a) timescales. From all of these data (Figures 7, 9, and 10a), one can also see the solar cycle behavior of the radiation belts.

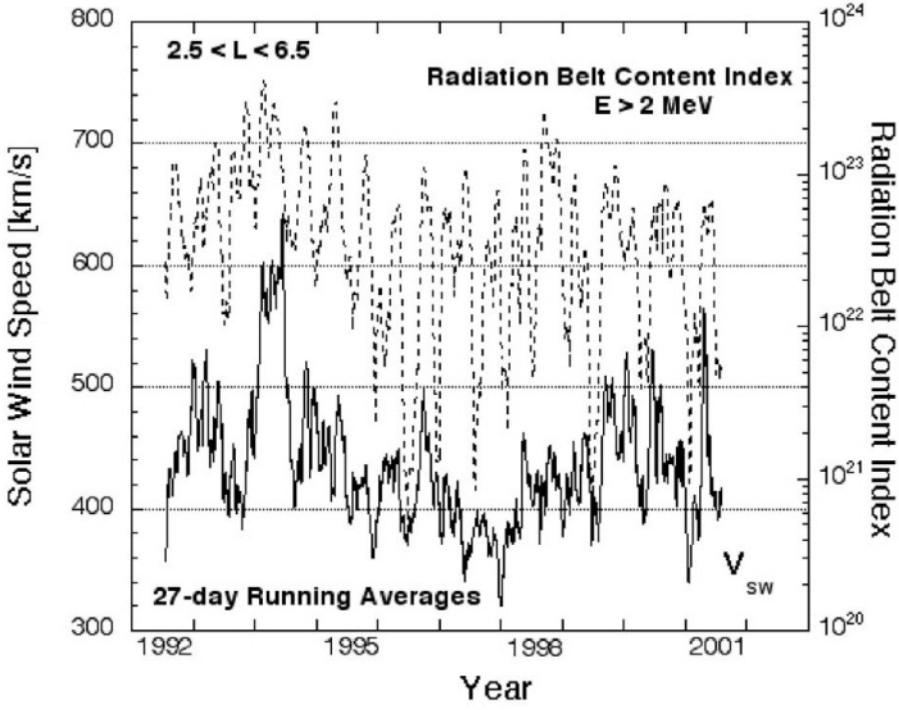

Figure 11a. Comparison of the Radiation Belt Content Index with concurrent solar wind speed (V_{SW}) for the years 1992-2001 (from Baker et al., 2003).

5. SOLAR WIND DRIVING OF THE RADIATION BELTS

As is well established by past work (e.g., Baker et al., 1987; Li and Temerin, 2001) and as alluded to above, the outer radiation belt energetic electrons are rather clearly accelerated within the magnetosphere by the action of high-speed solar wind. The Radiation Belt Content Index is useful for dramatizing that point.

Figure 11a (from Baker et al., 2003) shows the RBC Index, plotted again as a 27-day running average from 1992 to 2001 (upper curve). Plotted below this is the 27-day running average of the wind speed, V_{SW}. (The value of V_{SW} is obtained from the NASA omnitape data set). It is quite striking that the running-averaged values of V_{SW} were significantly greater than ~500 km/s only in 1994. This was also the time of the highest radiation belt electron content as well. As is evident, the smoothed values of V_{SW} do not show a very prominent seasonal increase (nor would we expect that), even though

the RBC Index shows this seasonal effect prominently. Obviously, internal magnetospheric acceleration is key to radiation belt electron enhancements.

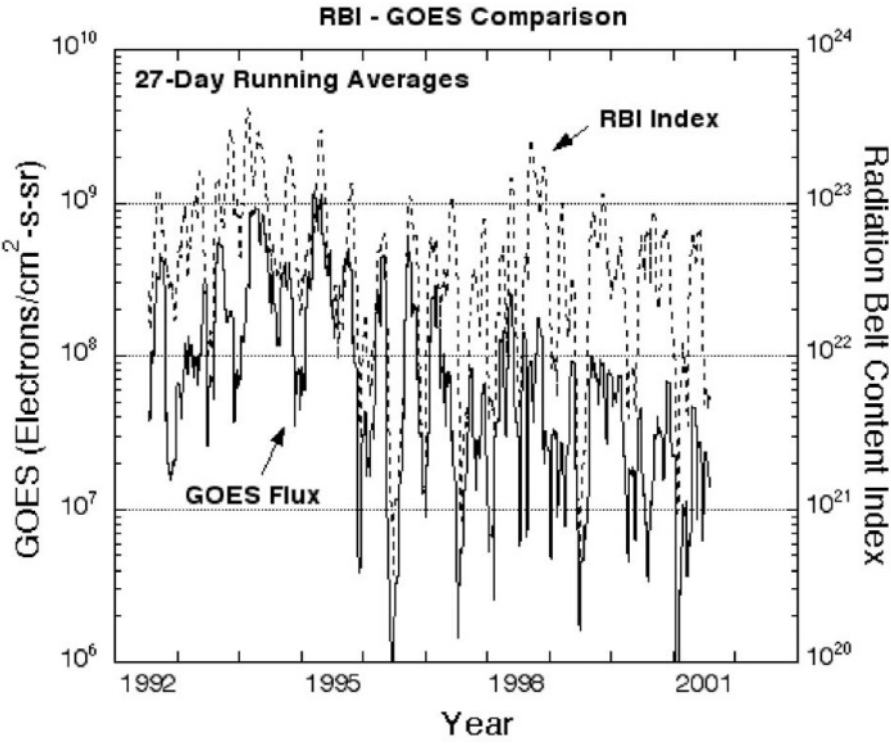

Figure 11b. Similar to (a) comparing RBC Index to concurrent GOES geostationary orbit data (from Baker et al., 2003).

We expect the RBC Index to be a useful comparator to other commonly used indicators of radiation belt behavior. Figure 11b, for example, overplots GOES data with the RBC Index data (27-day smoothed in both cases) (see Baker et al., 2003). Note that the RBC and GOES scale almost perfectly with one another in 1992-1997. Then, for some reason the two data sets diverge and show very large differences in 1998-2001. We are exploring the reasons for this divergence, but it may involve differences in the properties of E>2 MeV electrons which are produced by high-speed streams (1993-1995) and those produced mostly by CMEs (1998-2001).

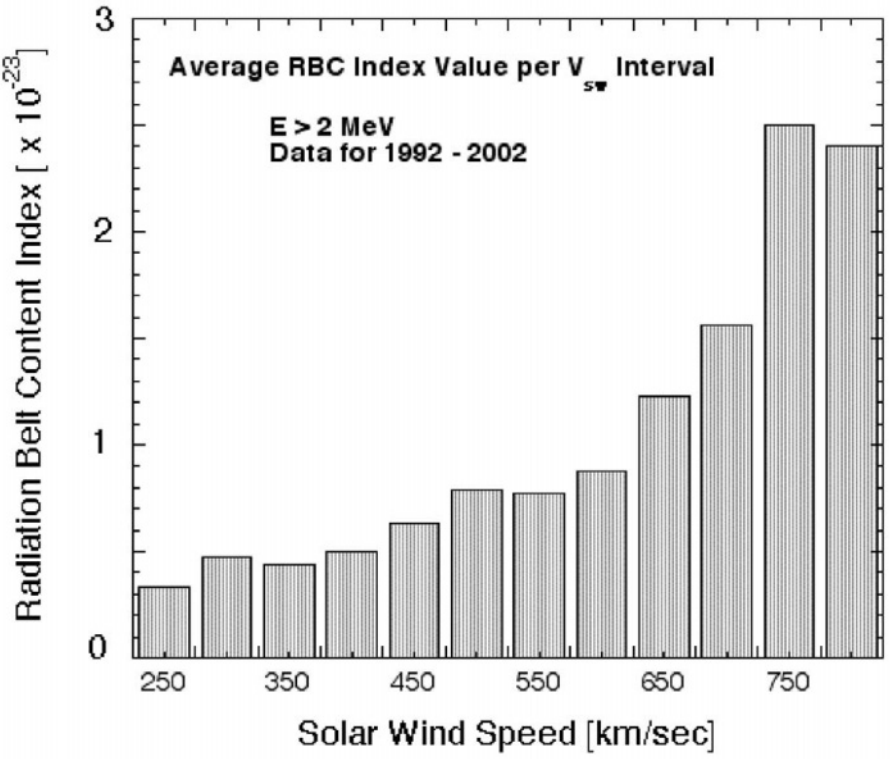

Figure 12. Average Radiation Belt Content index values versus solar wind speed for the period 1992-2002.

It is interesting to look at the dependence of the RBC index on solar wind speed in a more average sense than portrayed in Figure 11a. In Figure 12, we plot the histograms which show the average values of the RBC index for every 50 km/s interval of solar wind speed from 250 km/s to 800 km/s. Data were used for the period 1992-2002. We see from Figure 12 that there was a progressive increase of the average RBC value with increasing solar wind speed. For V_{SW} in the 700-750 km/s range, the average RBC value was nearly ten times greater than for V_{SW} in the 250-300 km/s range.

Finally, Figure 13 (taken from Baker et al. (2001)) shows another use for the RBC Index. By spectrally extending the Index downward in energy to E=0.5 MeV, we are able to calculate the entire relativistic electron energy content in the outer belt (2.5<L<6.5). If we calculate this relativistic electron power and compare to the concurrent solar wind input power to the magnetosphere, we get the electron accelerator "efficiency". As shown by the figure, Baker et al. (2001) determined this efficiency to be ~0.1% for the years 1993-1998.

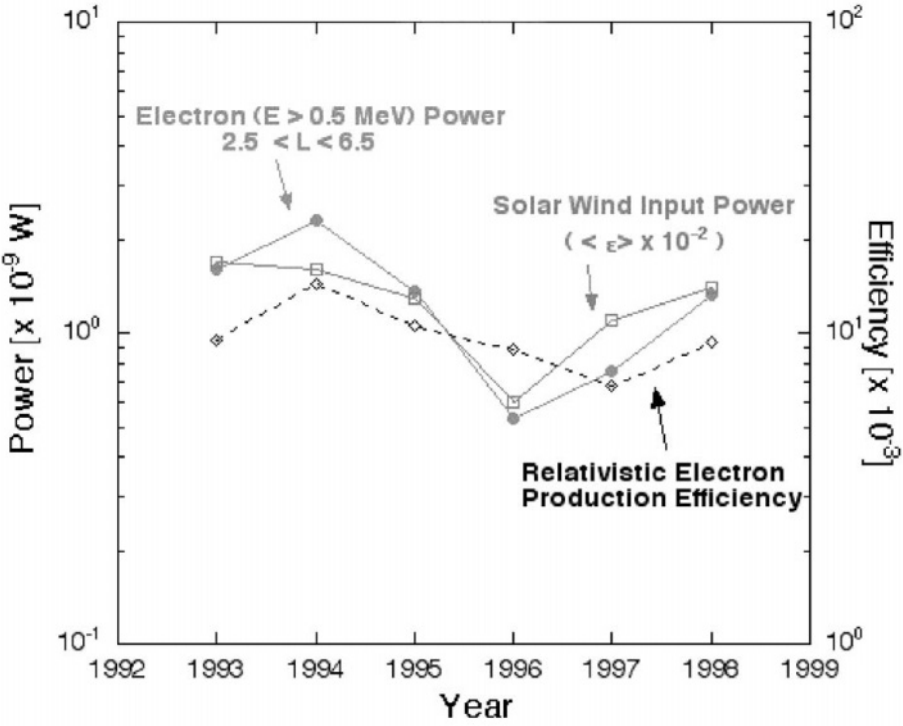

Figure 13. The annual values of solar wind input power (open squares) and magnetospheric relativistic electron power (solid circles) for the years 1993-1998 (scale to the left). Also shown by the diamonds and dashed curve is the calculated relativistic electron production efficiency (defined in the text) with the scale shown to the right (from Baker et al., 2001).

6. MAGNETOSPHERIC SUBSTORMS

A significant effect of moderate geomagnetic activity ("magnetospheric substorms") from the standpoint of space operations is the occurrence of spacecraft surface charging (see Rosen, 1976). During a surface-charging event, insulated regions on a spacecraft may charge to several kilovolts potential (usually negative relative to the ambient potential). This charging occurs because of a lack of current balance between the ambient plasma medium and the spacecraft surface (as illustrated in Figure 14). When a spacecraft is immersed in a cool, dense plasma, the incident particles (electrons and ions) as well as secondary emitted particles, photoelectrons, and backscattered electrons, all balance. This gives a low net spacecraft potential. However, in a very hot, tenuous plasma, current balance can be difficult to achieve and large potentials can build up.

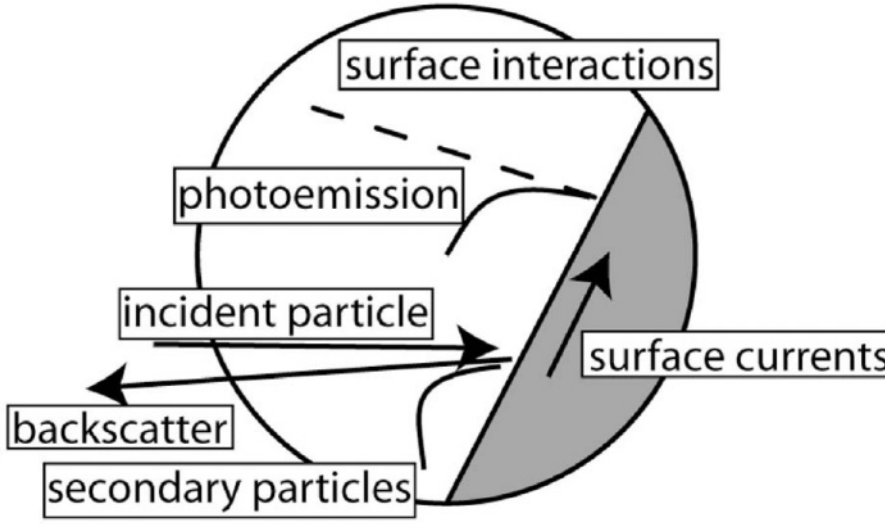

Figure 14. Surface charging mechanism for spacecraft materials (adapted from Robinson, 1989).

Figure 14 shows the interactions at the surface of a spacecraft. This points out that there are currents near the surface of the spacecraft due to incident, backscattered, and photo-emitted particles. These can, in principle, be examined to calculate the charge configurations for a particular spacecraft. A sheath region that forms around the spacecraft is a volume strongly affected by the satellite itself. The plasma there is distorted by electric fields due to the distributed charges. The sheath region can also be affected by operational activity such as thruster firings. These can extend the influence of the spacecraft far into the plasma volume (e.g., Robinson, 1989). The sheath is complex in shape and depends on the motion of the spacecraft through the plasma as well as the plasma properties and the surface materials of the vehicle. From an operational standpoint, differential charging of satellite sur faces can lead to significant discharges.

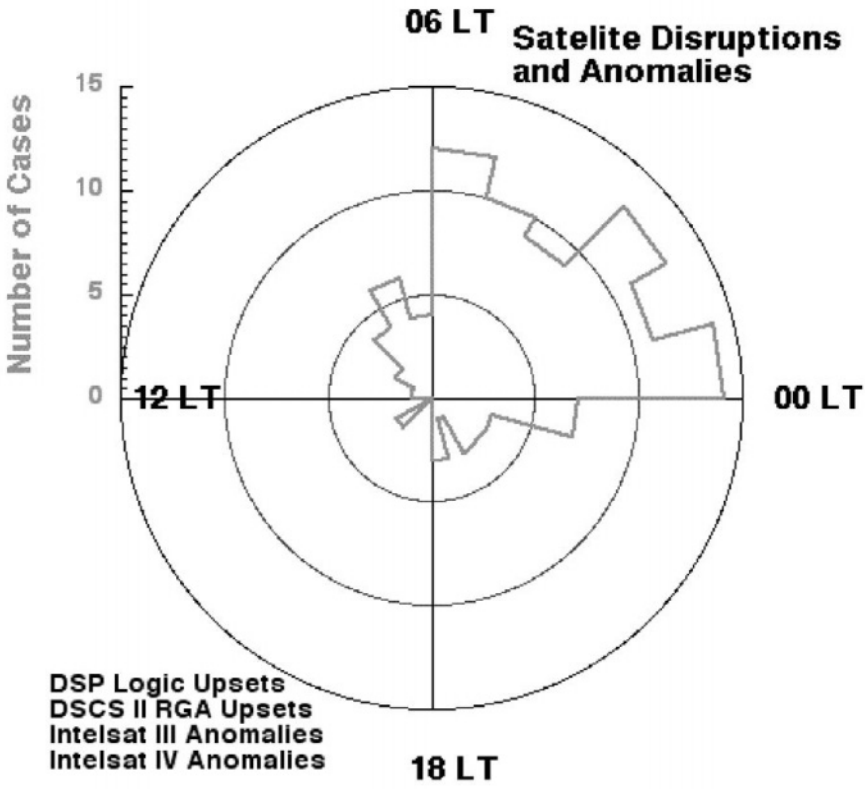

Figure 15. Local time distribution of satellite disruptions and anomalies showing a strong occurrence frequency peak in the midnight and local morning hours (data from Rosen, 1976).

Discharges introduce noise into subsystems and may interrupt normal spacecraft operations or represent a false command. In the process of discharge breakdown, physical damage may occur. This may change the physical characteristics (thermal properties, conductivity, optical parameters, etc.) of the satellite. Furthermore, the release of material from the discharge site has been suggested as a contamination source for the remainder of the vehicle (see Baker, 1998 and references therein).

Figure 15, adapted from data presented in Rosen (1976), shows the number of spacecraft anomalies detected at geostationary orbit as a function of spacecraft local time. The anomalies include logic upsets as well as other significant operational problems both for military (DSP and DSCS) and commercial (Intelsat) spacecraft. As may be seen, there is a strong local time asymmetry in the number of anomalies with the majority occurring between local midnight and local dawn. This is where substorm-injected electrons are seen most prominently (e.g., Baker, 1998, and references therein) and the LT

distribution shown in Figure 15 supports the view that surface charging has constituted a major cause of operational anomalies near geostationary orbit.

7. MODELING THE SPACE WEATHER ENVIRONMENT

The Earth's surface is encircled by communication links, power grids, and numerous new technological systems that did not even exist a decade ago. Considering the range of satellites orbiting the Earth from low to high altitudes, it is evident that there is effectively a "cyberelectric" cocoon that wraps around the Earth. Most elements of this web are susceptible in one way or another to space weather effects (Baker, 2002). Clearly, modern communication systems rely heavily on elements than include both ground links and satellite links. Recent assessments have shown that world-wide communication systems can be detrimentally affected by adverse space weather (Lanzerotti, 2001; Singer et al., 2001). The failure and loss of even one key communication satellite – as occurred on May 19, 1998 with the Galaxy IV satellite failure – can affect millions or tens of millions of customers relying on telephones, pagers, and other communications technologies (Baker et al., 1998).

The first defense for human technology systems against the effects of space weather is to build robust systems that readily withstand space weather disturbances. To a large extent, this has already been done. If not, there would be many more space weather-induced failures than presently are seen. Ground communications links, national power grids, and military installation – which must all withstand hurricanes, earthquakes, and floods – are very resilient and robust systems. Also, it is evident that today there are many hundreds of satellites in Earth orbit fulfilling a wide variety of military and civilian purposes. Few of these fail catastrophically due to space weather. On the other hand, some spacecraft do fail suddenly due to space weather effects and nearly all spacecraft eventually fail due to the rigors of the hostile space environment. Thus, we need to know more about the nature of space weather elements, we need to specify better what the space environment is at any point in space, and we ultimately want to be able to predict (i.e., forecast) what the space weather environment will be anywhere in Earth's neighborhood many hours or days in the future. This is the goal of the U.S. National space Weather Program (NSWP) (National space weather Program Strategic Plan, 1995; Robinson and Behnke, 2001).

Space weather has become a major unifying theme and a uniting force for the entire solar-terrestrial research community. Understanding and predicting such events is a challenge of great scope and complexity (Singer et al., 2001). The National Aeronautics and Space Administration (NASA) has

now undertaken a major new initiative called "Living With a Star" (LWS) to observe systematically the disturbances arising on the Sun and to follow these space weather drivers all the way to their ultimate dissipation in Earth's atmosphere. The National Science Foundation (NSF) has also been a leading agency in the development of the National Space Weather Program (Robinson and Behnke, 2001). The NSF has now selected a consortium of universities, industry partners, and national laboratories to form a Science and Technology Center dedicated to space weather. This "Center for Integrated Space-Weather Modeling" (CISM) is funded at several million dollars per year for the next 5-10 years and will have as its goal the building of physics-based models all the way from the Sun to the Earth's atmosphere. It involves numerous institutions all around the U.S., and it works closely with the National Oceanic and Atmospheric Administration (NOAA).

A starting point for CISM efforts has been to emplace a sequence of solar-terrestrial models to capture the variability in the connected Sun-Earth system. This requires coupling together empirical and semi-empirical models to describe the end-to-end Sun-Earth plasma chain. Figure 16 illustrates the Phase I CISM models. The effort starts (as shown to the left) with solar observations to provide "solar boundary measurements". These starting conditions are used in the Wang-Sheeley method, as modified by Arge (see Arge, 2000; Wang et al., 2002), to propagate the solar wind to the Earth's vicinity. The Wang-Sheeley-Arge (WSA) method gives information about solar wind speed and some information about IMF properties. This forecasted set of solar wind properties can then be used as a "driver" for empirical models that predict geomagnetic indices (such as Ap or Dst), or

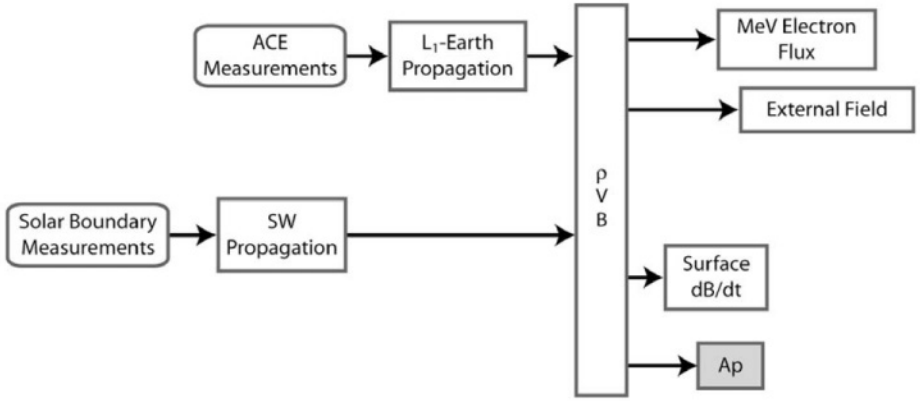

Figure 16. A schematic diagram showing end-to-end linked solar terrestrial models intended to provide forecasting capability in the CISM program of the NSF (see text for details).

radiation belt electron properties, or even localized variations in surface magnetic field fluctuations. As shown in Figure 16, the empirical models can

also be driven (with shorter predictive lead times, but higher confidence factors) by using solar wind (IMF) data from the ACE spacecraft at the upstream Lagrange point, L1. The WSA method can give 2-3 day forecasts, while the L1 inputs give 1 hour forecasts.

CISM is a new (and very important) part of a broader space weather model development program in the U.S. As shown by Figure 17, at its base the U.S. has many NASA, DoD, and NSF programs and projects funded within the space weather research community. These projects develop modeling ideas that need to be evaluated, tested, and validated as effective space weather tools. These models can be evaluated in relatively objective ways (as at the Community Coordinated Modeling Center (CCMC), at NASA's Goddard Space Flight Center). Models can also be tested and documented at NOAA's or DoD's Rapid Prototyping Centers (RPCs). Those models that are most useful, robust, and ready for operational application can then be moved into the user realm. The transition from research to applications is one of our biggest space weather challenges (Singer et al., 2001).

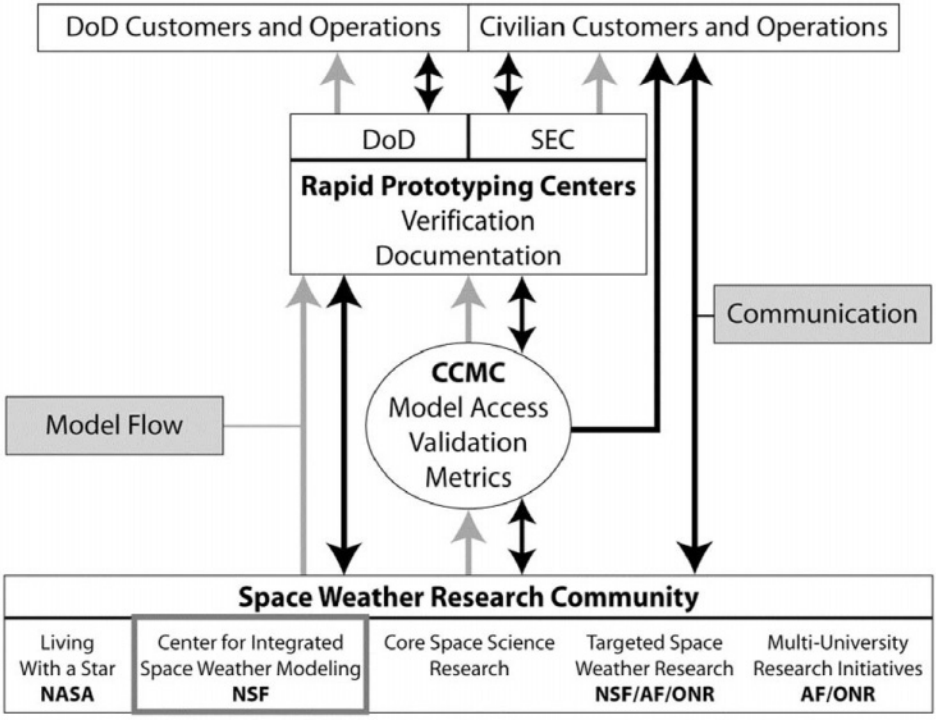

Figure 17. A diagram showing the flow of models from the space research community (bottom) to the DoD and civilian user communities (top).

8. ACKNOWLEDGMENTS

This work was supported by the NSF Space Weather Program and by the NSF Center for Integrated Space Weather Modeling (CISM). Research was also supported in part by NASA. The author thanks S. Kanekal and R. Weigel for major contributions to this paper.

9. REFERENCES

Akasofu, , S.-I., Energy coupling between the solar wind and the magnetosphere, Space Sci. Rev., 28, 121, 1981.

Arge, C.N., and V.J. Pizzo, Improvements in the prediction of solar wind in the prediction of solar wind conditions using near-real time solar magnetic field updates, J. Geophys. Res., 105, 10,465, 2000.

Baker, D.N., et al., Deep dielectric charging effects due to high energy electrons in earth's outer magnetosphere, J. Electrostatics, 20, 3, 1987.

Baker, D.N., What is space weather?, Adv. Space Res., 23, 1,7,1998.

Baker, D.N., et al., Equinoctial and solstitial averages of magnetospheric relativistic electrons: A strong semiannual modulation, Geophys. Res. Lett., 26, 20, 3193-3196, 1999.

Baker, D.N., The occurrence of operational anomalies in spacecraft and their relationship to space weather, IEEE Trans. on Plasma Sci., 28, 6, 2000.

Baker, D.N., et al., The global efficiency of relativistic electron production in the Earth's magnetosphere, J. Geophys. Res., 106, A9, 19,169-19,178, 2001.

Baker, D.N., How to cope with space weather, Science, 297, 30 Aug 2002.

Baker, D.N., S.G. Kanekal, and J.B. Blake, Characterizing the Earth's outer Van Allen zone using a radiation belt content (RBC) index, Space Weather, in press, 2003.

Gosling, J.T., et al., Geomagnetic activity associated with Earth passage of interplanetary shock disturbances, J. Geophys. Res., 96, 7831, 1991.

Lanzerotti, L.J., in Space Weather, (Geophys. Monograph 125), p. 11, Amer. Geophys. Union, Washington, DC, 2001.

Li, X., and M.A. Temerin, The electron radiation belt, Space Sci. Rev., (1-2) 95, 569-580, 2001.

National Space Weather Program Strategic Plan, Office Fed. Coord. For Met. Services, NOAA, Silver Spring, MD (1995).

Reagan, J.B., et al., Space charging currents and their effects on spacecraft systems, ISEE Trans. Elec. Insul., E1-18, 354,1983.

Robinson, P.A., Jr., Spacecraft environmental anomalies handbook, JPL Report GL-TR-89-0222, Pasadena, CA,1989.

Robinson, R.M., and R.A. Behnke, in Space Weather, (Geophys. Monograph 125), p. 1, Amer. Geophys. Union, Washington, DC, 2001.

Rosen, A. (editor), Spacecraft charging by magnetospheric plasmas, AIAA, 47, New York, 1976.

Singer, H.J., et al., in *Space Weather*, (*Geophys. Monograph 125*), p. 23, Amer. Geophys. Union, Washington, DC, 2001.

Vampola, A.L., The aerospace environment at high altitudes and its implications for spacecraft charging and communications, *J. Electrostat., 20*, 21, 1987.

Vassiliadis, D., et al., Long-term-average, solar cycle, and seasonal response of magnetospheric energetic electrons to the solar wind speed, J. Geophys. Res., 107, A11, 1383, 2002.

Wang, Y.-M, N.R. Sheeley, Jr., and M.D. Andrews, Polarity reversal of the solar magnetic field during cycle 23, J. Geophys. Res., 107, A12, 1465, doi:10.1029/2002JA009463, 2002.

Wrenn, G.L., Conclusive evidence for internal dielectric charging anomalies on geosynchronous communications spacecraft, *J. Spacecraft and Rockets, 32*, 514, 1995.

Chapter 2

Geospace storm dynamics

Ioannis A. Daglis
Institute for Space Applications and Remote Sensing, National Observatory of Athens
Penteli, 15236 Athens, Greece

Abstract Geospace storms, also known as space or magnetic storms, interconnect the Sun and interplanetary space with the terrestrial magnetosphere, ionosphere, and atmosphere – and often even the surface of the Earth – in a uniquely global and synergistic manner. Energy from the Sun drives a continuous interaction of these distinct but coupled regions. Geospace storms have traditionally been called geomagnetic storms, because of the defining feature of global geomagnetic field disturbances that they induce. However, observations over four decades of space-borne instrumentation have shown that storms involve more than just variations in the geomagnetic field: they involve acceleration of charged particles in the magnetosphere, modification of the electrodynamic properties of the ionosphere, heating of the upper atmosphere, and creation of geomagnetically induced currents on the ground. This chapter attempts a synoptic discourse of geospace magnetic storm history, the classical perception of magnetic storm dynamics, and deviations from long-time accepted paradigms. In particular, we review in some detail one of the critical issues of storm dynamics, namely the storm-substorm relationship.

Keywords Geospace storm, magnetic storm, ring current, radiation belts, geospace, magnetosphere-ionosphere coupling, particle acceleration, space-atmosphere coupling, space weather, space hazards.

1. INTRODUCTION AND HISTORICAL CONTEXT

The first priority of NASA's Living With a Star Geospace Mission (Kintner et al., 2001) is to understand the acceleration, global distribution and variability of energetic electrons and ions in the inner magnetosphere. Given that the most distinct result of geospace storms in the near-Earth space environment is the intensification of the radiation belts and of the ring current, geospace storms are an object of special interest within the Living With a Star Program.

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 27-42.

The geospace storm is the most complex collective phenomenon in the near-Earth space environment. It encompasses a large number of physical processes and effects in near-Earth space environment: Acceleration of charged particles in space; intensification of electric currents in space, in the upper atmosphere and on the ground; impressive aurora displays, which expand equatorwards; global magnetic disturbances on the Earth surface, which have actually been the defining storm feature and the origin of the classical denomination "magnetic storms". Despite their complexity, geospace storms have been identified by a rather simple pattern that is imposed by their development on the time profile of the magnetic disturbances measured on the ground (see Figure 1 and discussion in section 2).

At this point I consider it noteworthy to remark on names, because I disagree with the notion "nomina nuda tenemus" (i.e., "we hold naked [empty] words", Bernard de Morlaix, in *De contemptu mundi*, 12th century). A few years ago I had suggested and had used the term "space storm" (Daglis, 1997b; 1999a; 1999b; 2001a), but I was criticized that "space storm" sounds too general and non-specific, or simply trendy. More recently I proposed in a forum article in Eos (Daglis, 2003) the expression "geospace storm" instead of "magnetic storm". Accordingly, this chapter uses the term "geospace storm".

The eminent German explorer Alexander von Humboldt was probably the first to use the expression "magnetic storms" for the definition of an intense geomagnetic phenomenon. However, von Humboldt was not the father of the magnetic storm concept; he used the term to describe time intervals of intense magnetic fluctuations rather than a prolonged worldwide weakening of the horizontal component, *H*, of the geomagnetic field. This is, of course, expectable, as there was no way for him to know about worldwide negative *H*-excursions. In his letter to Prof. Paul Erman, published in *Annalen der Physik* (von Humboldt, 1808), von Humboldt had described a night of impressive observations in Berlin during the night of December 20-21, 1806, remarking that "there was no magnetic storm", since "the (magnetic) fluctuations were not especially intense" ("Dabei fand kein magnetisches Ungewitter statt; die Schwankungen waren nicht besonders stark").

Even in the 19th and early 20th century, scientists used "magnetic storm" for periods of intense geomagnetic variations in general, and not for what we identify as magnetic storms nowadays. Birkeland (1908), for example, used the term "magnetic storm", to collectively describe five distinct types of magnetic perturbations (details by Chapman and Bartels, 1940).

The ground manifestation of geospace storms as we perceive it today emerged from a discussion of the long series of Bombay, India, magnetic

data by the Indian scientist Dr. Nanabhoy Ardeshir Framji Moos, Director of the Colaba-Alibag Observatories. Although it had generally been known (e.g. Broun, 1861; Adams, 1892) that for some time after a period of great geomagnetic disturbance the *H*-component of the geomagnetic field is reduced below its mean value, this critical information became much more complete by the work of Moos, which was decisive for the identification of the now familiar storm pattern in the time variation of *H* (Moos, 1910).

Later Sydney Chapman, one of the great pioneers and founders of modern solar-terrestrial research, who led much of the early work on magnetic storms, applied Moos' methods to study the average features of moderate storms at many stations in different geographic latitudes. Chapman (1919) demonstrated the global aspect of magnetic storms and named the storm-time variation of *H* "*Dst* variation" (meaning "Disturbance Storm-Time"). The characteristic average variation of *Dst* led Chapman to regard the storm geomagnetic variations as a unity, with a beginning, middle and end. In fact, Chapman was the one who combined the method and the name in his seminal statistical work and established the present concept of magnetic storm (S.-I. Akasofu, personal communication).

The foundations of modern geospace storm research were laid by Chapman and Ferraro (1930, 1931), who proposed a transient stream of outflowing solar ions and electrons to be responsible for terrestrial magnetic storms. Chapman and Ferraro claimed that once the solar stream had reached the Earth, charged particles would leak into the magnetosphere and drift around the Earth, creating a current whose field would oppose the main geomagnetic field. This is astoundingly close to what we believe today. The only major element of Chapman's theory that has changed is the existence of a continuous - instead of transient - stream of ionised gas from the Sun. This stream was christened "solar wind" by Eugene Parker (Parker, 1958) and its existence was later confirmed by measurements performed by the Venus-heading Mariner 2 spacecraft (Neugebauer and Snyder, 1962).

The basic idea of the Chapman and Ferraro theory was that the physical reason for the magnetic perturbation on the Earth's surface is a huge "ring current" in space circling the Earth. This idea was further elaborated by Singer (1956, 1957) and was eventually confirmed by in situ spacecraft measurements. The first measurements in space were conducted by Geiger Mueller tubes of the group of James Van Allen on board the first Explorer satellites in the end of the 1950s. Van Allen interpreted those measurements as the result of intense corpuscular radiation (Van Allen et al., 1958; Van Allen, 1959).

The ability of the geomagnetic field to trap relativistic electrons was experimentally verified by the Argus experiment, which was proposed by Nicholas C. Christofilos in 1957 and carried out in 1959 (Christofilos, 1959).

Christofilos, an unconventional Greek scientist who had been working as an engineer designing elevator systems in Athens before migrating to the US in 1953, had actually communicated to the US Army in the early 1950s that many charged particles, due to the dipole magnetic field, could be trapped around the Earth. He further proposed that an artificial radiation belt, due to beta decay, could be created by exploding a nuclear bomb at high altitude. This proposal evolved into Argus - the first active experiment in space.

2. THE CLASSICAL PICTURE AND PARADIGM SHIFTS IN GEOSPACE STORM DYNAMICS

During the two decades that followed the dawn of space exploration, experimental data led to a rough morphological model of geospace storms, which was in general agreement with the theoretical postulations of Chapman and Singer.

In the past, the solar antecedents of storms were thought, and are still sometimes erroneously considered, to be strong solar flares. The obvious reason is that the observability of flares permitted their identification and connection to geomagnetic storm disturbances as early as Richard Carrington's solar flare observations during the superstorm of September 1859 (Carrington, 1863). In the 1990s however, Gosling (1993) questioned this paradigm. He argued on the basis of accumulated observational indications that the coronal mass ejection (CME) is the solar event, which is the origin of large geospace storms.

Nevertheless, not every CME leads to a storm in geospace (Tsurutani, 2001). The decisive interplanetary condition for a storm to develop is a prolonged, southward-directed IMF. Russell et al. (1974) had suggested threshold values of $B_z \leq -5$nT and $\Delta T \geq 2$ hours for the development of moderate storms with -100 nT $<$ peak *Dst* ≤ -50nT. Gonzalez and Tsurutani (1987) have empirically found that interplenatary disturbances with $B_z \leq -10$nT and $\Delta T \geq 3$ hours lead to intense storms with peak *Dst* < -100nT.

Even impulsive solar events, like great solar flares, or large CMEs, are not geoeffective if the IMF does not turn southward near 1 AU to permit magnetic reconnection at the dayside magnetopause. A characteristic example is the superflare of November 4, 2003, which was classified as an X28 flare and became the most powerful in recorded observational history (Simpson, 2003).

The classic graphical representation of measurable geospace storm effects on the surface of Earth, is the time profile of the *Dst* index - a geomagnetic index commonly used as a measure of storm intensity (Figure 1). The index

is produced from data provided by low-latitude ground magnetometers, which record the decrease of the horizontal component of the geomagnetic field due to the westward-flowing ring current (Sugiura, 1964).

In accordance with Singer's and Chapman's ideas about the reason for the magnetic perturbations on the Earth's surface, the *Dst* index was conceived as a ring current measure. The concept was based on the assumption that the global decrease of the geomagnetic *H*-component is solely due to an external westward electric current system (the ring current), which encircles the Earth symmetrically (Akasofu and Chapman, 1961). This paradigm has been questioned both by spacecraft observations and by simulations: the storm-time ring current is often highly asymmetric in the main phase and becomes symmetric only in the late recovery phase (e.g., Kozyra et al., 2002; Daglis et al., 2003).

The general morphology of a geospace storm in terms of *Dst* variations is shown in Figure 1: a relatively sharp and large decrease of *Dst* signifies the "main phase" of the storm, and the subsequent slow increase of *Dst* marks the storm recovery. Some storms, especially the largest ones, begin with a sudden impulse (positive excursion of *Dst*), which marks the arrival of an interplanetary shock.

The *Dst* index is widely used to monitor and predict magnetic storm activity and therefore attracts special attention. As mentioned, the original assumption and corresponding paradigm was that *Dst* is influenced only by the ring current fluctuations. Today this paradigm is also under question. The prevailing perception is that there are other magnetospheric currents (cross-tail current, substorm current wedge, magnetopause current, Birkeland field-aligned currents), which also fluctuate during geospace storms and influence the ground magnetic field and, consequently, the *Dst* index (Liemohn and Kozyra, 2003).

The most distinct result of geospace storms in the near-Earth space environment is the intensification of the ring current and of the radiation belts. As perceived by scientists before the space era, and confirmed by spacecraft observations afterwards, energetic charged particles can be trapped by the geomagnetic field and thereafter perform a drift motion around the Earth. The most energetic of these trapped particles comprise the Van Allen radiation belts (Figure 2), which include high-energy ions and relativistic electrons (energies above several hundred keV). Although the acceleration of ions to the moderately high ring current energies (up to a few hundred keV) is firmly associated with storm development, there are still ambiguities about how efficient storms are in terms of radiation belt intensification. For instance, Reeves et al. (2002) showed that storms can either increase or decrease the fluxes of relativistic electrons in the radiation belts.

Although the abundance of these particles is relatively low, their impacts on space technological systems are appreciable, and at times severe (Baker, 2004).

According to the classical picture, radiation belts have a fairly stable basic structure, with a well-known stationary radial profile of the flux represented by standard radiation models (e.g., Walt, 1994). However, inner magnetosphere missions in the 1990s, like SAMPEX and, in particular, CRRES, have demonstrated the highly dynamic and complex nature of the electron radiation belts (e.g., Lemaire, 2001). The shift from the classical radiation belt paradigm is discussed by Vassiliadis et al. in chapter 3 of this book.

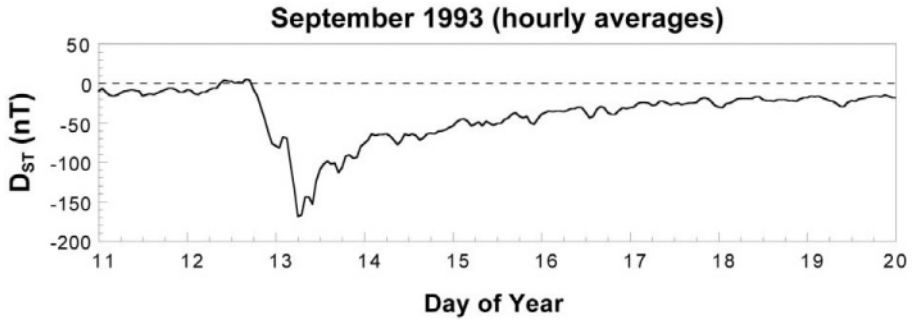

Figure 1. Typical time profile of the *Dst* index - a geomagnetic index commonly used as a measure of storm intensity (courtesy of J. K. Arballo, Jet Propulsion Laboratory). Despite their complexity, geospace storms have been identified by the rather simple pattern that is imposed by their development on the *Dst* time profile.

Ions in the medium-energy range of ~10 keV to a few hundreds of keV constitute the terrestrial ring current (see recent reviews by Daglis et al., 1999; Daglis, 2001b). Long-standing paradigms pertaining to the ring current have included its exclusively solar origin, its build-up through substorms and its decay through charge exchange ("trinity of ring current life", Daglis, 2001a). All of these paradigms have been questioned. The solar origin paradigm, for example, persisted for a couple of decades, from the dawn of the space era to the mid-1980's, when conclusive composition measurements covering the whole energy range important for the storm-time ring current, were performed by the AMPTE mission (Krimigis et al., 1985). The other two ring current paradigms are addressed in more detail in the following sections.

3. IMPORTANCE OF SUBSTORMS IN PARTICLE ACCELERATION AND STORM DEVELOPMENT

One of the oldest and rather classic storm paradigms is the role of substorms in storm development. Sydney Chapman introduced the term "substorm" to suggest that geospace storms and their ring current are the result of a series of intense substorms (Chapman, 1962; Akasofu et al., 1965). Chapman noted that the same storms, which at near-equator magnetic observatories (e.g. in Hawaii) followed simple curves of growth and decay, in Alaska seemed to consist of a number of distinct "sub-storms". We now know that substorms are independent processes and exist at other times as well. Substorms do not need much of a stimulus: during times of southward interplanetary field, the magnetotail seems to quickly reach the rim of instability, and small changes in the solar wind can then trigger a substorm.

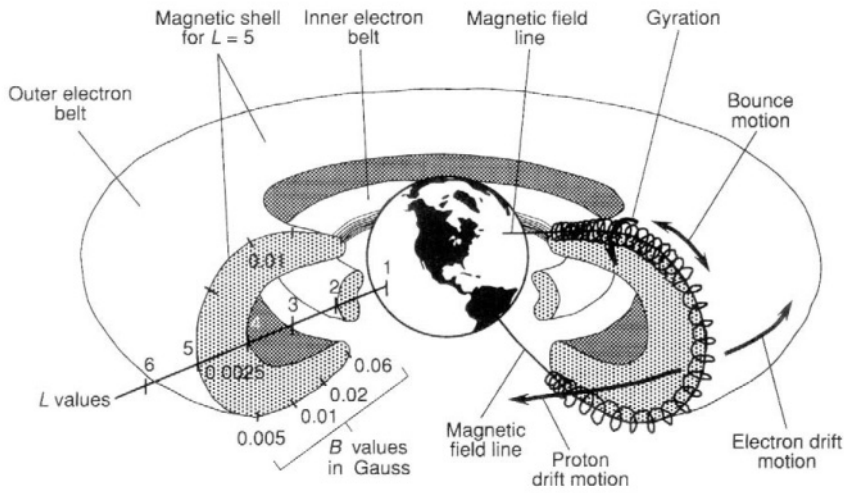

Figure 2. A three-dimensional representation of the inner and outer radiation belts around the Earth (Mitchell, 1994).

Chapman and Akasofu had postulated that the bulk acceleration of particles during storms is the additive result of “partial” acceleration during consecutive substorms (e.g., Akasofu, 1968). This paradigm has been heavily disputed during recent years. A new line of thought is that substorm acceleration may be sufficient to produce individual high-energy particles that create auroras and possibly harm spacecraft, but it cannot produce the massive acceleration that builds-up the storm-time ring current. In other words, it has been suggested that substorm occurrence during storms is

incidental and does not have any causal relation to storm development (Kamide, 1992).

The storm-substorm relation paradigm according to which storms are the result of a superposition of successive "sub-storms" has been addressed by several studies recently (e.g., Kamide et al., 1998). No conclusive evidence has been obtained yet. Studies opposing the "Chapman-Akasofu paradigm" have claimed that substorm occurrence is incidental to the main phase of storms, and that ion transport into the ring current is accomplished solely by enhanced large-scale convection electric fields, with little contribution from substorms if any.

The storm-substorm dispute as appearing in published papers relates to two coupled, yet distinct, issues which are often confused: the effects of substorms on the ring current growth and the effects of substorms on *Dst* variations. The two issues are distinct because the ring current growth is how a storm materializes, while the *Dst* variations is how a storm is measured. We will here discuss only the issue of substorm influence on the bulk particle acceleration that leads to the build-up of the ring current.

Both geospace storms and magnetospheric substorms are characterized by the efficient acceleration of charged particles and their subsequent injection into the inner magnetosphere. However, non-storm substorms are of notably lesser efficiency in the extent of acceleration and inward penetration of charged particles, as compared to geospace storms.

The case against substorms as building blocks of the storm-time ring current is based on the a priori assumption that storm-time substorms do not differ from non-storm substorms, hence the "inability" of non-storm substorms to produce significant ring currents, condemns all substorms to "storm-impotence". However, there are no sound research results that could justify this assumption and therefore it is still too premature to dismiss substorms as particle accelerating processes significant for the storm-time ring current.

The dispute actually refers to the relative efficiency of the large-scale convection electric field and of the substorm-associated impulsive electric fields in accelerating and transporting ions into the ring current. Short-lived impulsive electric fields are induced by magnetic field reconfigurations at substorm onset: i.e., "dipolarizations" from a stretched tail-like configuration to a dipole-like configuration. Wygant et al. (1998) showed that during the large March 1991 storm, the large-scale convection electric field penetrated Earthward, maximizing between L=2 and L=4 with magnitudes of 6 mV/m. Such magnitudes are 60 times larger than quiet-time values. During magnetic field dipolarizations in the inner magnetosphere (i.e., during substorm expansions or intensifications) Wygant et al. also observed strong impulsive electric fields with amplitudes of up to 20 mV/m, which is more

than three times the largest convection electric field. Consequently, substorm induced electric fields can certainly compete with the convection electric field in ion acceleration during storm development. Substorm electric fields may be episodic, but they are much stronger.

In order to reach a conclusion on this issue, it is of essential importance to assess the efficiency of substorm-induced electric fields in ring current development. The problem has been addressed by a number of simulations with contrasting results. While Chen et al. (1994) and Fok et al. (1996) suggested that the substorm contribution is subtle and possibly negative to the development of a ring current, a more recent study by Fok et al. (1999) suggested that the substorm-associated induced electric fields significantly enhance the ring current by redistributing plasma pressure Earthward.

Another approach to this problem relates to the aspect of compositional changes. Massive outflow and preferential acceleration of ionospheric O^+ ions is outstanding during intense storms (Figure 3), when the oxygen to proton energy density ratio can reach values of up to 400% (Daglis, 1997a,b; Daglis et al., 1999b).

As a matter of fact, the O^+ abundance increases with the intensity of storms (Daglis, 1997a). Remarkably, this is also a feature of magnetospheric substorms, certified by relevant studies with measurements from the AMPTE and CRRES missions (Daglis et al., 1994, 1996). Consequently, the magnetosphere-ionosphere coupling may be the final tuning factor of solar wind driving of the two main dynamic magnetospheric phenomena, storms and substorms, in the sense that it regulates their eventual intensity. We shall elaborate a little bit on this suggestion.

Recent modeling of ring current dynamics, based on observational support (Daglis et al., 2003), has put constraints on our empirical recognition that the prolonged southward orientation of the interplanetary magnetic field (IMF) is the main driver of geospace storms. The simulations have shown that this driver is conditioned by internal magnetospheric conditions: the plasma sheet density is of critical importance to the eventual result of the interplanetary drivers, as measured by storm intensity (e.g., Kozyra et al., 1998). Variations in the plasma sheet density significantly modify the geoeffectiveness of southward IMF: high plasma sheet densities result in stronger ring currents. The outflow of ionospheric O^+ ions increases the plasma sheet density. In this sense, the magnetosphere-ionosphere coupling represents a tuning factor for storm development.

Obviously, one must consider and explain the efficient acceleration of relatively cold O^+ ions. The O^+ acceleration is moreover preferential with regard to H^+ and He^{++}, and therefore cannot be accounted for by simple convection. An analysis of single-particle dynamics in simulations of magnetic field dipolarizations (Delcourt, 2002) revealed prominent short-

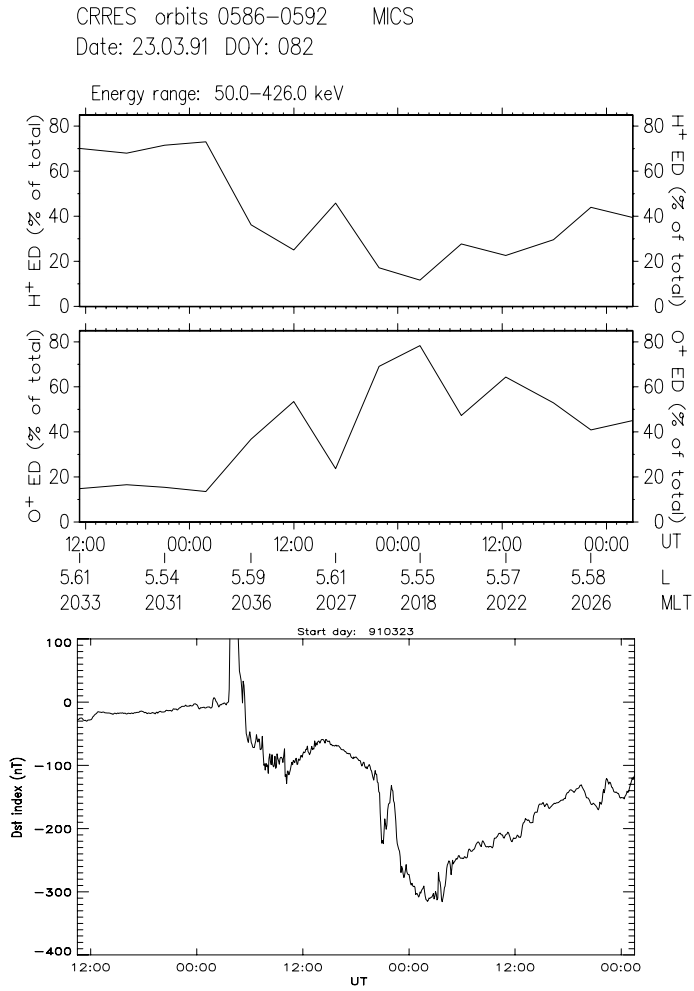

Figure ıner magnetosphere during the main phase of intense storms. During the March 23-25, 1991 storm, O^+ ions contributed nearly 80% of the ring current energy density at storm maximum. This figure shows the time profile of the H^+ (top panel) and O^+ (middle panel) contribution to the total ion energy density in the *L*-range 5-6. The bottom panel shows the time profile of the 5-min resolution *Dst* index. It is remarkable that the *Dst* minima are concurrent with O^+ maxima, implying that the ring current intensifications are due to the acceleration and transport of new ionospheric ions into the inner magnetosphere (adopted from Daglis et al., 1999b).

lived acceleration of plasma sheet ions during the expansion phase of substorms.

Under the effect of the transient impulsive electric fields induced by relaxation of the magnetic field lines, ions with gyro-periods comparable to the field variation time scale can experience dramatic non-adiabatic heating. For example, when considering a 1-min magnetic reconfiguration, low-energy O^+ ions originating from the terrestrial ionosphere are found to be accelerated up to a few hundreds of keV during Earthward injection. These ions evidently can provide a significant or even major part of the ring current. This tells us that, in principle, inductive electric fields, and therefore substorms, are of considerable importance for the storm-time particle acceleration and ring current dynamics.

O^+ is interesting and important not only because of its role in increasing the plasma sheet density or the ring current density itself, but also because of its role in a storm-substorm relationship scenario, which we will outline here. Observations and simulations have indicated a feedback between O^+ injections and substorm breakups moving progressively duskward and Earthward (Baker et al., 1982, 1985; Rothwell et al., 1988). Such a feedback will substantially contribute to a rapid enhancement of the ring current. This feature is also consistent with a relatively old storm study (Konradi et al., 1976), which had shown that the substorm injection boundary was displaced Earthward with each successive substorm during the storm.

Combining model predictions with observations, and considering the fact that O^+ abundance increases with storm size (Daglis, 1997a), we suggest a scenario of a feedback between enhanced (in quantity and spatial extension) O^+-feeding of the plasma sheet and/or the inner magnetosphere and series of intense substorms occurring at progressively lower *L*-shells. Such a combination of successive substorms and continuous O^+ supply can facilitate successive inward penetration of substorm ion injections, consistent with the model of Rothwell et al. (1988) and with the observations reported by Daglis (1997a) and Konradi et al. (1976). The result of successive inward penetration of substorm injections would be the transport of increasingly more energetic ions into the inner magnetosphere, resulting in the intensification of the storm-time ring current. This scenario can explain why some substorms seem to influence the storm time ring current growth, while others don't: substorms resulting in weak inward penetration of injections will not contribute much to the ring current growth. An experimental verification will be possible through detailed global imaging of storms with sufficient composition information.

4. STORM RECOVERY AND RING CURRENT DECAY

According to the traditional *Dst*-profile storm representation, storm recovery manifests itself as the increase of *Dst* from its low, negative values reached during the storm main phase, to its pre-storm level around zero. As already mentioned, Sugiura (1964) designed *Dst* as a ring current measure, according to the original storm - ring current paradigm of Chapman and Singer. Therefore, storm recovery as seen in the *Dst* profile has been traditionally linked to ring current decay. Nevertheless, we now know that *Dst* change does not necessarily reflect changes in ring current intensity, but may signify changes in the intensity of other magnetospheric currents (e.g., Ohtani et al., 2001). This is another important paradigm shift in geospace storm dynamics.

Here we will briefly discuss ring current decay and storm recovery, and we will also refer to some recent results that further modify the classical picture. The loss mechanisms of ring current ions include charge exchange, convective drift losses through the dayside magnetopause, Coulomb collisions with thermal plasma, and wave-particle interactions that cause pitch angle scattering into the atmospheric loss cone. The main mechanism of ring current decay is generally believed to be the charge exchange of ring current ions with cold hydrogen atoms of the geocorona. The geocorona is an exospheric extension of relatively cold (~1000 K) neutral atoms, which resonantly scatter solar Lyman-α radiation, thus optically resembling the solar corona.

All ring current ions are subject to charge-exchange decay, although the decay rate depends on the ion mass and energy. While the O^+-H charge exchange cross section hardly depends on ion energy, the H^+-H charge exchange cross section is dramatically reduced with increasing energy, resulting in much longer charge-exchange lifetimes for higher energy H^+. While at 50 keV H^+ and O^+ lifetimes are comparable, at 100 keV they already differ by an order of magnitude (Smith and Bewtra, 1978). Furthermore, the charge exchange decay rate grows with exospheric hydrogen density, i.e. at lower altitudes. Therefore, ions with mirror points at lower altitudes (i.e., ions with smaller equatorial pitch angles) will charge-exchange easier.

Accordingly, high-energy O^+ will be lost much faster than H^+, and field-aligned pitch-angle distributions will experience larger losses than pancake pitch-angle distributions. It is noteworthy that storm-time O^+ distributions tend to be more field-aligned than H^+ ones (Daglis et al., 1993). Consequently, the storm-time O^+ population will additionally experience faster charge-exchange decay because of their smaller pitch angles.

Therefore, the ring current composition plays a significant role in storm evolution (Daglis, 1997a; Daglis et al., 2003).

Nevertheless, a paradigm shift regarding the extent of charge exchange losses has been introduced during the last few years. Simulation studies have pointed out that convective drift losses through the dayside magnetopause can be the dominant loss process during the storm main phase (e.g., Liemohn et al., 1999; Kozyra et al., 2002). Accordingly, it has been suggested that convective drift loss out the dayside magnetopause has been suggested as the dominant (and fast) process in removing particles from the inner magnetosphere during the main phase and the initial recovery phase of storms. Furthermore it has been suggested that a combination of convective drift loss with a sharp drop in plasma sheet density at the end of the storm main phase, results in a rapid initial *Dst* recovery, as seen during many intense storms that exhibit a two-step recovery (Kamide et al., 1998).

5. ACKNOWLEDGEMENTS

I would like to thank Syun-Ichi Akasofu, Yohsuke Kamide and Dimitris Vassiliadis for their helpful comments.

6. REFERENCES

Adams, W.G., Comparison of simultaneous magnetic disturbance at several observatories, *Phil. Trans. London (A)*, *183*, 131, 1892.

Akasofu, S.-I., *Polar and Magnetospheric Substorms*, D. Reidel, Boston, MA, 1968.

Akasofu, S.-I., and S. Chapman, The ring current, geomagnetic disturbance, and the Van Allen radiation belts, *J. Geophys. Res.*, *66*, 1321-1350, 1961.

Akasofu, S.-I., S. Chapman, and C.-I. Meng, The polar electrojet, *J. Atmos. Terr. Phys.*, *27*, 1275-1305, 1965.

Baker, D. N., E. W. Hones Jr., D. T. Young, and J. Birn, The possible role of ionospheric oxygen in the initiation and development of plasma sheet instabilities, *Geophys. Res. Lett.*, *9*, 1337-1340, 1982.

Baker, D. N., T. A. Fritz, W. Lennartsson, B. Wilken, H. W. Kroehl, and J. Birn, The role of heavy ions in the localization of substorm disturbances on March 22, 1979: CDAW 6, *J. Geophys. Res.*, *90*, 1273-1281, 1985.

Baker, D. N., et al., Recurrent geomagnetic storms and relativistic electron enhancements in the outer magnetosphere: ISTP coordinated measurements, *J. Geophys. Res*., *102*, 14,141-14,148, 1997.

Baker, D. N., Specifying and forecasting space weather threats to human technology, this volume, 2004.

Birkeland, K., The Norwegian Aurora Borealis Expedition, H. Aschehoug and Co., Christiania, sect. 1, 1908.

Broun, J.A., On the horizontal force of the Earth's magnetism, *Proc. Roy. Soc. Edinburgh*, *22*, 511, 1861.

Carrington, R. C., *Observation of the Spots on the Sun from November 9, 1853, to March 24, 1863, Made at Redhill*, William and Norgate, London and Edinburgh, 167, 1863.

Chapman, S., An outline of a theory of magnetic storms, *Proc. R. Soc. London. A95*, 61, 1919.

Chapman, S., Earth storms: Retrospect and prospect, *J. Phys. Soc. Japan*, *17* (Suppl. A-I), 6-16, 1962.

Chapman, S., and J. Bartels, *Geomagnetism*, pp. 321-327, Oxford University Press, London, 1940.

Chapman, S., and V. C. A. Ferraro, A new theory of magnetic storms, *Nature*, *126*, 129-130, 1930.

Chapman, S., and V. C. A. Ferraro, A new theory of magnetic storms, I. The initial phase, *Terrest. Magn. Atmosph. Elec*., *36*, 77-97, 1931.

Chen, M. W., L. Lyons, and M. Schultz, Simulations of phase space distributions of storm time proton ring current, *J. Geophys. Res.*, *99*, 5745-5759, 1994.

Christofilos, N. C., The Argus Experiment, *J. Geophys. Res*., *64*, 869-875, 1959.

Daglis, I. A., The role of magnetosphere-ionosphere coupling in magnetic storm dynamics, in *Magnetic Storms*, *Geophys. Monogr. Ser.*, vol. 98, edited by B. T. Tsurutani, W. D. Gonzalez, Y. Kamide, and J. K. Arballo, pp. 107-116, American Geophysical Union, Washington, DC, 1997a.

Daglis, I. A., Terrestrial agents in the realm of space storms: Missions study oxygen ions, *Eos Trans. AGU*, *78* (24), 245-251, 1997b.

Daglis, I. A., Space storms and space weather hazards, *Proposal for an Advanced Study Institute to the NATO Scientific and Environmental Division*, Athens, February 1999a.

Daglis, I. A., Space Storms, Human Potential Research Training Network Proposal RTN1-1999-00285, Athens, May 1999b.

Daglis, I. A., Space storms, ring current and space-atmosphere coupling, in *Space Storms and Space Weather Hazards*, edited by I. A. Daglis, pp. 1– 42, Kluwer Academic Publishers, Dordrecht, 2001a.

Daglis, I. A., The storm-time ring current, *Space Science Reviews*, *98*, 343-363, 2001b.

Daglis, I. A., Magnetic Storm – still an adequate name?, *Eos Trans. AGU*, *84* (22), 207-208, 2003.

Daglis, I. A., and W. I. Axford, Fast ionospheric response to enhanced activity in geospace: Ion feeding of the inner magnetotail, *J. Geophys. Res.*, *101,* 5047-5065, 1996.

Daglis, I. A., E. T. Sarris, and B. Wilken, AMPTE/CCE observations of the ion population at geosynchronous altitudes, *Ann. Geophys.*, *11*, 685-696, 1993.

Daglis, I. A., S. Livi, E. T. Sarris, and B. Wilken, Energy density of ionospheric and solar wind origin ions in the near-Earth magnetotail during substorms, *J. Geophys. Res.*, *99,* 5691-5703, 1994.

Daglis, I. A., W. I. Axford, S. Livi, B. Wilken, M. Grande, and F. Søraas, Auroral ionospheric ion feeding of the inner plasma sheet during substorms, *J. Geomagn. Geoelectr.*, *48*, 729-739, 1996.

Daglis, I. A., R. M. Thorne, W. Baumjohann, and S. Orsini, The terrestrial ring current: Origin, formation, and decay, *Rev. Geophys*, *37*, 407-438, 1999a.

Daglis, I. A., G. Kasotakis, E. T. Sarris, Y. Kamide, S. Livi, and B. Wilken, Variations of the ion composition during a large magnetic storm and their consequences, *Phys. Chem. Earth*, *24*, 229-232, 1999b.

Daglis, I. A., J. U. Kozyra, Y. Kamide, D. Vassiliadis, A. S. Sharma, M. W. Liemohn, W. D. Gonzalez, B. T. Tsurutani, and G. Lu, Intense space storms: Critical issues and open disputes, *J. Geophys. Res*., *108*(A5), 1208, doi:10.1029/2002JA009722, 2003.

Delcourt, D. C., Particle acceleration by inductive electric fields in the inner magnetosphere, *J. Atmos. Sol. Terr. Phys.*, *64*, 551–559, 2002.

Fok, M.-C., T. E. Moore, and M. E. Greenspan, Ring current development during storm main phase, *J. Geophys. Res.*, *101*, 15,311-15,322, 1996.

Fok, M.-C., T. E. Moore, and D. C. Delcourt, Modelling of inner plasma sheet and ring current during substorms, J. Geophys. Res., *104*, 14,557-14,569, 1999.

Gonzalez, W. D., and B. T. Tsurutani, Criteria of interplanetary parameters causing intense magnetic storms (Dst < −100nT), *Planet. Space Sci*., *35*, 1101-1109, 1987.

Gosling, J. T., The solar flare myth, *J. Geophys. Res*., 98, 18,937-18,949, 1993.

Kamide, Y., Is substorm occurrence a necessary condition for a magnetic storm?, *J. Geomagn. Geoelectr.*, *44*, 109-117, 1992.

Kamide, Y., N. Yokoyama, W. D. Gonzalez, B. T. Tsurutani, I. A. Daglis, A. Brekke, and S. Masuda, Two-step develoment of geomagnetic storms, *J. Geophys. Res.*, *103*, 6917-6921, 1998.

Kintner, P.M., R. Meier, and J. Spann, Living With a Star, the Geospace Mission Definition Team and Aeronomy, AGU Fall Meeting Abstracts, 2001AGUFMSA41C, 2001.

Konradi, A., C. L. Semar, and T. A Fritz, Injection boundary dynamics during a geomagnetic storm, *J. Geophys. Res.*, *81*, 3851-3865, 1976.

Kozyra, J. U., V. K. Jordanova, J. E. Borovsky, M. F. Thomsen, D. J. Knipp, D. S. Evans, D. J. McComas, and T. E. Cayton, Effects of a high-density plasma sheet on ring current development during the November 2-6, 1993, magnetic storm, *J. Geophys. Res.*, *103*, 26,285-26,305, 1998.

Kozyra, J. U., M. W. Liemohn, C. R. Clauer, A. J. Ridley, M. F. Thomsen, J. E. Borovsky, J. L. Roeder, V. K. Jordanova, and W. D. Gonzalez, Multistep *Dst* development and ring current composition changes during the 4 – 6 June 1991 magnetic storm, *J. Geophys. Res*., *107*(A8), 1224, doi:10.1029/2001JA000023, 2002.

Krimigis, S. M., G. Gloeckler, R. W. McEntire, T. A. Potemra, F. L. Scarf, and E. G. Shelley, Magnetic storm of 4 September 1985: A synthesis of ring current spectra and energy densities measured with AMPTE-CCE, *Geophys. Res. Lett., 12*, 329-332, 1985.

Liemohn, M. W., and J. U. Kozyra, Lognormal form of the ring current energy content, *J. Atmos. Sol. Terr. Phys.*, *65* (5), 871-886, 2003.

Liemohn, M. W., J. U. Kozyra, V. K. Jordanova, G. V. Khazanov, M. F. Thomsen, and T. E. Cayton, Analysis of early phase ring current recovery mechanisms during geomagnetic storms, *Geophys. Res. Lett., 26*, 2845-2849, 1999.

Mitchell, D. G., The space environment, in *Fundamentals of space systems*, edited by V. L. Pisacane and R. C. Moore, pp. 45-98, Oxford Univ. Press, Oxford, 1994.

Moos, N. A. F., Colaba magnetic data 1846 to 1905. Part I: magnetic data and instruments. Part II: the phenomenon and its description, Bombay, India, 1910.

Neugebauer, M., and C. Snyder, The mission of Mariner II: Preliminary observations: Solar plasma experiments, *Science*, *138*, 1095-1096, 1962.

Ohtani, S., M. Nosé, G. Rostoker, H. Singer, A. T. Y. Lui, and M. Nakamura, Storm-substorm relationship: Contribution of the tail current to *Dst*, *J. Geophys. Res.*, *106*, 21199-21209, 2001.

Parker, E. N., Newtonian development of the dynamical properties of the ionised gases at low density, *Phys. Rev.*, *107*, 924-933, 1957.

Reeves, G. D., K. L. McAdams, R. H. W. Friedel, and T. P. O'Brien, Acceleration and loss of relativistic electrons during geomagnetic storms, *Geophys. Res. Lett.*, *30*, 1529-1532, 2003.

Rothwell, P. L., L. P. Block, M. B. Silevitch, and C.-G. Fälthammar, A new model for substorm onsets: The pre-breakup and triggering regimes, *Geophys. Res. Lett.*, *15*, 1279-1282, 1988.

Simpson, S., Massive solar storms inflict little damage on Earth, *Space Weather*, *1*, 1012, doi:10.1029/2003SW000042, 2003.

Smith, P. H., and N. K. Bewtra, Charge exchange lifetimes for ring current ions, *Space Sci. Rev.*, *22*, 301-318, 1978.

Sugiura, M., Hourly values of the equatorial *Dst* for IGY, in *Ann. Int. Geophys. Year*, *Vol. 35*, pp. 945-948, Pergamon Press, Oxford, 1964.

Tsurutani, B. T., The interplanetary causes of magnetic storms, substorms and geomagnetic quiet, in *Space storms and space weather hazards*, edited by I. A. Daglis, pp. 103-130, Kluwer Academic Publishers, Dordrecht, 2001.

Van Allen, J. A., The geomagnetically trapped corpuscular radiation, *J. Geophys. Res.*, *64*, 1683-1689, 1959.

Vassiliadis, D., A.J. Klimas, S.F. Fung, D.N. Baker, R.S. Weigel, and S. Kanekal, Structure and dynamics of the outer radiation belt, this volume, 2004.

von Humboldt, A., Die vollständigste aller bisherigen Beobachtungen über den Einfluss des Nordlichts auf die Magnetnadel, *Annalen der Physik*, *29*, 425-429, 1808.

Walt, M., *Introduction to Geomagnetically Trapped Radiation*, Cambridge University Press, Cambridge, UK, 1994.

Wygant, J., D. Rowland, H. J. Singer, M. Temerin, F. Mozer, and M. K. Hudson, Experimental evidence on the role of the large spatial electric field in creating the ring current, *J. Geophys. Res*., *103*, 29,527-29,544, 1998.

Chapter 3

Structure and Dynamics of the Outer Radiation Belt

Implications for Space Weather Modeling and Forecasting

D. Vassiliadis[1], A.J. Klimas[2], S.F. Fung[2], D.N. Baker[3], R.S. Weigel[3], S. Kanekal[4]

1. *USRA at NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA*
2. *NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA*
3. *LASP, University of Colorado, Boulder, CO 80309, USA*
4. *Catholic University of America, Washington, DC 20007*

Abstract Since the early 1990s a series of spacecraft missions have completely transformed our view of the electron radiation belts. This paper summarizes a number of new results on the structure and dynamics of the belts obtained from those measurements. First, the structure of the outer electron belt is discussed with emphasis on the regions (P_0-P_2) distinguished on the basis of the time variations of the electron flux. Each region is characterized by distinct set of acceleration and loss processes. While these processes are traditionally represented by diffusion models, new empirical models have emerged in the last decade. These models are developed from the observed dynamics of the flux as a function of L shell and energy. We have developed such models in each P_i region, and introduce methods of writing them as empirical diffusion-convection models. Since any realistic space weather model must be driven by interplanetary activity parameters, we discuss the development of input-output models (filters) focusing on those driven by the solar wind velocity. In addition to the plasma velocity, other solar wind and IMF parameters are important for each outer-belt region. Taken together, these parameters describe geoeffective solar wind structures. This precursor information can be used to advance the forecast lead time. The expected impact of these modeling approaches to radiation belt forecasting is discussed.

Keywords Magnetic storms, radiation belts, particle acceleration, MHD waves and instabilities, solar wind, magnetosphere-inner, solar wind/magnetosphere interactions, solar cycle, forecasting, space weather, time series analysis.

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 43-64.

1. RADIATION BELT MODELING: FROM SPACE PHYSICS TO SPACE WEATHER

Inner-magnetosphere spacecraft missions of the last decade have highlighted the dynamic and complex nature of the electron radiation belts [Lemaire, 2001]. They have led to new questions regarding the relative efficiency of acceleration and loss processes, and the conditions under which these processes are activated [Li and Temerin, 2001; Friedel et al., 2002]. At the same time the need for space weather forecast products has steadily increased [Baker et al., 2001]. Modern empirical models have been developed to address that need, using the significant knowledge base of mission datasets. The first attempts in real-time prediction at geosynchronous orbit are promising [Baker et al., 1990; Li et al., 2001], while future capabilities and accuracy levels are expected to increase with the advent of data-assimilation techniques [Moorer et al., 1999; Rigler et al., 2004].

Probably more than any other mission, the Combined Release and Radiation Effects Satellite (CRRES) spacecraft demonstrated the dynamic character of the radiation belts. One of the best-studied events has been the March 24, 1991 storm highlighted by the rapid formation of a new radiation belt during storm sudden commencement during the passage of a high-amplitude interplanetary shock [Li et al., 1993]. After its formation deep in the inner magnetosphere at L=2.5, normally a region of relatively low flux due to enhanced precipitation and loss, the belt persisted at least until the mission's end, 6 months from the time of its creation. Following CRRES, the Polar spacecraft [Blake et al., 1995] and Cluster constellation [Daglis et al., 1995] have provided new insights in the composition and energy spectrum of the trapped population.

While unique events, such as the March 1991 storm, reveal novel aspects of the acceleration processes, comprehensive modeling of the radiation belts needs to be based on statistical and comparative studies, and this is the perspective emphasized in the present paper. Statistical and dynamical modeling makes use of the long-term monitoring of the radiation belts by the Solar, Anomalous, and Magnetospheric Particle Explorer (SAMPEX), EXOS-D (Akebono) and its predecessors, the GPS constellation, and the numerous spacecraft at geosynchronous orbit (GEO).

Several of the results reviewed below are based on analysis of data by SAMPEX. This spacecraft is a Small Explorer Mission launched in 1992 into a polar, circular, low Earth orbit (LEO) [Baker et al., 1993]. One might presume that the usefulness of flux measurements by LEO spacecraft like SAMPEX and Akebono may be limited, because they have access only to the off-equatorially-mirroring part of the trapped distribution rather than the full trapped distribution measured by spacecraft at an equatorial orbit, such

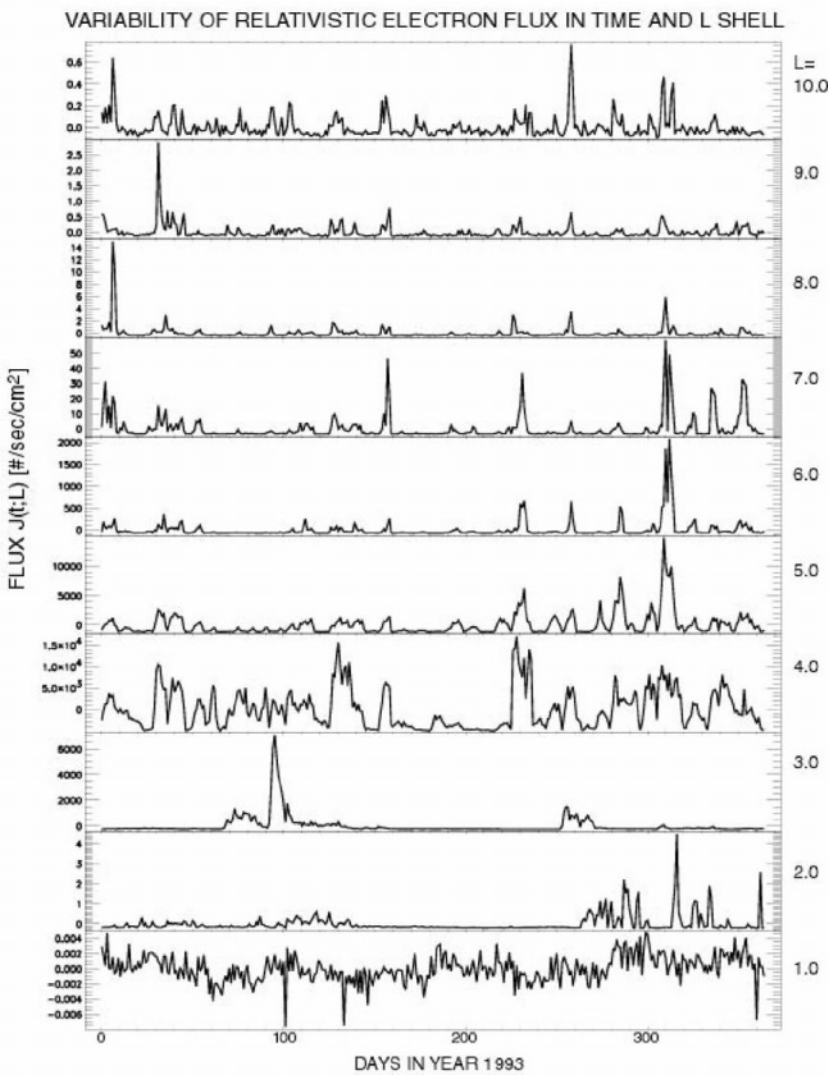

Figure 1. Time and L shell var iability of the omnidirectional relativistic-electron flux j_e(t;L=const.). SAMPEX/PET flux data are shown over the course of one year (1993) and at equidistant L shells.

as CRRES or various GEO spacecraft. Comparison between simultaneous flux measurements, however, has shown that the pitch angle equilibration is rapid, of the order of a few hours, and leads to a coherent response of the inner magnetosphere [Kanekal et al., 2001]. In addition, LEO spacecraft cross a wide range of L shells every few hours, in a traversal, which is fast enough to produce synoptic coverage of the entire radial range of the inner magnetosphere at the timescale of one day. Due to its fast period and solar-

cycle-long mission duration, SAMPEX has captured a plethora of radiation-belt phenomena triggered by the passage of various interplanetary structures.

In general both low and high-altitude measurements have demonstrated the extent to which the energetic electron flux varies in a complex manner in both time and space. For instance, Fig. 1 shows the electron flux from SAMPEX's Proton-Electron Telescope (PET) [Cook et al., 1993] at energies of 2-6 MeV. The flux variations are shown at fixed, equidistant L shells ($\Delta L=1.0$) over the course of one year. Note that high-amplitude relativistic electron flux events (electron "storms") may sometimes appear simultaneously over many radial distances, but in general their amplitude and duration differs significantly with altitude. Some storms are observed only in a subset of latitudes and, to complicate things further, measurements at high L shells can sometimes miss signatures of storms entirely while in other regions the instrument may be contaminated or its detectors may be saturated. The spatial and temporal complexity and the small number of simultaneous observational platforms make it difficult to conclusively identify source and loss regions and mechanisms. These difficulties are compounded by the current lack of an accurate magnetic-field model [Selesnick and Blake, 2000].

In the following sections this paper reviews the structure and dynamics obtained from various types of flux data analysis and the prospects of physical and empirical modeling. The following topics are discussed:

- Structure of the outer zone. Section 2 examines the electron flux $J_e(L;E)$ and its correlation as a function of radial distance.
- Flux dynamics resulting from acceleration, loss, and transport. Section 3 is an introduction to the basic processes, and presents the development of empirical models of the flux variations.
- Coupling to interplanetary activity. Geoeffective structures in the solar wind couple to each radial region of the inner magnetosphere in different ways. In Section 4, filter and precursor analysis are used to identify relevant interplanetary inputs and the mechanisms that they drive. At longer timescales, the methods are used to measure the geoeffectiveness of solar wind inputs as a function of solar cycle and season.

2. RADIAL STRUCTURE OF THE RADIATION BELTS

The electron radiation belts constitute a complex plasma system. Its most salient radial feature is the division into the inner and outer belt by the slot region in which electron scattering and loss are maximized. The stationary (time-averaged) radial profile of the flux is well known [Walt et al., 1994] and represented by standard radiation models [e.g., Vette, 1991; Heynderickx, 2002]. A better resolution of the radial structure can be obtained

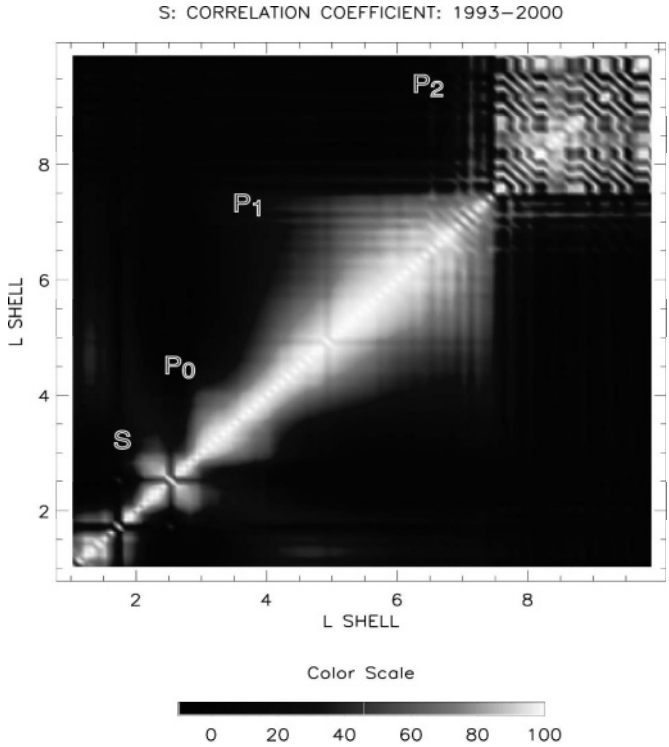

Figure 2. The radial correlation function (1) is calculated from SAMPEX/PET fluxes in the observation interval 1993-2000. Note the division of the outer belt in 3 "blocks," or structures P_0-P_2 of different dynamics. The quadrupole-shaped region below P_0 is the slot (S); the quadrupole shape means that the flux dynamics at the slot edges are strongly correlated with each other suggesting loss processes that operate coherently across the slot.

by examining the flux variation of the radial profile [Vassiliadis et al., 2003b], as will be reviewed below. On the other hand, local time and polar-angle variations of the flux are smaller, because of the high drift velocity and bounce frequency, respectively. Nevertheless, acceleration and loss processes are highly structured in MLT. For instance, the growth of ULF waves, involved in radial diffusion and/or acceleration, is prominent in the dayside and the dawn regions [Anderson et al., 1990; Engebretson et al., 1998].

We discuss the radial structure of the belts in terms of the omnidirectional flux $J_e(t;L;E)$ at shell L and energy E. The time variation of the log-flux at fixed L and E produces the time series $j_e(t;$ L=const.; E=const.). Using low-time-resolution (daily-average) fluxes reduces the significance of individual injections and adiabatic effects. Under those conditions, the time variation of the flux is indicative of the global dynamics of the belts [see also Baker et al.,1999; Kanekal et al., 2001].

To measure the spatial variation of $J_e(t;L)$ we use the radial correlation function:

$$C(L_1,L_2)=\frac{1}{T\sigma_{J_e(L_1)}\sigma_{J_e(L_2)}}\int_{t=0}^{T}\delta J_e(t;L_1)\,\delta J_e(t;L_2)\,dt \quad (1)$$

where δJ_e is the fluctuation of the flux $J_e(t;L=\text{const})$ from its long-term average, $\sigma_{J_e(L_i)}$ is the standard deviation of the flux distribution in shell L_i; and time T is a long interval of activity (here: years 1993-2000). The L shell range is [1,10] while the energy is fixed at the 2-6 MeV channel of PET. This standard form of correlation takes values in the range [-1,1].

The radial correlation function (1) consists of discrete regions, seen as diagonal blocks in Fig. 2, of high correlation values. Outside these blocks the correlation is low, meaning that the time variations of $J_e(t;L=\text{const})$ varies significantly from one region to another. This view of the outer belt's radial structure stands in direct contrast to the time-averaged profile $J_e(L)$ which varies much more slowly with L than C(L,L).

Region P_1 at L=4.1(±0.2)-7.5(±0.4), is the "heart" of the outer belt, containing by far the largest amount of trapped radiation flux than the other two regions due to its size and flux amplitude. Because L=6.6 falls within that range, the flux variation at the geosynchronous orbit is on average similar to the variation in other L shells, all the way down to L=4. However, solar cycle variations affect the flux at the geosynchronous region much more than the flux at the lower part of region P_1 (see discussion below in this section). Also the geosynchronous orbit is differentiated by other factors from the rest of P_1 such as the response to IMF inputs [Vassiliadis et al., 2004].

Closer to Earth than L=4 is region P_0, at L=3.0(±0.1)-4.0(±0.2) with clearly different dynamics from P_1. The difference between the two regions arises because of different acceleration and loss processes, as well as different types, and degrees, of coupling to the solar wind. The differences are further discussed in Sections 3 and 4, respectively.

At higher L shells than P_1, L>7.5(±0.4), is region P_2 featuring lower-amplitude, transient fluxes. The dynamics of this marginally trapped population is most probably dependent on the plasma sheet penetration into the inner magnetosphere. An alternative view suggests that a significant part of these electrons are accelerated in the cusp [Fritz, 2001; Sheldon et al., 1998]. A third possibility is that the flux dynamics in P_2 is related to the "leakage", or outward transport, of electrons from the main portion of the outer belt [Blake et al., 2002].

Below region P_0, the slot (S) at L=2.0(±0.1)-3.0(±0.1) has a characteristic quadrupolar shape in the correlation function (Fig. 2). This type of shape means that the slot edges are highly correlated and particle loss occurs simultaneously on average. The interpretation is that the causes of

loss processes, such as VLF wave fields, vary coherently over a radial range comparable to the slot width.

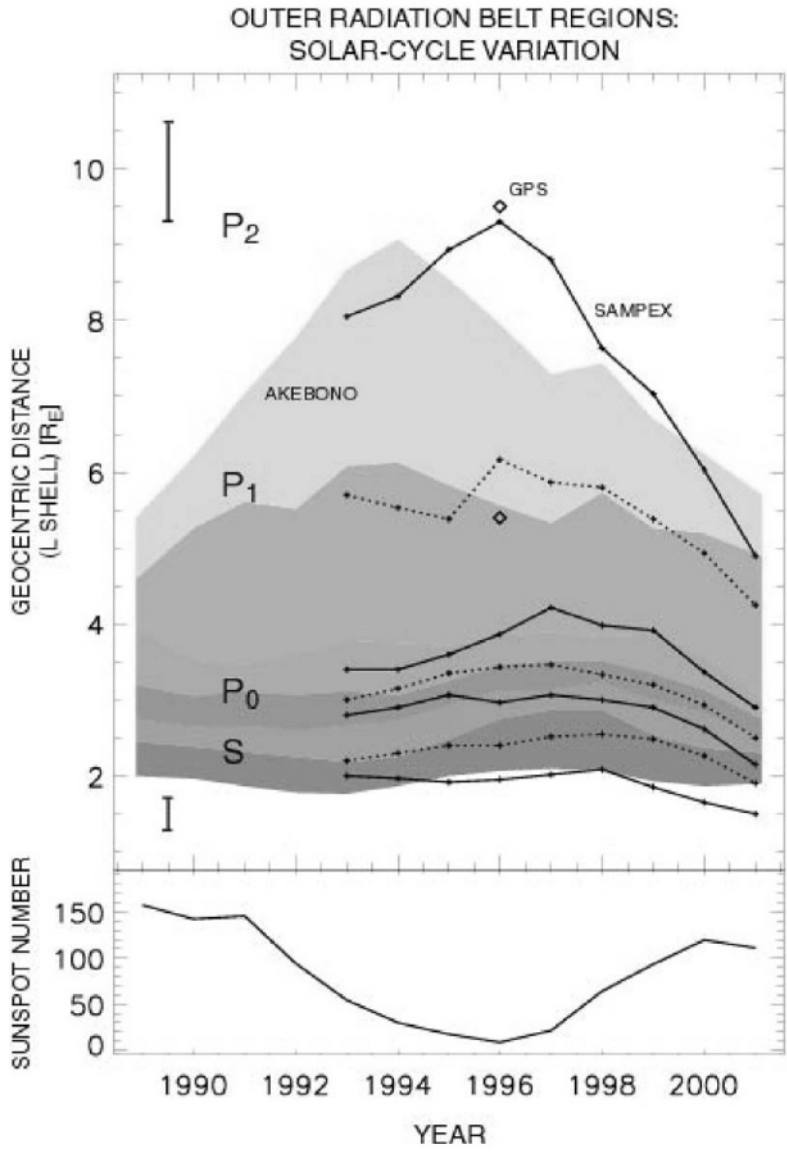

Figure 3. Upper panel: Solar cycle variation of the size of the three regions P_i and the slot. The region boundaries are obtained by Akebono/RDM measurements (regions shown in color; including white for P_2 and black for the slot) and SAMPEX/PET (shown as dotted lines). In 1996, the GPS NS-33 spacecraft provided two measurements for L>4 (shown as diamonds). Within a region, the L shell with the widest correlation length is also indicated (for Akebono: changes in color shading; and for SAMPEX: solid lines). Lower panel: Sunspot number. Note the erosion of P_1 during solar maximum in favor of the quasitrapping region P_2.

Similar results are obtained from three other spacecraft: EXOS-D/Akebono (instrument: RDM), EXOS-C/Ohzora (HEP), and GPS NS-33 (BD II) [Vassiliadis et al., 2003b]. These spacecraft have operated at different altitudes and solar cycle phases, and their instruments have covered different energy ranges.

The time variation of the P_i regions can be followed as a function of solar cycle phase thanks to the long-term observations of Akebono and SAMPEX (Fig. 3). The boundaries and "centers" (points of widest correlation in L) of the three regions are variable with time. The largest radial displacement is that of the P_1-P_2 boundary: during solar maximum, buffeting of the magnetosphere by shocks, CMEs, and other solar wind ejecta reduces the long-trapping region P_1 in favor of the quasi-trapping region P_2. The variability is evidence that the geosynchronous orbit is dominated by different populations and dynamics in the course of the solar cycle. Therefore a realistic model of the geosynchronous region must vary with solar cycle phase.

A similar oscillation can be seen at the boundary between region P_0 (whose greatest radial extent occurs approximately 3 years before solar minimum) and the slot (3 years before solar maximum). In addition to the region boundaries, one can discern a solar-cycle variation in the L shell of the broadest correlation (indicated on Fig. 3).

3. FLUX DYNAMICS DUE TO ACCELERATION AND LOSS PROCESSES

Electron acceleration occurs due to a variety of processes [Friedel et al., 2002]. Important among those are the interaction with low-frequency waves which scatter the electrons in energy and/or pitch angle. Other major processes are direct injection and nonlinear diffusion.

A standard scenario involving ultra-low-frequency (ULF) waves begins with reconnection producing a seed population of electrons (10-100 keV) during storms and substorms. Increases in solar wind velocity V_{SW} excite ultra-low-frequency (ULF) waves in the dayside magnetopause and the flanks [Engebretson et al., 1998; Vennerstrom, 1999] primarily as shear-flow instabilities [Farrugia et al., 2001] and compression. These are consistent with the ULF wave distribution as determined from in situ measurements [Anderson et al., 1990] and remote sensing [e.g., Pilipenko and Engebretson, 2002] (note that there are significant differences between space and ground observations because of ionospheric absorption of the waves). At the end of this growth stage the wave power reaches its peak after 1 day [Rostoker et al., 1998; Mathie and Mann, 2000; O'Brien et al., 2003].

In the second stage the waves accelerate the seed electrons to MeV energies possibly through resonant acceleration. In this type of acceleration there is a resonance

$$\omega - m\omega_d = 0 \tag{2}$$

between the wave frequency ω and the electron drift frequency ω_d. The observational evidence is complemented by numerical experiments in which fields from global MHD simulations are used to drive guiding-center particle codes [Hudson et al., 1999; Elkington et al., 1999, 2003].

Fluxes in the few-MeV range reach their peak after 2-3 days at the geosynchronous region [Paulikas and Blake, 1979; Baker et al., 1990] and more generally in the P_1 region [Vassiliadis et al., 2002]. For a monochromatic wave, the maximum energy gain ΔE_m is the half-width around the resonant frequency ω [Elkington et al., 2003]:

$$\Delta E_m = \sqrt{\frac{2e\delta E_{rm}\delta r}{m\left[\frac{\partial(\ln\omega_d)}{\partial E}\right]_{E=E_m}}}$$

Other factors determining the efficiency of resonant acceleration are the type of resonance (toroidal or poloidal) [Chan et al., 1989; Elkington et al., 1999], the effect of solar wind pressure, etc. Further improvements are expected in the near future since the current global MHD models do not accurately represent the inner magnetosphere structure (e.g., they do not include a ring current), or time dependence (their effective time resolution is typically ~1 min).

In addition to resonant acceleration, other candidate acceleration mechanisms are large- and small-scale recirculation [Nishida et al., 1976], cusp acceleration [Sheldon et al., 1998; Fritz, 2001]; direct injection during substorms [Ingraham et al., 2001]; and enhanced diffusion [Hilmer et al., 2000]. The March 1991 event mentioned in Section 1 highlighted what is probably one of the fastest and most efficient acceleration processes, involving a magnetospheric compression by a high-amplitude interplanetary shock, and the resulting impulsive injection [Li et al., 1993]. ULF-wave-related mechanisms include magnetic pumping via pitch-angle scattering and flux tube motion associated with the waves [Liu et al., 1999]; and cyclotron interaction between trapped electrons and a fast-mode ULF wave [Summers and Ma, 2000]. Finally, loss mechanisms are numerous as well with electron scattering off waves (VLF chorus, whistlers) being two of the most important ones.

While in all these processes, one or more adiabatic invariants are violated resulting in higher fluxes at relativistic energies, adiabatic processes simply displace the particles while preserving the invariants. In the latter case, phase space density is preserved and the electron displacement is temporary. Thus in order to develop a realistic model based on the observed

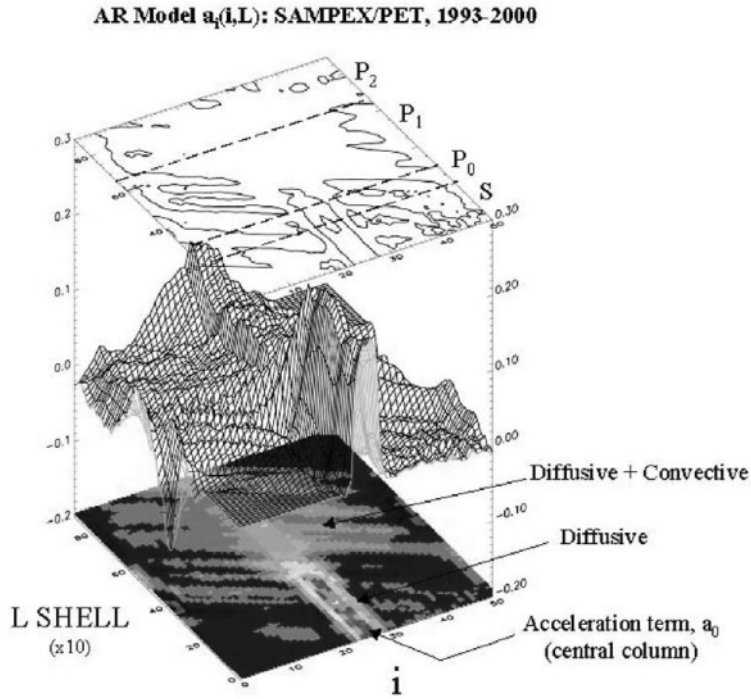

Figure 4a. Radial transport coefficients a_i from AR model (3) as a function of i and L. The left-hand (right-hand) part of the figure indicates transport from lower (higher) L shells, or i<0 (i>0).

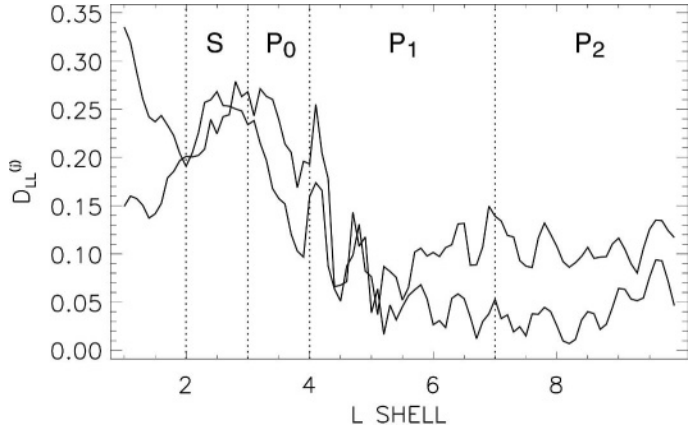

Figure 4b. Diffusion coefficient $D_{LL}^{(j_e)}$ for j_e. Two different ways of calculating it from (5) are shown.

flux variations, the effects of the adiabatic changes need to be removed first. The most important adiabatic effects are the displacement of electrons by an increase in the ring current (the "D_{st} effect") or the tail current. The D_{st} variation accounts only for a small fraction of the storm-time electron flux change [McAdams and Reeves, 2001].

In the following we develop an dynamical empirical model for the time variations of the logarithmic flux (or log-flux) $j_e(t;L;E) = \log_{10} J_e(t;L;E)$ parameterized by L. Adiabatic effects are minimized by the choice of a low time resolution (1 day). We write the log-flux on day t+1 as a function of the log-flux at nearby L shells on day t:

$$j_e(t+1;L) = \sum_{i=-N}^{N} a_i j_e(t;L+i\delta L) \tag{3}$$

where N is a free parameter and δL=0.1 is the resolution in L shell. The 0-th term represents the effect of local acceleration on the flux at fixed L over the course of one day. The terms with $i \neq 0$ indicate transport from lower and higher L shells. The radial range over which $a_i(i)$ is significant determines N, and therefore $N\delta L$ is a effectively a length scale of spatial correlation (compare Eq. (3) with (1)). This type of model is called autoregressive (AR).

In its current version, the model (3) ignores any effects due to interplanetary input or magnetospheric-activity parameters. These effects are examined in Section 4. In addition, the electron energy is fixed at 2-6 MeV for all L shells. In the next version of the model, energy changes as a function of L will be included.

Solving Eq. (3) for the radial coupling coefficients a_i we find that they are functions of i and L (Fig. 4a). A coefficient at position (i,L) indicates the amplitude of the coupling between $j_e(t;L)$ and $j_e(t;L+i\delta L)$ as follows:

In regions S and P_0 the correlation length is small, N=3. The coefficients $a_i(i)$ are large only close to i=0 and symmetric. Their profile indicates a diffusive process. The effective diffusion coefficient will be discussed below.

On the other hand, regions P_1 and P_2 are characterized by a much larger correlation length ($N\delta L$ = 2.0-2.2). The coefficients $a_i(i)$ are asymmetric with the parameter i, and are broadly distributed around the central column, i=0. Both these features cannot be explained by diffusion alone and need to include convection as well.

To see the connection between model (3) and diffusive-reactive radial transport consider such a process for j_e:

$$\frac{\partial j_e(t;L)}{\partial t} = \frac{\partial}{\partial L}\left(D_{LL}^{(j_e)} \frac{1}{L^2} \frac{\partial\left(L^2 j_e\right)}{\partial L} \right) + V_{conv} \frac{\partial j_e}{\partial L} + S(t;L) \tag{4}$$

where the right-hand side contains a diffusive term with coefficient $D_{LL}^{(j_e)}$. We also include a convective term with speed V_{conv}, and a source term S. A diffusion equation is obtained for the phase space density $f = j_e / p^2$, where p is the relativistic momentum, as an approximation to the Fokker-Planck equation. In that case, however, convective effects are typically neglected [Walt, 1994].

We consider that an injection has just taken place and will not be followed by other injections for some time. In modeling the time decay of the flux we can thus neglect the source term S(t;L). Discretizing (4) in t and L with steps $\delta t=1$ and δL, respectively, assuming that D_{LL} varies slowly with L, we obtain Eq. (3) with

$$a_0 = 1 + 2\left[\frac{1}{L}\frac{\partial D_{LL}}{\partial L} - D_{LL}\left(\frac{1}{\delta L^2} + \frac{1}{L^2}\right)\right] \simeq 1 - 2D_{LL}\left(\frac{1}{\delta L^2} + \frac{1}{L^2}\right)$$

$$a_i = \pm\frac{1}{2\delta L}\frac{\partial D_{LL}}{\partial L} + D_{LL}\left(\frac{1}{\delta L^2} \pm \frac{1}{L\delta L}\right) \pm V_{conv}\frac{1}{2\delta L} \simeq D_{LL}\left(\frac{1}{\delta L^2} \pm \frac{1}{L\delta L}\right) \pm V_{conv}\frac{1}{2\delta L} \quad i \gtrless 0 \tag{5}$$

A comparison between (5) and (3) yields an effective diffusion coefficient $D_{LL}^{(j_e)}$ for the log-flux (Fig. 4b). Its scaling with L differs significantly from the conventional D_{LL} for the phase space density. Similar scalings are obtained also for the effective convection velocity, V_{conv}.

4. DYNAMIC INPUTS: INTERPLANETARY AND MAGNETOSPHERIC PARAMETERS

Acceleration and loss processes are driven or modulated by changes in the solar wind and the interplanetary magnetic field (IMF). It is therefore important to determine the empirical relations between interplanetary parameters and flux variations. Such empirical relations serve as baseline predictive models driven with time-dependent solar wind inputs [Baker et al., 1990] and can be incorporated in space weather models [Heynderickx, 2002]. In addition to driver inputs, the state of an empirical model can be specified more precisely by additionally using magnetospheric parameters (such as magnetic indices) which provide a specification of the activity level [Fung, 1996].

4.1. Filters

Several interplanetary parameters are important for radiation-belt dynamics. Here we examine the solar wind plasma velocity, V_{SW}. Others are examined elsewhere [Blake et al., 1997; Fung and Tan, 1998; Tsutai et al., 1999; Vassiliadis et al., 2004].

The solar wind velocity is the single most important interplanetary input to the radiation belts. The relative importance of this parameter can be assessed by measuring the higher prediction capability of electron flux dynamics using the V_{SW} input compared to any others. The significance of V_{SW} was determined early on through comparisons between V_{SW} and subsequent relativistic electron fluxes at geosynchronous orbit [e.g., Paulikas and Blake, 1979]. Physically, increases in V_{SW} lead to momentum and energy transfer on the dayside through compressions and at the magnetospheric flanks through shear-flow instabilities such as Kelvin-Helmholtz [e.g., Farrugia et al., 2001]. The effects of the viscous processes are most clearly observed under conditions of a zero or weakly positive IMF B_z. In either scenario, the compression or instability drives ULF waves such as Pc5 (period of 2-10 min) into the magnetosphere [Anderson et al., 1990; Engebretson et al., 1998; Vennerstrom, 1999]. The waves grow significantly ~1 day before the rapid increase of energetic electrons [Rostoker et al., 1998].

The effective coupling between velocity and fluxes is a complex function of L shell. At geosynchronous-orbit altitudes an increase in $V_{SW}(t)$ produces an increase in $j_e(t;L=6.6)$ 2-3 days later [Baker et al., 1990]. Using SAMPEX/PET data we extend the modeling for all L shells in the range L=[1-10] [Vassiliadis et al., 2002]. As a model we use a linear finite-impulse-response (FIR) filter of the form:

$$j_e(t;L_i) = \int_{-T_S}^{T} H(\tau;L_i) V_{SW}(t-\tau) d\tau \qquad (6)$$

where the log-flux $j_e(t;L) = \log J(t;L)$ as before, the impulse response $H(\tau;L_i)$, parametrized by L_i, is convolved with V_{SW}; the coupling starts at $-T_s$ and ends at time T, both measured in reference to the time of arrival of solar wind at the magnetopause.

The impulse response at geosynchronous-orbit altitude, $H(\tau;L=6.6)$, peaks at τ=3 days (Fig. 5a) in agreement with earlier studies which analyzed geostationary satellite data [e.g., Baker et al., 1990]. The response is calculated from 8 years (1993-2000) of daily averages of V_{SW} and SAMPEX flux measurements so it represents a long-term average [Vassiliadis et al., 2002]. The function becomes negative at τ=7 days meaning that from that time on, variations in V_{SW} are anticorrelated with increases of $j_e(t;6.6)$, or

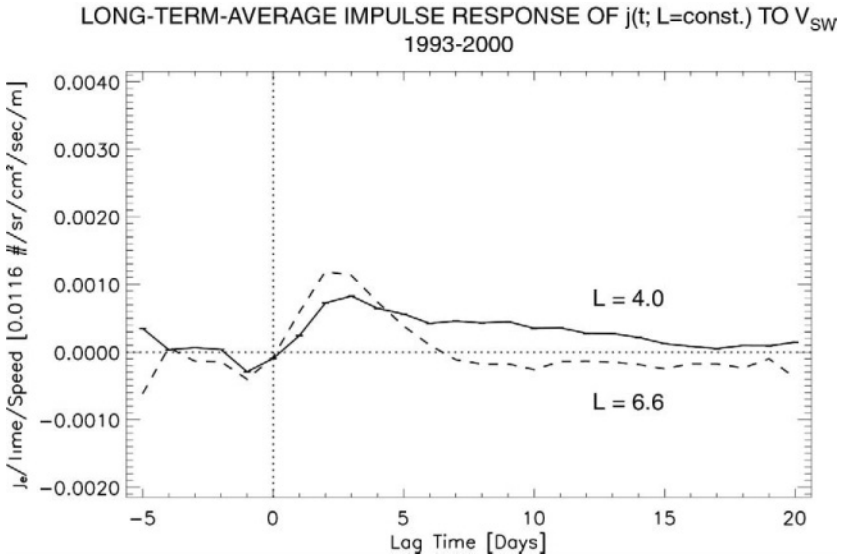

Figure 5a. Impulse response function for V_{SW} from the FIR model (6) for individual L shells, L=4.0 and 6.6. Note the similarity in the location of the τ=3 day peak, subsequent decay, as well as the minimum at τ=-1 day.

particle loss. An earlier minimum at τ=-1 days is due to the adiabatic displacement of electrons by the growth of the ring current. The current intensifies because of enhanced convection caused by a Southward interplanetary magnetic field component. The intensification occurs initially at approximately L=5.5 and then expands radially Earthward and outward.

The second curve in Fig. 5a shows the flux response at L=4.0. The peak position is located close to that of L=6.6; note, however, that H(τ;4.0) remains positive long after τ=7 indicating stable trapping at this L shell.

Making use of SAMPEX's continuous and broad coverage in L shell, we synthesize a composite impulse response H(τ;L) from individual response functions calculated at fixed L. The composite function expresses the amplitude and time of the coupling V_{SW}, as well as the radial location where it occurs (Fig. 5b). Peaks P_0 and P_1 correspond to the blocks in the radial correlation graph of Fig. 2. Thus there is a good correspondence between the dependence of long-term flux dynamics on L shell and the flux response to V_{SW} (and other inputs). The correspondence suggests that the dynamics are determined to a great extent by the external forcing rather than by internal processes.

The filters (6) can be interpreted as direct local acceleration due to processes activated by the solar wind speed (such as ULF wave acceleration). Alternatively, increases in j_e(t;L=const.) can be due to transport from higher L shells. The ambiguity between the two inter-

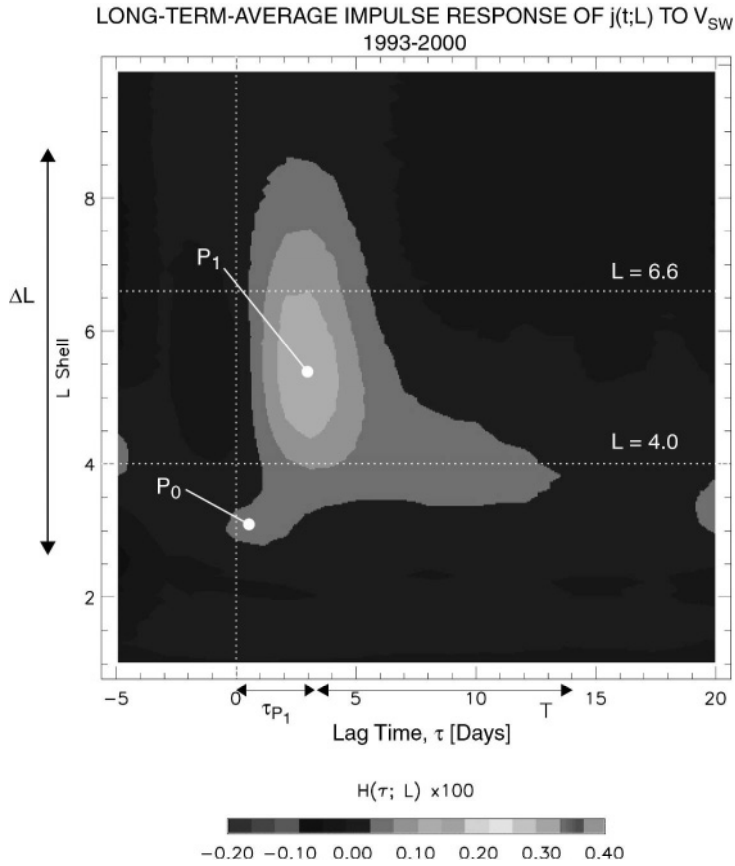

Figure 5b. Impulse responses H(τ;L) for all shells in the L range [1,10]. Note the position and extent of region P_1 (response to high-speed streams) and P_0 (response to CMEs).

pretations arises because the modeling is applied on fluxes at a single energy range, and can be resolved by modeling of flux measurements at multiple energies.

In addition to solar wind velocity, other interplanetary variables are important in controlling the flux dynamics. Earlier studies have shown the significance of the IMF B_z component and density [Blake et al., 1997] as well as magnetospheric indices [e.g., Tsutai et al., 1999]. Vassiliadis et al. [2004] have examined the response of the electron flux in terms of solar or interplanetary variables, or geomagnetic indices which we use as proxies for the regional electrodynamic activity.

We find that a total of 17 such parameters fall into three categories which affect each P_i region in a different way. Hydrodynamic parameters such as V_{SW}, ρ_{SW}, and P_{SW} can predict up to 36% of the variance in P_1 and a smaller amount in P_1. The IMF B_z component, and magnetic indices that

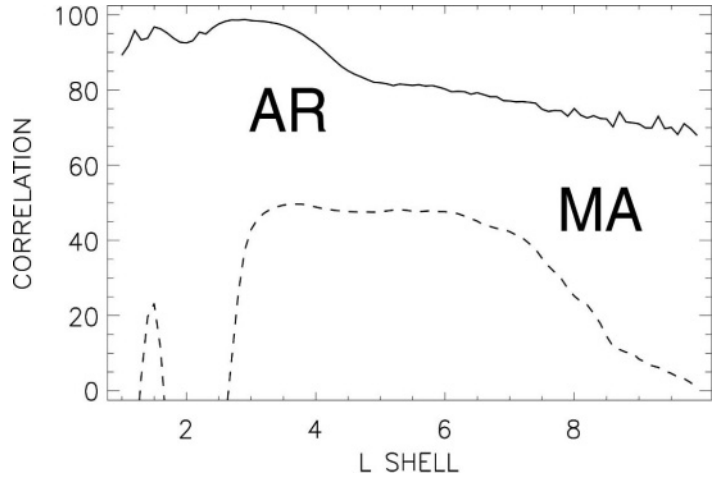

Figure 6. Comparison of the data-model correlations for the AR and FIR (denoted as MA) models. The square of the correlation is the percentage of the variance in the flux data explained by the model.

depend on it, form a second category. The IMF B_z regulates, primarily through the reconnection rate at the magnetopause, the intensity of currents systems such as the ring and tail currents, field-aligned currents, and ionospheric currents. The electrodynamic activity is quantified in terms of regional or planetary measures of geomagnetic activity, which are the geomagnetic indices. Both the IMF B_z and the indices predict the variance of fluxes in P_0 and P_1 in a very similar manner. The indices such as K_p and the polar cap index, PC, are much more accurate, however, predicting the variance of P_1 at a 25% level and the variance of P_0 at a 50% level. Much higher percentages can be explained by AR models such as (3) as Fig. 6 shows. The combination of models (3) and (6) are expected to further increase the explained variance of $j_e(t;L)$.

4.2. Precursors

For an externally driven system like the radiation belts, determining the geoeffective precursor activity is a more direct method than filter analysis for forecasting and modeling. The precursors to an electron acceleration event can be physically interpreted as structures in the solar wind. We briefly sketch out below the precursor analysis which is described at length in [Vassiliadis et al., 2003a]. We denote the daily average $J(t_i;L_j)$ as an event of that amplitude at shell L_j. The precursor to that event, in terms of the solar wind velocity, is the activity vector

$$\mathbf{I}^{(V_{SW})}(t) = [V_{SW}(t-(T-1)),\ldots,V_{SW}(t-(T-1)),V_{SW}(t-T),\ldots,V_{SW}(t+T_s)] \quad (7)$$

or a window in the velocity time series with width $T+T_s$, where generally $T \gg T_s \simeq 0$. Similar precursors can be formed for other solar wind and IMF variables.

Precursors to events of similar activities are averaged together in a superposed-epoch-type analysis. The key is to identify intervals of similar activity over a large and comprehensive dataset. As such a database, we use the SAMPEX/PET daily flux measurements from 1993 to 2000.

The measurements are sorted in order of increasing amplitude and divided in 4 quartiles (indexed by q=1,2,3,4), each comprising ~730 days. The average flux in the q-th quartile is $\langle j_e(L)\rangle_{(q)}$. The average precursor, $\langle \mathbf{I}^{(V_{SW})}(t;L)\rangle_{(q)}$, a vector of length $T+T_s$, is obtained by averaging over the corresponding individual precursors (7) for each $j_e(t;L)$ in the q-th quartile. Note that geoeffectiveness is defined as the flux over a given L shell range, and therefore the precursor depends on the choice L shell range.

The most geoeffective precursor is that with the highest q-value (here: q=4) corresponding to the top quartile. It is by construction the average of the solar wind conditions that result in the highest flux at a given L shell, or range. Fig. 7 shows the velocity, IMF B_z (in the GSM coordinate system), and solar wind density V_{SW} for q=4. Precursors are shown for 3 L shell regions: P_1, P_2, and the entire inner magnetosphere (L=1-10). The precursor for P_0 is not different from P_1 at the daily resolution so it is not shown.

First, there is a strong similarity between the precursors of the flux in region P_1 and those of the flux in the entire inner magnetosphere. This is because the total electron flux in P_1 is much higher than for any other region in the outer belt; thus, defining geoeffectiveness for P_1 is generally very similar to defining geoeffectiveness for the outer belt, and even in the entire inner magnetosphere.

Second, the precursors for P_1 and P_2 have opposite orientations in solar wind velocity V_{SW} and density ρ_{SW}. In addition the P_1 precursor has a Southward IMF B_z (middle panel of Fig. 7), consistent with energization through dayside reconnection and, eventually, production of seed electrons. The P_2 precursor is characterized by a Northward IMF B_z., producing a weak reconnection poleward of the cusp. Acceleration in the cusp is well established [Sheldon et al., 1998; Fritz, 2001] and is best observable under Northward B_z conditions. It is also of interest that the most geoeffective precursor for P_2 is a low-speed, high-density structure (top and bottom panels of Fig. 7). In summary, the most geoeffective precursors for P_1 are the least geoeffective for P_2, and vice versa. The contrast between the two regions is discussed in more detail elsewhere [Vassiliadis et al., 2003a].

5. SUMMARY AND DISCUSSION

Input-output analysis of the fluxes in terms of the preceding solar wind/IMF variations shows that geoeffective solar wind inputs are different for each region and at varying degrees: the variance in P_1 flux is partly explained by changes in the solar wind velocity, V_{SW}. The impulse response $H(\tau;L=L_{P1})$ shows a τ=2-3 day delay relative to the arrival of the solar wind. The response occurs for high-speed streams and is particularly well-known for $j_e(L=6.6)$. The impulse response in P_0 peaks more rapidly, at $\tau<1$ day, a response which is consistent with recent observations during magnetic cloud and interplanetary CME passages. A different set of precursors is found for the third region, P_2. Thus the input-output analysis suggests that, as the turbulent interplanetary input varies randomly, it excites the three regions as nonlinear modes of the inner magnetosphere.

Synoptic flux measurements over the entire radial extent of the radiation belts allow us to model the temporal dynamics of the flux and determine the corresponding spatial structure in unprecedented ways. Correlation analysis has revealed three regions of distinct dynamical behavior, identified in regions P_0 (L=3-4), P_1 (L=6-7.5), and P_2 (L>7.5). Flux dynamics in the slot S (L=2.0-3.0) are different from any of the three regions. The dynamics itself as obtained from AR modeling can be classified as diffusion- or convection-like. The functional form in regions S and P_0 is consistent with diffusion while regions P_1 and P_2 include additional convection terms in their equations.

In practical terms, a first notable point is that the differences between regions P_i should be taken into account in regional modeling. A model which is accurate in reproducing fluxes at L=6.6 will be significantly than a model at lower L shells in P_1 or P_2. Second, a combination of the FIR and AR models (Eq. (6) and (3)) is expected to result in more sophisticated models with better prediction capabilities for $j_e(t;L)$.

More generally, dynamic modeling is expected to contribute to our understanding and lead to improvement in the predictive capabilities of radiation environment models. These models are used in tandem with radiation effects models to quantify the space weather hazards on specific spacecraft components. Current models have evolved from the static NASA AE-8 model [Vette, 1991], but are still limited in spatial and temporal coverage [Heynderickx, 2002]. This is the main reason due to which the traditional models are still the industry standard. Integrated approaches such as the interaction between empirical and physical models promoted by the Center of Integrated Space Weather Modeling (CISM) are expected to further improve the predictability of the radiation belt environment.

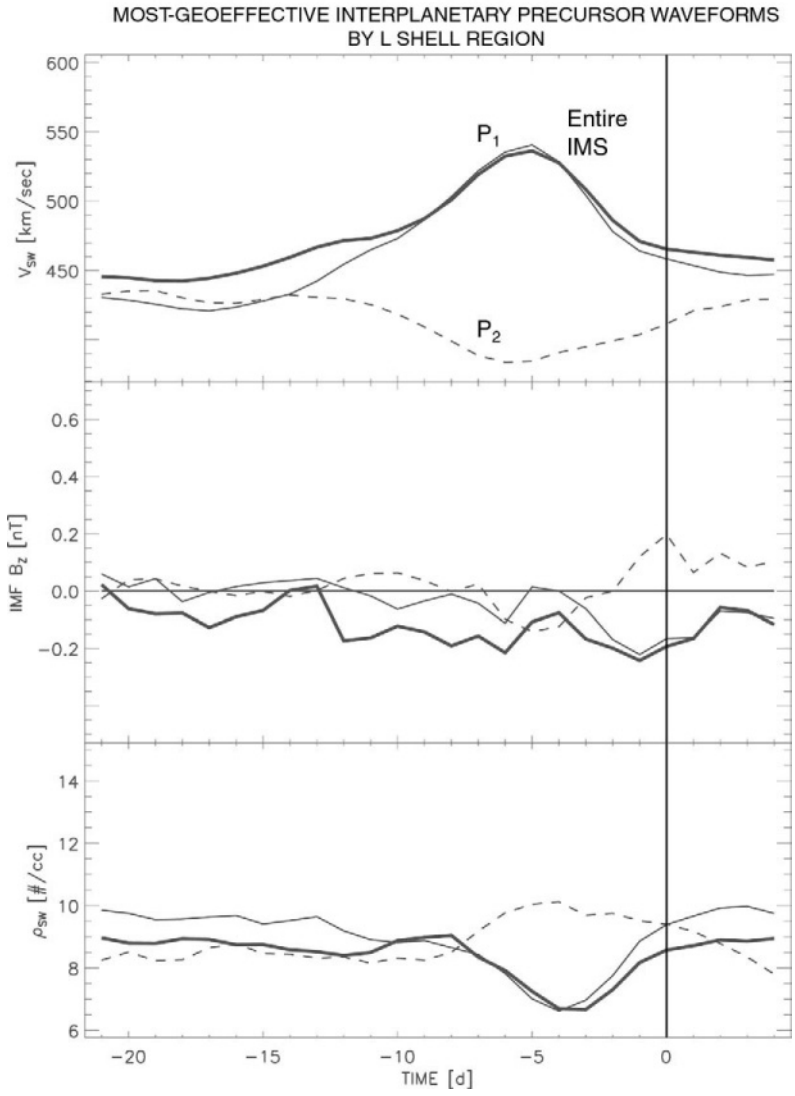

Figure 7. Precursors for P_1, P_2, and the entire inner magnetosphere (L=[1-10]). Precursors are shown in terms of V_{SW}, IMF B_z, and ρ_{SW} for the last T=20 days before a high-amplitude event (also shown are the T_s=5 days following the event).

6. ACKNOWLEDGMENTS

We are thankful to the ESPRIT organizers, and especially Ioannis Daglis, for the invitation to present this review to the Advanced Research Workshop. We thank J. Fennell for discussions, and the referee for a number of comments and suggestions. T. Nagai and R. Friedel provided flux measurements from several spacecraft. Research was supported by the NASA/LWS TR&T program and the NASA/NSSDC research program. We also thank the data providers in NSSDC, and the Kyoto and Copenhagen World Data Centers.

7. REFERENCES

Anderson, B.J., M.J. Engebretson, and S.P. Rounds, A statistical study of Pc 3-5 pulsations observed by the AMPTE/CCE magnetic fields experiment 1. Occurrence distributions, J. Geophys. Res. 95, A7, 10,495-10,523, 1990.

Baker, D. N., R. L. McPherron, T. E. Cayton, R. W. Klebesadel, Linear prediction filter analysis of relativistic electron properties at 6.6 RE, J. Geophys. Res. 95, A9, 15,133-15,140, 1990.

Baker, D.N., G.M. Mason, O. Figueroa, G. Colon, J.G. Watzin, R.M. Aleman, An overview of the Solar, Anomalous, and Magnetospheric Particle Explorer (SAMPEX) mission, IEEE Trans. Geosci. Rem. Sens., 31, 3, 531-541, 1993.

Baker, D.N., S.G. Kanekal, A.J. Klimas, D. Vassiliadis, T.I. Pulkkinen, Collective phenomena in the inner magnetosphere, Phys. Plasmas 6, 11, 4195-4199, 1999.

Baker, D.N., Satellite anomalies due to space storms, in: *Space Storms and Space Weather Hazards*, I.A. Daglis (ed.), pp. 285-311, NATO Science series, Kluwer Academic Publishers, Dordrecht, 2001.

Blake, J.B., J.F. Fennell, L.M. Friesen, B.M. Johnson, W.A. Kolasinski, D.J. Mabry, J.V. Osborn, S.H. Penzin, E.R. Schnauss, H.E. Spence, D.N. Baker, R. Belian, T.A. Fritz, W. Ford, B. Laubscher, R. Stiglich, R.A. Baraze, M.F. Hilsenrath, W.L. Imhof, J.R. Kilner, J. Mobilia, D.H. Voss, A. Korth, M. Gull, K. Fisher, M. Grande, and D. Hall, CEPPAD, in: Russell, C.T., (ed.), *The Global Geospace Mission*, Kluwer Academic, Dordrecht, the Netherlands, 1995.

Blake, J.B., D.N. Baker, N.Turner, K.W. Ogilvie, R.P. Lepping, Correlation of changes in the outer-zone relativistic-electron population with upstream solar wind and magnetic field measurements, Geophys. Rev. Lett. 24, 8, 927-929, 1997.

Blake, J.B., J.L. Roeder, R.J. Selesnick, D.N. Baker, P. Daly, M. Grande, M. Carter, Cluster-Polar simultaneous observations of energetic particles in the plasma sheet – evidence for radiation belt leakage?, Eos Trans. AGU 83 (47), Fall Meet. Suppl. SM51A-0502, 2002.

Chan, A.A., L. Chen, and R.B. White, Nonlinear interaction of energetic ring current protons with magnetospheric hydromagnetic waves, Geophys. Res. Lett. 16, 1133-1136, 1989.

Cook, W.R., A.C. Cummings, J.R. Cummings, T.L. Garrard, B. Kecman, R.A. Mewaldt, R.S. Selesnick, E.C. Stone, D.N. Baker, T.T. Von Rosenvinge, J.B. Blake, L.B. Callis, PET- A Proton Electron Telescope for studies of magnetospheric, solar, and galactic particles, IEEE Tras. Geosc. Rem. Sens. 31, 3, 565-571, 1993.

Daglis, I.A., B. Wilken, and E.T. Sarris, *Proceedings of the Cluster-RAPID Science Workshop*, *MPAE-W-056-95-01*, Max-Planck-Institut für Aeronomie, Katlenburg-Lindau, Germany, 1995.

Elkington, S.R., M.K. Hudson, A.A. Chan, Acceleration of relativistic electrons via drift-resonant interaction with toroidal-mode Pc5 ULF oscillations, Geophys. Res. Lett. 26, 21, 3273-3276, 1999.

Elkington, S.R., M.K. Hudson, and A.A. Chan, Resonant acceleration and diffusion of outer zone electrons in an asymmetric geomagnetic field, J. Geophys. Res. 108, A3, 1116, doi: 10.1029/2001JA009202, 2003.

Engebretson, M., K.-H. Glassmeier, M. Stellmacher, W.J. Hughes, H. Luehr, The dependence of high-latitude Pc5 wave power on solar wind velocity and on the phase of high-speed solar wind streams, J. Geophys. Res. 103, A11, 26,271-26,283, 1998.

Farrugia, C.J., F.T. Gratton, R.B. Torbert, Viscous-type processes in the solar wind-magnetosphere interaction, Space Sci. Rev. 95, 1-2, 443-456, 2001.

Friedel, R.H.W., G.D. Reeves, T. Obara, Relativistic electron dynamics in the inner magnetosphere – a review, J. Atmos, Sol.-Terr. Phys. 64, 2, 265-282, 2002.

Fritz, T.A., The cusp as a source of magnetospheric energetic particles, currents, and electric fields: a new paradigm, Space Sci. Rev. 95, 469-488, 2001.

Fung, S.F., Recent developments in the NASA trapped radiation models, in: *Radiation Belts: Models and Standards*, J.F. Lemaire, D. Heynderickx, and D.N. Baker (eds.), Geophysical Monograph 97, American Geophysical Union, Washington, DC, 1996.

Fung, S.F., and L.C. Tan, Time correlation of low-altitude relativistic trapped electron fluxes with solar wind speeds, Geophys. Res. Lett. 25, 13, 2361-2364, 1998.

Heynderickx, D., Radiation belt mod elling in the framework of space weather effects and forecasting, J. Atm. Sol.-Terr. Phys. 64, 1687-1700, 2002.

Hilmer, R.V., G.P. Ginet, T.E. Cayton, Enhancement of equatorial energetic electron fluxes near L=4.2 as a result of high-speed solar wind streams, J. Geophys. Res. 105, 23311-23322, 2000.

Hudson, M.K., S.R. Elkington, J.G. Lyon, C.C. Goodrich, T.J. Rosenberg, Simulation of radiation belt dynamics driven by solar wind variations, in: *Sun-Earth Plasma Connections*, J.L. Burch, R.L. Carovillano, and S.K. Antiochos (eds.), Geophysical Monograph 109, American Geophysical Union, Washington, DC, 1999.

Ingraham, J.C., T.E. Clayton, R.D. Belian, R.H.W. Friedel, M.M. Meier, G.D. Reeves, and M.G. Tuszewski, Substorm injection of relativistic electrons to geosynchronous orbit during the great magnetic storm of March 24, 1991, J. Geophys. Res. 106, 25,759-25,776, 2001.

Kanekal S.D., D.N. Baker, J.B. Blake, Multisatellite measurements of relativistic electrons: global coherence, J. Geophys. Res. 106, A12, 29,721-29,732, 2001.

Lemaire, J.F., From the discovery of radiation belts to space weather perspectives, in: *Space Storms and Space Weather Hazards*, I.A. Daglis (ed.), NATO Science Series, 79-102, 2001.

Li, X., I. Roth, M. Temerin, J.R. Wygant, M.K. Hudson, and J.B. Blake, Simulation of the prompt energization and transport of radiation belt particles during the Marth 24, 1991 SSC, Geophys. Res. Lett. 30, 22, 2423-2426, 1993.

Li, X., and M.A. Temerin, The electron radiation belt, Space Sci. Rev. 95, 1-2, 569-580, 2001.

Li, X., M. Temerin, D.N. Baker, G.D. Reeves, D. Larson, Quantitative prediction of radiation belt electrons at geostationary orbit based on solar wind measurements, Geophys. Res. Lett. 28, 9, 1887-1890, 2001.

Liu, W.W., G. Rostoker, and D. N. Baker, Internal acceleration of relativistic electrons by large-amplitude ULF pulsations, J. Geophys. Res. 104, A8, 17,391-17,407, 1999.

Mathie, R.A., and I.R. Mann, A correlation between extended intervals of ULF wave power and storm-time geosynchronous relativistic electron flux enhancements, Geophys. Res. Lett. 27, 3261-3264, 2000.

Moorer, D.F., D.N. Baker, and S.F. Fung, Estimating outer radiation belt electron flux by data assimilation (unpublished), 1999.

McAdams, K.L., and G.D. Reeves, Non-adiabatic response of relativistic radiation belt electrons to GEM magnetic storms, Geophys. Res. Lett. 28, 1879-1882, 2001.

Nishida, A., Outward diffusion of energetic particles from the Jovian radiation belt, J. Geophys. Res. 81, 1771-1773, 1976.

O'Brien, T.P. K.R. Lorentzen, I.R. Mann, N.P. Meredith, J.B. Blake, J.F. Fennell, M.D. Looper, D.K. Milling, and R.R. Anderson, Energization of relativistic electrons in the presence of ULF power and MeV microbursts: evidence for dual ULF and VLF acceleration, J. Geophys. Res. 108, A8, Art. No. 1329, 2003.

Paulikas, G.A., and J.B. Blake, Effects of the solar wind on magnetospheric dynamics: energetic electrons at the sunchronous orbit, in: *Quantitative Modeling of Magnetospheric Processes*, W.P. Olson (ed.), Geophysical Monograph 21, American Geophysical Union, Washington, DC, 1979.

Pilipenko, V.A., and M.J. Engebretson, Ground images at high-latitudes of ULF wave processes in the outer magnetosphere, J. Atmos. Sol.-Terr. Phys. 64, 2, 183-201, 2002.

Rigler, E.J., D.N. Baker, R. Weigel, D. Vassiliadis, A. Klimas, Adaptive Linear Prediction of Radiation Belt Electrons Using the Kalman Filter, Space Weather (in print), 2004.

Rostoker, G., S. Skone, and D.N. Baker, On the origin of relativistic electrons in the magnetosphere, Geophys. Res. Lett. 25, 3701-3704, 1998.

Selesnick, R.S., and J.B. Blake, On the source location of the radiation belt relativistic electrons, J. Geophys. Res. 105, A2, 2607-2624, 2000.

Sheldon, R.B., H.E. Spence, J.D. Sullivan, T.A. Fritz, J. Chen, The discovery of trapped energetic electrons in the outer cusp, Geophys. Res. Lett. 25, 1825-1828, 1998.

Summers, D., and C.-Y. Ma, Rapid acceleration of electrons in the magnetosphere by fast-mode MHD waves, J. Geophys. Res. 105, 15,887, 2000.

Tsutai, A., C. Mitsui, T. Nagai, Predictions of a geosynchronous electron environment with in situ magnetic field measurements, Earth, Plan. Space 51, 3, 219-233, 1999.

Vassiliadis, D., A. J. Klimas, S. G. Kanekal, D. N. Baker, R. S. Weigel, Long-term average, solar-cycle, and seasonal response of magnetospheric energetic electrons to the solar wind speed, *J. Geophys. Res.*, 10.1029/2001JA000506, 2002.

Vassiliadis, D., R.S. Weigel, A.J. Klimas, S.G. Kanekal, R.A. Mewaldt, Modes of energy transfer between the solar wind and the inner magnetosphere, Phys. Plasmas 10 (2), 463-473, 2003a.

Vassiliadis, D., A.J. Klimas, R. S. Weigel, D. N. Baker, E. J. Rigler, S. G. Kanekal, T. Nagai, S. F. Fung, R. W. H. Friedel, and T. E. Cayton, Structure of Earth's outer radiation belt inferred from long-term electron flux dynamics, Geophys. Res. Lett., 30(19), 2015, doi:10.1029/2003GL017328, 2003b.

Vassiliadis, D., S.F. Fung, and A.J. Klimas, Interplanetary and magnetospheric state parameters for radiation belt electron flux models, submitted, 2004.

Vennerstrom, S., Dayside magnetic ULF power at high latitudes: A possible long-term proxy for the solar wind velocity?, J. Geophys. Res. 104, A5, 10,145-10,157, 1999.

Vette, J.I., The AE-8 trapped electron model environment, NSSDC WDC-A-R&S, 91-24, 1991.

Walt, M., *Introduction to Geomagnetically Trapped Radiation*, Cambridge University Press, Cambridge, UK, 1994.

Chapter 4

The Ion Radiation Belts: Experiments and Models

Mikhail I.Panasyuk

D.V.Skobeltsyn Institute of Nuclear Physics
M.V. Lomonosov Moscow State University,119992 Moscow, Russia

Abstract Several years after the discovery of the Earth's radiation belts (in the mid-sixties) the model describing their formation due to geomagnetic field fluctuations and/or the electrostatic convection field was developed. This model permitted to explain many regularities of the experimentally recorded radiation belt parameters. By the present time vast experimental data on the structure, dynamics and composition of the radiation belts has been acquired. These data permit to make more detailed comparison of physical model predictions and experimental results. This report analyses the experimental data on the ion composition of the radiation belts and corresponding physical models of ion sources and transport in the trapped radiation zone.

Keywords Radiation belts, ions, protons, cosmic rays, radial diffusion.

1. INTRODUCTION

During several years after the discovery of the Earth's radiation belts (RB) (see the survey by J. Lemaire, 2001) the main attention of both experimental physicists and theoreticians was paid to the studies of the electron and proton trapped radiation components. Primarily this was caused by inadequate development of experimental techniques, which did not permit to measure ions, heavier than protons. As a result our understanding of the RB at that time was that their predominant component consisted of protons.

However, somewhat later experimental evidence was obtained, that heavy ions play a significant role in the development of several dynamic processes in the magnetosphere. Experimental studies of ions (with different mass, charge states and energy) are an efficient tool for revealing the regularities in the formation of the spatial-energetic trapped radiation

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 65-90.

structure in the vicinity of the Earth. These studies also permitted to establish the possible sources of trapped radiation in the Earth's vicinity.

It is generally accepted that the lower energy boundary for RB particles are tens of keV, i.e. particles with energies close to those of the ring current. The minimum energy of RB ions can be determined as the energy of a population of particles which are subject predominantly to magnetic drift in the geomagnetic field, unlike ring current particles, since their dynamics is determined by drift in the geomagnetic and electric convection fields. The upper energy of RB ions is determined by the condition of stable trapping in the geomagnetic field according to the Alfven criteria:

$$\rho_L / \rho_m << 1 \qquad (1)$$

Where ρ_L is the particle Larmor radius and ρ_m is the curvature radius of magnetic field line. The maximum particle energy for ions depends on L-shells, and charge state. For protons these energies are of the order of hundreds MeV in the inner radiation zone, i.e. the maximum energy of RB protons coincides with the minimum energies of galactic cosmic rays (GCR). It were GCR that became the first candidate for the source of RB particles, however, later it was discovered that they were by far not the only one. We will consider the possible sources of RB ions and the experimental data confirming their existence.

The structure of the ion RB unlike that of the electron RB, is described by single-maximum profiles of particle intensity with E=const. According to the theory of Tverskoy (1965), the location of ion maximum intensity L_{jm} is determined by the balance equation (see Fig.1):

$$\tau_t(L) \approx \tau_s(L) \qquad (2)$$

where $\tau_t(L)$ is the particle lifetime on a given L - shell controlled by losses, and $\tau_t(L)$ is the particle transport time from the RB boundary to a given $L = L_0$.

Studies of RB ions besides their fundamental aspect also have important applications in such aspects of space physics as 'space weather'. In particular, these applications include phenomena associated with radiation effects (dose effects and single event upsets) and human safety in space (see e.g. the surveys by Panasyuk, 2001and Baker, 2001).

This paper gives an overview of the main experimental results and corresponding physical models, which permit to explain the possible sources, acceleration and transport mechanisms of RB ions.

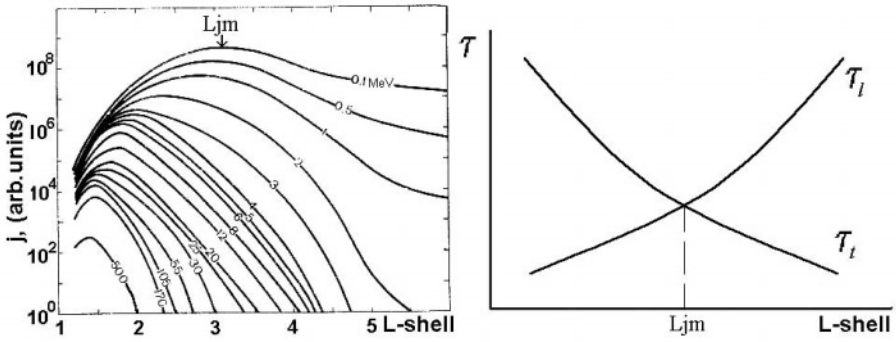

Figure 1. Radial profiles of the equatorial intensities of RB protons with different energies (left panel) and schematic drawing of the transport $\tau_t(L)$ and loss times $\tau_l(L)$ (right panel). The balance $\tau_t \approx \tau_l$ determines the location of maximum intensity for the radial profile L_{jm} of protons (and other ions).

2. SOURCES

2.1 Galactic Cosmic Rays

At present the acquired experimental data is sufficient to prove that this space radiation component as an important source of RB particles. The CRAND (Cosmic Ray Albedo Neutron Decay) mechanism – neutron decay leading to _ generation of secondary protons and electrons ($n \rightarrow p + e + \nu + 782keV$) became the first physical mechanism describing the origin of RB particles.

GCR particles, which have significant momentum, penetrate inside the Earth's magnetic field, reaching the atmosphere they interact with atmospheric nuclei producing neutrons. The products of neutron decay (protons and electrons) become trapped in the geomagnetic field and undergo diffusion transfer, getting accelerated due to betatron acceleration. This scenario was simulated in numerous works and the results give good agreement with experimental data. Apparently there are no reasons to doubt that GCR particles are responsible for the population of the inner RB region (on L <2) with energies exceeding tens of MeV. (see e.g. the model Beutier, et al., 1995).

2.2 Anomalous Cosmic Rays

GCR have a low-energy anomalous component (ACR) (^{16}O, ^{14}N, ^{20}Ne and other ions with energies of 10-20 MeV/nucl which can penetrate inside the geomagnetic field like the main GCR component, since, as a rule, they have the minimum charge state of Q=1+. However, the mechanism of 'secondary' particle production in this case is different. Reaching the atmosphere ACR undergo charge-exchange on neutrals. The charge-exchange products (stripped heavy ions) are trapped by the geomagnetic field, forming a radiation belt. This mechanism was suggested by Blake and Friesen (Blake, Friesen (1977)) and was confirmed in a number of experiments. The first experimental indications of the possibility of the formation of a RB containing ACR ions appeared in the works of Biswas (see e.g. Biswas et al. (1980)). The final proof of the existence of such belts was given in the works of the Russian-American collaboration (Grigorov et al., 1991) and later in the experiment on SAMPEX (Selesnick et al., 1997).

Trapped ACR form a radiation belt on $2< L <3$ and their flux exceeds the flux of ACR in the interplanetary medium by a factor of hundreds. Their lifetimes are quite short (~ months), which is confirmed by practically full coincidence of solar-cycle variations of trapped and interplanetary ACR ions (see Grigorov, et al., 1991).

2.3 Albedo Particles

Besides albedo neutrons from GCR and ACR, undergoing charge-exchange in the atmosphere, there is also experimental evidence confirming the existence of proton and other ion fluxes under the RB, in regions below the geomagnetic cut-off threshold. The origin of these fluxes in not fully understood. These particles can be considered as a potential source of trapped particles in the RB, however, the efficiency of this process also needs additional investigation.

The first indication of the existence of the so-called 'excess radiation' appeared as far back as the middle of the 60-ies . According to data of the 'Cosmos-721' satellite (Basilova, et al. (1973)) the total flux of albedo protons with rigidities R >8-10 GV at altitudes of 200 km near the equator exceeds the flux of primary GCR. According to data of the 'Coronas-I' satellite (Kuznetsov, et al. (2002)) on L =1.1 and 1.6 significant fluxes of albedo protons with E > 80 MeV were observed. These fluxes displayed significant azimuthal asymmetry.

Recently the AMS-01 experiment recorded a very accurate proton spectrum in the region below the geomagnetic cut-off threshold. It is noted, that the fluxes of secondary particles are concentrated in the vicinity of the

equator with significant azimuthal asymmetry. Calculations show, (Plyaskin, 2001) that the presence of these particles is due to the complicated GCR drift trajectories in the geomagnetic field. The long-living component of this population of particles occupies an intermediate position between stably trapped and quasi-trapped particles. They can be considered as the source of particles for the inner RB zone. However, the problems associated with stable trapping of particles with such large rigidities should be thoroughly investigated.

2.4 Solar Energetic Particles

Solar energetic particles (SEP) generated on the Sun during flares or in the process of coronal mass ejections (CME) are surely among the main candidates for the RB particle source. SEP can penetrate inside the magnetosphere to relatively deep L-shells; and during powerful geomagnetic disturbances SEP fluxes are observed in low-altitude (~ 400 km) orbits (LEO).

However, the issue arises of how efficient the mechanism of SEP trapping is for increasing the population of the RB. Some doubts arise as a result of estimating the expression ρ_L / ρ_m (see (1)) for these particles. It should be expected, that the lifetimes of SEP particles inside the RB should be relatively small due to their large rigidities. For SEP protons this effect is vividly demonstrated in many experiments: the gyro-radius of energetic protons is comparable with the curvature radius of the field line in their penetration region (as a rule it is the outer region of RB), therefore, protons leave the trapping region. This effect is enhanced during magnetic storms, when a depression of the magnetic field in the outer RB regions is observed. Hence, the probability of observing long-living SEP protons inside the RB after disturbances on the Sun is small.

We would also like to draw attention to another effect, giving evidence in favor of this statement.

It is known, that for SEP the ratio He/H ~4÷5%. Therefore, it seems reasonable, that for those RB particles which are not subject to losses (e.g. $\tau_t(L) > \tau_s(L)$) similar abundances of these elements should be observed, if their source is SEP.

A characteristic feature of the RB ion energy spectra is their form which displays a maximum. The location of the maximum E_m is determined by losses (Coulomb losses and charge-exchange). At $E > E_m$ the relative abundances of ions should be determined by their source, i.e. the spectrum at the RB boundary. The ‘footprints’ of SEP inside the RB should be sought in the energy range of several MeV and more. Fig. 2 demonstrates the energy

spectra of near-equatorial protons and He ions on L=2.5; 2 and 1.7 (the lower part of Fig.2).

The upper panel of the figure shows the He/H ratio for E/A=const. As it can be seen from the He/H energy dependence, in the energy range $E > E_m$ the value of He/H hardly reaches 10^{-1}. Therefore, if the influence of SEP on the stationary structure of the ion RB can be manifested, it is only on the high energy tails of the energy distributions in the inner RB zone.

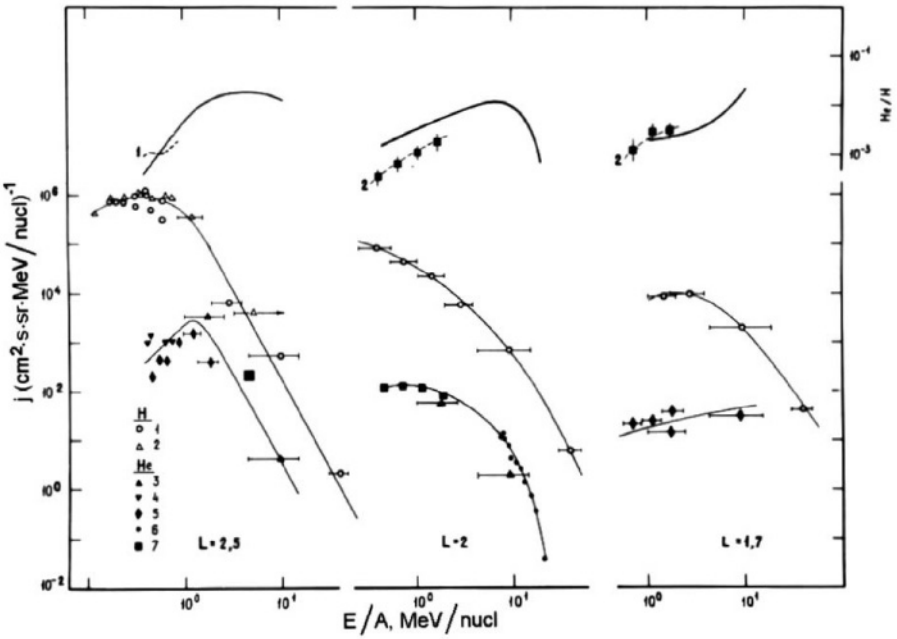

Figure 2. The differential energy spectra of near equatorial H and He fluxes in the inner zone of the RB (the lower part of the figure). References to experiments (1-7) can be found in (Panasyuk, 1983). The upper part of the figure shows the dependence He/H(L) obtained from experimental data approximations.

2.5 Nuclear Reaction Products

The radial intensity profiles of the ion RB have a typical maximum (see Fig.1). The location of the maximum on a certain L-shell is determined by the balance between the transport velocity and particle losses during radial diffusion. However, in a number of experiments measuring ions of He and heavier elements a second intensity maximum, located on L-shells deeper than the diffusion maximum, was discovered. One of the first observations

of this maximum (see Fig.3) was the made in the experiment on 'Intercosmos-17' (Kuznetsov, 1988).

The authors suggested a mechanism for the trapping of nuclear reaction products, generated as a result of high energy (~100 MeV) RB proton interactions with the residual atmosphere (thermal oxygen and helium). Proof of the existence of heavy nuclei fluxes (He, CNO) in the inner RB zone was obtained in other experiments, e.g. on SAMPEX (Cummings, et al. 1994). Therefore, the existence of heavy ions as the products of nuclear reactions in the residual atmosphere can be considered proven.

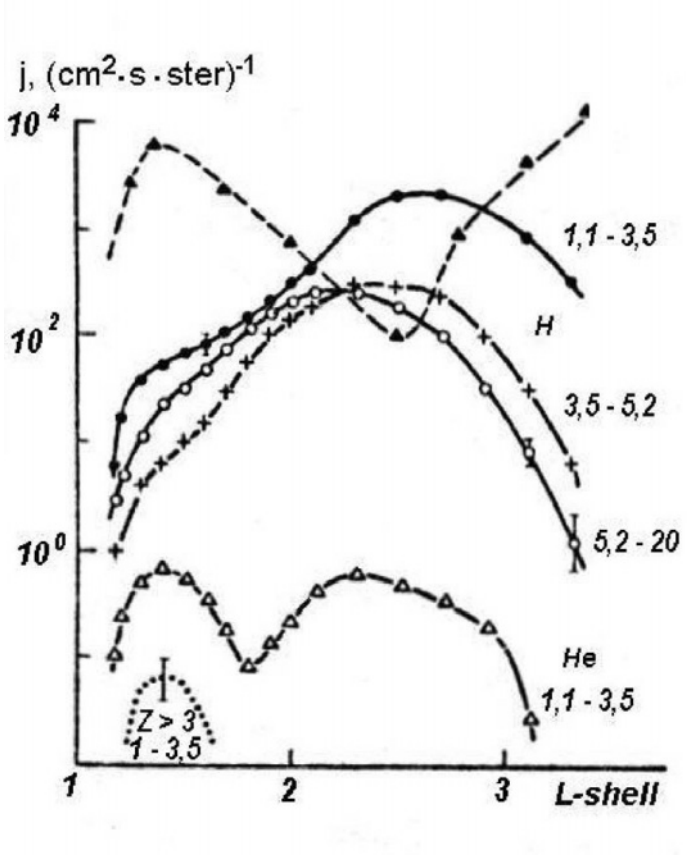

Figure 3. The radial intensity profiles for H, He, and ions with Z>3 of different energies (MeV/nucl) according to data of Intercosmos-17 (Kuznetsov, 1988) at altitudes of ~500 km.

2.6 Plasma Sources

The plasma sources of the ion RB include plasma of solar origin (solar wind) and ionospheric plasma. The major difference between these sources is the elemental composition and ionisation state of the particles.

Solar plasma is characterised by a high content of ^{12}C ($C/O \sim 1$) and high charge states of heavy ions. The ionosphere has low relative content of ^{12}C ($C/O \sim 10^{-5}$) and low charge state of such ions as He, N and O (see the survey by Kremser, et al., 1989). It is these parameters of the ion composition that can serve as indicators of the efficiency of these sources. Evidence in favour of high efficiency of the solar source comes from measurements of the relative content of energetic (in the MeV range) C and O ions in the RB. Fig. 4 demonstrates the energy spectra of H, He, C and O in the near-equatorial plane on L=4 and L=5. The data of ISEE-1 (Hovestadt , et al., 1978) on C and O confirm a high (close to ~1) relative content of C and O, which corresponds to the solar source (both for solar wind and SEP).

Other proof of the efficiency of the solar source could be obtained from determining the charge state of heavy ions in the RB. There are no direct experimental techniques for determining the charge states of ions in the MeV energy range. Therefore, we can once again use the criterion for stable trapping of particles (1) in order to estimate their charge states. From (1) we can obtain:

$$E_c\,(\text{MeV}) \approx 2 \cdot 10^3 Q^2 / A L_c^4 \qquad (3)$$

In expression (3) Q is the charge state of ions, A is the ion mass number, L_c is the outer edge of the intensity profile for particles with E=const. An estimate of L_c according to experimental data (for references see Panasyuk, 1982) gives the possibility to determine Q for different ions. For the first time L_c was used as the stable trapping boundary for protons with a fixed energy in the work of Ilyin, et al., (1986) to determine the constant $\varepsilon = \rho_L / \rho_m$, which was found to be ≈ 0.1.

The results of estimating Q using experimental data on L_c for near equatorial He, C and O ions are shown in Fig. 5 These results give convincing evidence that $Q(He)$=2+ and $Q(C,O)$>5-6+ are close to their maximum values. These estimates of Q for He, C and O in the MeV energy range agree with the results of determining energetic ion charge states according to such events as the ‘drift echo’ in geostationary orbits (see e.g. Sibeck, et al., 1988).

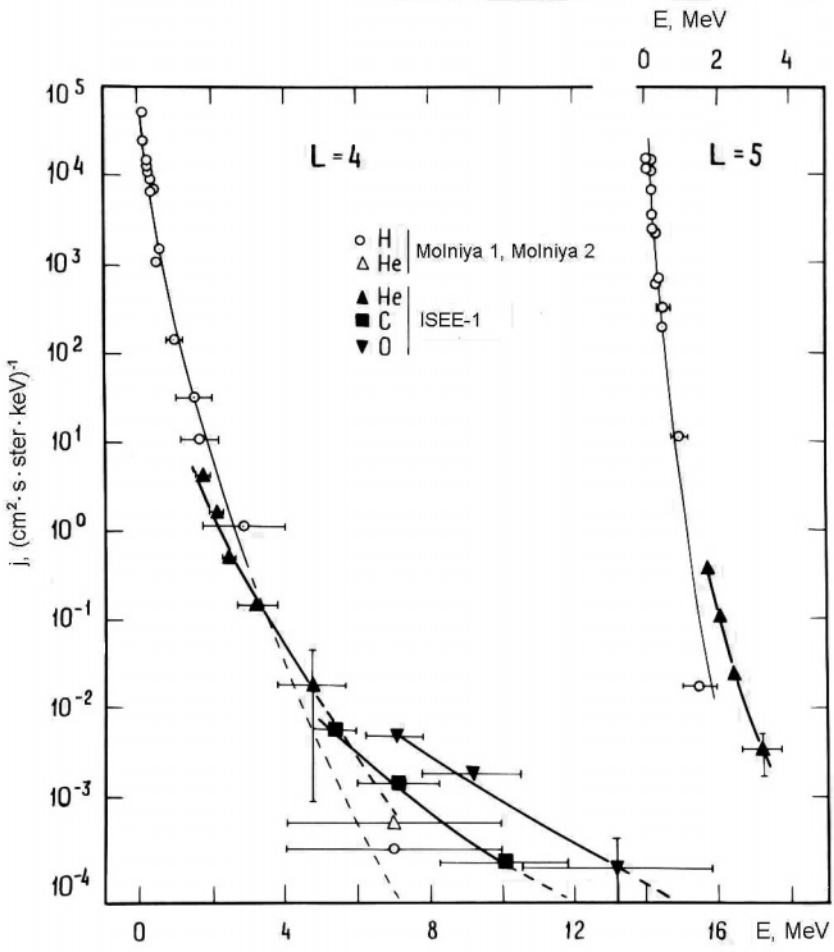

Figure 4. Averaged equatorial spectra of H and He ions on L=4 and L=5 according to data of the Molniya-1 and ISEE satellites. The solid lines are an approximation of the exponential function $j_i(E) \propto \sqrt{E/E_i^*}\exp(-\sqrt{E/E_i^*})$, which demonstrates, scaling of the energy spectra $E_i^* = QE_H^*$, where Q is the charge state of ions in the solar source, E_H^* is the characteristic energy for H.

Estimates of charge states obtained using the dispersion pattern of particle drift also gave Q=2+ for He with energies of hundreds of keV and Q=5+ for $[C, N, O]$ with E>1 MeV. The presence of hot *ionospheric plasma* (i.e. particles with energies E>10 keV) inside the trapped radiation zone has been confirmed in many experiments (see e.g. the survey by Daglis, 2001).

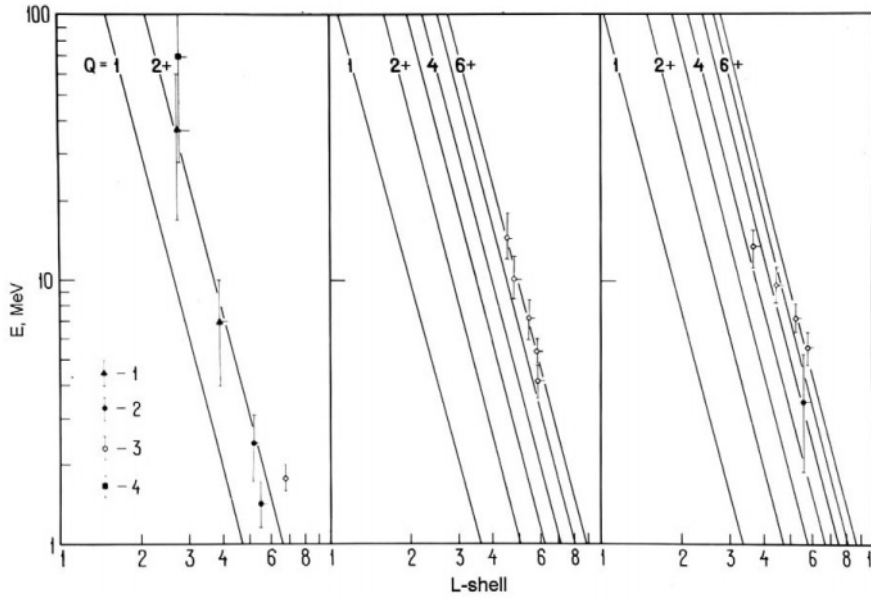

Figure 5. Determining the charge state of ions using expression (3). The lines are calculation results for different values of Q. The values corresponding to the locations of the outer edges of the ion RB - L_c were taken from experiments 1-4 (see text)

The ring current injected inside the RB during magnetic storms mostly consists of H and O ions. The charge distribution shows that the ring current contains both multiply and singly charged ions (Kremser, et al. 1989). This could serve as evidence that the ring current contains ions of both solar and ionospheric origin. However, charge-exchange processes, leading to both increase and decrease of the ion charge states, transform the spectrum typical for a certain source. Therefore, estimates of the relative contribution of these two different sources according just to their charge states can hardly be accurate. Estimates of the contributions of solar and ionospheric sources were also made using comparison of the relative content of He^{2+}/H. However in this case the final conclusions can also hardly be made, since it is impossible to distinguish between ionospheric and solar H, and for He^{2+} the comments made above are also true.

Nevertheless, on the outer shells of the trapping region (e.g. in geostationary orbits), i.e. where transport processes dominate over losses, such estimates can be valid.

At present many authors assert (see the survey of Daglis, 2001), that the relative contribution of the two sources varies depending on the magnitude of magnetic storms. With increasing values of the Dst index the contribution of the ionospheric source (which mostly consist of O^+ ions) increases. In other words, in large and gigantic geomagnetic storms ions of terrestrial origin dominate. The mixture of plasma particles of solar and ionospheric origin in the ring current is definitely one of the main sources of RB particles. Here once again the issue which of the ring current populations has predominant influence on the stationary structure of the ion RB arises.

For the first time this issue was studied by Spjeldvik and Fritz, who developed models of the heavy ion RB (see e.g. Spjeldvik, Fritz, 1978). The author simulated RB, consisting of heavy ions. The spectrum in the diffusion equation was taken as containing either particles of solar (He^{2+}, O^{8+}) or ionospheric (He^+, O^+) ions. The obtained result confirms that the initial charge states of ions are transformed during their transport inside the RB. Hence, it was shown, that the main charge states of O ions with energies exceeding hundreds of keV in the core of the RB is ~ 4+. This makes it impossible to identify experimentally the distribution of particles associated with charge-exchange in those energy ranges where losses dominate.

Similar modeling of He and O in the RB taking into account the solar source, or the ionospheric source, or both these sources of particles was made in Belyaev et al. (1995), but with diffusion coefficients different from those used in the model of Spjeldvik. Diffusion coefficient D_m and D_e corresponding to experimental data on the structure of the ion RB were used. The obtained result is close to the conclusions made by Spjeldvik: the nature of the source does not significantly influence either the charge or energy distributions of ions in the space-energy region where losses dominate (particles with energies exceeding several hundreds of keV on $L \leq 3$).

2.7 The Form of the Injection Spectrum

The currently available experimental data confirm the existence of stationary fluxes of energetic He, C and O ions in the RB predominant over H fluxes at constant energies (see Panasyuk et al., 1977; Spjeldvik, Fritz, 1978; Panasyuk, 1982). In geostationary orbit (L=6.6) an excess of He over H is observed at $E \geq 1$ MeV. In this region the fluxes of O and C exceed the fluxes of He, and $O/H > C/H > He/H$ (Konradi, et al. 1980). Inside the RB a similar pattern is observed for more energetic particles (see Fig. 4).

The author of this work suggested an interpretation of the energetic ion space-energy structure, basing on the existence of E/Q scaling for the exponential energy distributions of ions in the RB (Panasyuk, 1982). In

scope of this idea the existence of dominating ion fluxes at E=const can be easily explained. Simulations have shown, that the use of such solar wind parameters as the charge states Q=2+ for He and $Q = 5 \div 6$ for C and O, along with relative concentrations of these ions give satisfactory agreement with the spectral characteristics of these ions inside the RB. In other words, the energy distribution of RB ions are invariant in the E/Q representation, where Q is the ion charge state, typical for the solar wind. In Fig. 4 an example of such simulation and comparison for L=4 and 5 are shown.

The temperature of the solar corona mainly determines the charge state of the solar wind. Therefore, the observed E/Q scaling of the energy distributions is evidence, that they are a 'response function' of the solar corona temperature.

Can this conclusion be extended to smaller energies, e.g. the ring current or the plasma sheath? We can mention a number of results confirming the existence of E/Q scaling of the ion energy distributions, for both the plasma sheet and the ring current ions. (See e.g. Kremser, 1989). There are cases of E/Q invariant distributions of thermalised particles in the cusp (see Fritz et al., 2002). However, such structure of the distributions is not always observed and significantly depends on geomagnetic activity. (See Kovtyukh, 1999). So far, we can state, that E/Q scaling exists for the energetic component of RB particles, and extension of this conclusion to smaller energies is subject to discussion.

E/Q scaling of the energy distributions is of principle importance for RB particles, since solar plasma and SEP particles typically have energy distributions (distribution functions) in velocity. This can be an indication of the existence of a magnetospheric acceleration mechanism, transforming the initial distributions of solar particles, and forming the RB particle injection spectra with E/Q scaling.

3. ION TRANSPORT

3.1 Diffusion equation for ion transport in the radiation belts

The Fokker-Plank equation describes stationary space-energy distributions of particles inside the radiation belts. For particles with pitch-angles α=90° it has the form:

$$L^2 \frac{\partial}{\partial L}\left(\frac{D_{LL}}{L^2}\frac{\partial f_i}{\partial L}\right) - \frac{G}{\sqrt{\mu}}\frac{\partial f_i}{\partial \mu} - f_i \Lambda_{ce} = 0 \qquad (4)$$

In expression (1):

f_i is the distribution function for *i*- type ions; μ is the magnetic moment; G is the Coulomb factor, Λ is the term describing the charge-exchange process and D_{LL} is the radial diffusion coefficient.

The efficiency of particle transport inside the magnetosphere is determined by D_{LL}. The magnetic and electric diffusion coefficients D_m and D_e are determined by the power spectra of the azimuthally symmetrical parts of fluctuations for the electric $P_e(\nu)$ and magnetic $P_m(\nu)$ fields at ion azimuthal drift frequencies $\nu = \nu_d$.

Taking into account the first spatial harmonic of the Fourier spectrum for the asymmetric part of the field fluctuations of α=90° particles we can write D_{LL} (Falthhammar,1966) as:

$$D_m \propto \nu_d^2 P_m(\nu_d) L^{10} \qquad (5)$$
$$D_e \propto P_e(\nu_d) L^6$$

If $P_{m(e)}$ depends on the drift frequency as $P_{m(e)} \propto \nu_d^{-p(l)}$, where p and l are indices of the fluctuation power spectrum for magnetic and electric fields respectively. From (5) it follows, that for arbitrary values of $p(l)$ the diffusion coefficient $D_{m(e)}$ will differently depend on L, particle energy E and charge state Q, since $\nu_d \propto \mu / L^2 Q$, where μ is the ion magnetic moment.

From (5) the expression for $D_{m(e)}$ in general form can be written as:

$$D_{m(e)} = C_{om(e)} \left(\frac{\mu}{Q}\right)^{v_{m(e)}} L^{u_{m(e)}} \qquad (6)$$

where $v_{m(e)}$=2- p and u_m=2 $p+l$ for magnetic diffusion and $v_e = -l$, u_e=6+2 l for electric and $C_{oe(m)}$ are constants of the diffusion coefficients $D_{m(e)}$. The form of $D_{m(e)}$ according to (3) is valid for power-law fluctuation power spectra of the electric and magnetic fields.

Unlike the power spectra of the geomagnetic field, characteristics of electric convention fields have not been well studied experimentally. Our knowledge of these fields was mainly obtained in observations of cold and hot plasma, whistlers and other magnetospheric phenomena (see e.g. Carpenter ,1972 ; Caufman, Gurnett ,1972; Volland ,1973; Galperin et al., 1980; Mozer,1971; Holthworth, Mozer,1979; Andrews,1980).

From the point of view of revealing the relative importance of each of the transport mechanisms it is necessary to take into account the following

circumstances. According to a number of models, during the movement of hot plasma in the plasma sheath (ring current injection) for a uniform magnetic field a gradient drift (current) arises in the azimuthal direction. This drift causes depletion of the electric convection field in the inner regions of the RB and induces longitudinal currents near the inner boundary of the plasma sheath (see e.g. Alfven, Falthammar ,1967; Tverskoy,1970. There are numerous experimental results (see e.g. Gurnett, Frank,1973; Mozer, Lucht,1974; Southwood, Kaye,1979), which give evidence in favor of electric fields attenuation on L <4. Therefore, it can be expected that the efficiency of electric diffusion decreases in the inner RB regions. Therefore, postulating the uniformity of electric fields in the whole RB region as, for example, it was done in the ion RB models of Cornwall (1968); Cornwall(1971), and later in Spjeldvik (1977) and others is an idealization of the actual convection electric field distribution pattern inside the RB.

However, we should take into account, that for arbitrary field fluctuation power spectra, the structure of the ion RB should be determined by both the amplitude and indices of these spectra. Since $D_{m(e)}$ have different dependencies on L, E and Q, comparison of experimental data on the structure of energetic RB ions is important for determining the predominant mechanism responsible for radial transport.

3.2 Experimental verification of magnetic and electric diffusion efficiency

The particle transport time from the RB boundary (L_b) to a given $L = L_o$ can be written as:

$$\tau_t = \int_{L_b}^{L_0} (\partial L / \partial t)^{-1} \, \partial L \tag{7}$$

where $\partial L / \partial t = -2\partial D / \partial L + 2D / L$ is the particle radial transport velocity. Using expression (1,6) and (7) , and taking into account only the Coulomb losses, we obtain the dependence for $L_{jm}(E)^*$:

$$L_{jm}^{m(e)} = a_{m(e)} E^{S_{m(e)}} \tag{8}$$

where the $m(e)$ indices correspond to magnetic and electric diffusion respectively.

It should be mentioned, that the slopes $s_{m(e)}$ of the energy dependence $L_{jm}(E)$ are determined as:

$$s_m = (p - 3.5)/n - p + 10$$

$$s_e = (l - 1.5)/n - l + 4) \tag{9}$$

i.e. depend both on the spatial distribution of the concentration of cold electrons $N_e \propto L^{-n}$ and on the form of the non-stationary field power spectrum ($(P_{m(e)} \propto \nu^{-p(l)})$.

Besides, L_{jm} for different types of ions at s=const, E =const should differ and their ratio (r) can be described as:

$$r_{m(e)} = L_{jm}({}^{A}I^{Q+})/L_{jm}({}^{1}H^{+}) = \begin{cases} (A^{0.5}Q^{4-p})^{\frac{1}{n-p+10}} \\ (A^{0.5}Q^{2-e})^{\frac{1}{n-l+4}} \end{cases} \tag{10}$$

The conclusions which follow from expressions (7,8) are the following:

1. $s_{m(e)}$ and $r_{m(e)}$ decrease with increasing slope of the cold ion concentration profile $N_e(L) \propto L^{-n}$ and with softening of the $P_{m(e)}$ power spectra.

2. Taking into account that for $N_e(L)$ in the plasmasphere $n \geq 0$, and that definitely it can be expected that $10 \geq p \geq 0$ and $4 \geq l \geq 0$, we come to a more rigid constraint on the slope indices p and l:

a) if $s > 0$, then $p > 3.5$ and $l > 1.5$;

b) if $s < 0$, then $p < 3.5$ and $l < 1.5$.

It should be mentioned, that in a number of models it was assumed (see e.g. Cornwall,1971; Spjeldvik, 1977) that $p = l = 2$. This leads to the opposite slopes for the $L_{jm}(E)$ dependence: $s_m < 0$ and $s_e > 0$. A positive slope $s > 0$ contradicts experimental data. For joint action of both magnetic and electric diffusion this paradox can be resolved only by decreasing the efficiency of electric diffusion (in comparison with magnetic) for RB particles, i.e. a decrease of the electric diffusion coefficient relatively to the magnetic one.

In order to determine the efficiency of one or another transport mechanism it is very convenient to use the experimental energy distributions $L_{jm}(E)$. It is here, that the structure of the ion RB specifically helps to solve the multi-parametric problem of determining the input parameters for the particle transport equation. We will consider these results in more detail.

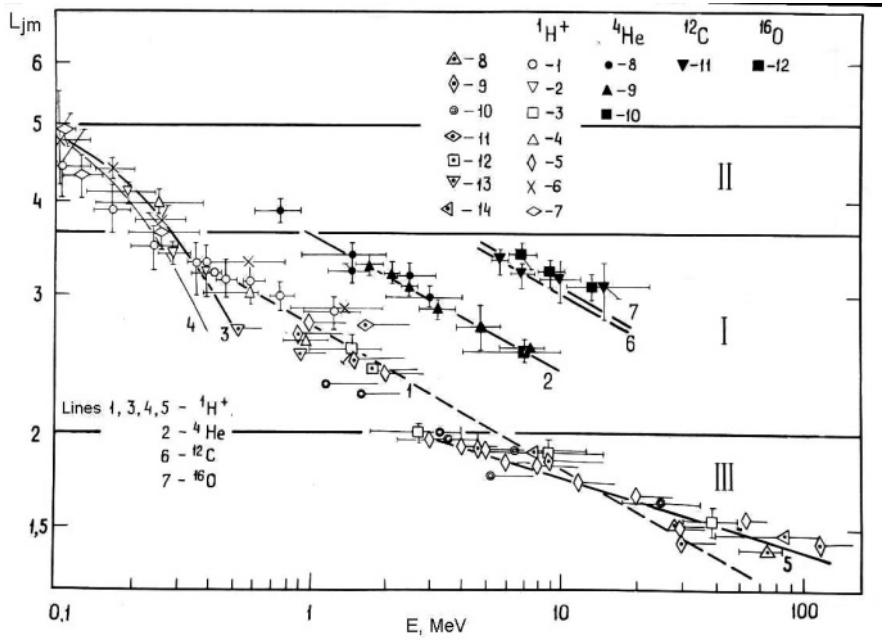

Figure 6. The energy dependence for the maximums of equatorial flux radial profiles L_{jm} for H, He, C and O according to data of experiments 1-12 (see text)

Figure 6 shows the data on L_{jm} for H, He, C and O ions during magnetically quiet times near the geomagnetic equator obtained in different experiments (for references see Panasyuk,1984). The main conclusions, which can be drawn from the analysis of these results are the following:

1. The L_{jm} dependence for protons cannot be represented by a single power law dependence $L_{jm} \propto E^s$ with s=const. Three energy ranges for protons where the s parameter is different (I, II, III) can be pointed out. The experimental data for He, C and O are insufficient for making a similar conclusion.

2. In the core of the RB on $3.5 \le L \le 2$ curves 1 and 2 correspond to the approximation $L_{jm}=2.8 \cdot E^{-0.18\pm0.03}$ for protons and $L_{jm}=3.5 \cdot E^{-0.18\pm0.03}$ for He ions. In concordance with this it follows from (8) that $r=1.3\pm0.02$ Using these estimates we defined the values of $\overline{p}$, $\overline{l}$ and $\overline{u}$ from expressions (7,8) (considering magnetic and electric diffusion to be independent). These results are brought together in Table 1.

Ions	$\bar{p}$	$\bar{l}$	$\bar{n}$	$C_{om}(e)$
H^+	2.2±0.5		0±1.4	$(1.8^{+8.1}_{-1.6})\cdot 10^{-14}$
H_e^{2+}	2.1±0.4		0±1.2	$(1.2^{+6.0}_{-0.5})\cdot 10^{-14}$
H^+		0.3±0.5	3.6±1.2	$(8.1^{+14.4}_{-5.0})\cdot 10^{-12}$
H_e^{2+}		0.1±0.4	3.8±1.2	$(3.3^{+4.4}_{-1.8})\cdot 10^{-12}$

Table 1. The calculation results for $\bar{p}$ $\bar{l}$ $\bar{n}$ and various space-weather environments and $C_{om}(e)$.

Using the obtained values of $\bar{p}$, $\bar{l}$, and $\bar{n}$ and assuming N_e=10^3 cm^{-3} on $L=3$ (which agrees with numerous experimental data, see e.g. Kawashima et al. (1984)) estimates of C_{0m} and C_{0e} were made. The value of C_{0m}=(1-10)·10^{-14} is in good agreement with average disturbed conditions in the magnetosphere at K_p=2÷3. (Panasyuk, Sosnovets ,1984). It is also in good agreement with the conclusions of the magnetic diffusion model developed by Tverskoy (1964, 1965), where C_{0m} was calculated using the statistical distribution of SSC geomagnetic disturbances. The values p=2 and n=0 also concur with this model. It should be noted, that the value of p=2 for the magnetic fluctuation spectrum assumes that D_m does not depend on E, Q and A of the ion components.

Basing on the conclusion that magnetic diffusion is predominant in region *I* of the RB and using expressions (7,8) we can obtain $L_{jm}(E)$ for C and O ions under the assumption that their mean charge states are $\bar{Q}$=5+ and 6+. Calculations give satisfactory agreement with the experiment (curves 6 and 7 for C and O in Fig. 6)

On the other hand, the attempt to explain the distribution $L_{jm}(E)$ for H and He using only electric diffusion encounters difficulties, since a flat spectrum of electric field fluctuations hardly corresponds to theoretical and experimental results (see, e.g., Andrews ,1980 where $l \approx 1$).

3. Assuming, that the values of the $C_{0m}(e)$ diffusion coefficients defined for region *I*, are preserved for the outer RB region *II* (L<3.5), the dependencies of $L_{jm}(E)$ were calculated for protons with account for their charge-exchange on neutrals of the exosphere with the temperature of T_{ex}=950K (Panasyuk,1984). Comparison of the curves describing magnetic diffusion – 3 and electric diffusion – 4 shows, that in the case of electric diffusion C_{0e} needs to be increased by a factor of ~5 in comparison to that of region *I* in order to achieve satisfactory agreement with experimental data.

On the other hand, magnetic diffusion with $C_{0m} \approx 2 \cdot 10^{-14}$ s^{-1} gives us satisfactory agreement both in region RB(*I*) and RB(*II*).

Summing up the analysis of the $L_{jm}(E)$ dependence for ions on $L > 2$, we can make the following conclusions:

Agreement between calculated and experimental data on L_{jm} for H, He, C and O ions on $2 < L < 3.5$ is achieved either for a flat electric field fluctuation spectrum with $l \approx 0$, or for a magnetic field fluctuation spectrum with $p \approx 2$. The shape of the electric field spectrum with $l \approx 0$ does not agree with the results of direct measurements on L=6 (Holthworth, Mozer, 1979) and on L=2,3 (Andrews ,1980). On the other hand, a softer spectrum of electric field fluctuations with $l > 0$ cannot explain the flux distribution at these L-shells for energies of hundreds of keV and more. In the outer region, as it was mentioned above, the presence of solely magnetic diffusion is sufficient to explain the proton RB structure. This leads to the conclusion that the role of electric diffusion on $L > 2$ is insignificant.

Let us now consider the inner zone of the RB - *III.*

4. In the inner region *III* a different slope is observed $s \approx -0.1$ (line 5). If we assume that the power spectra for both magnetic and electric fields preserve their shape in the inner region of the RB this leads to a cold plasma distribution which is too steep ($\bar{n} > 6$) and the necessity to increase $C_{0m(e)}$ by an order of magnitude (see (7)).

A possible mechanism of local increase of D_m (on $L \sim 2$) could be resonant interactions of particles with quasi-periodic fluctuations of the geomagnetic field with the period of several minutes (Panasyuk, Sosnovets 1984). The amplitude of these fluctuations exceeds the regular spectrum of fluctuation power $D_m(\nu)$ with $p \approx 2$. Such oscillations of the geomagnetic field (which can be assigned to the Pc 3 - Pc 4 type) with periodicity of 8-10 min are observed practically continuously both in space and on-ground and are a typical phenomenon for magnetically quiet ($K_p < 2$) periods. (see Fig. 9).

These data show that $P_m(\nu)$ can hardly be described by a single power law of the $P_m(\nu) \propto \nu^{-p}$ with $p \approx 2$ in the whole range of drift frequencies of RB particles.

The predominant spectral density at frequencies of several mHz (in comparison with that expected from the dependence $P_m(\nu) \propto \nu^{-2}$ should lead to an increase of D_m for the ion longitudinal drift period T_φ, corresponding to this frequency range.

Calculations of L_{jm} were made for protons near $L \approx 2$ (Panasyuk, Sosnovets (1984)). Protons with energies $E \approx$ 2-5 MeV in this region have drift frequencies close to characteristic magnetosphere oscillation periods, mentioned above. Therefore the shift of L_{jm} to smaller L for this range of energies E in the inner zone of the RB could be caused by the presence of this type of predominant spectral components of the geomagnetic field power spectrum.

The results of quantitative analysis for this assumption are shown in Fig. 8. In the calculations the predominant harmonics were simulated using gaussians of different amplitude. According to this the C_{0m} constants of the different diffusion coefficients (2,3,4 in Fig.8), which exceed the typical C_{0m} value equal to $2 \cdot 10^{-14}$ s^{-1} (dashed line 1) of the magnetic diffusion coefficient $D_m = C_{0m} L^{10}$ were calculated. In concordance with these C_{0m} values the model dependencies of $L_{jm}(E)$ were calculated. Comparison with experimental data actually shows, that calculations of $L_{jm}(E)$ in concordance with D_m, defined by power spectra 2 and 3 is close to experimentally observed values of L_{jm} in the $2 < E < 30$ MeV energy range for protons. Here an increase of D_m by a factor of 2-4 at frequencies of several mHz is observed in comparison to $C_{0m} = 2 \cdot 10^{-14}$ s^{-1} typical for $L > 2$.

However, this mechanism cannot be extended to protons with energies of tens and hundreds of MeV. In this energy range along with radial transport of particles from the RB boundary, there is a more efficient mechanism – particle injection due to decay of albedo neutrons usually called the CRAND mechanism (see section 1). Therefore, the deviations of experimental L_{jm} in the energy range of tens-hundreds MeV on $L < 2$ from those predicted by the dependence $L_{jm} \propto E^s$ in the $L > 2$ region could be associated with the CRAND mechanism. Actually, calculations of the proton belt structure in the inner RB zone, taking into account the CRAND mechanism (made by e.g. Beutier, et al. 1995), show satisfactory agreement with experimental data.

Above we analyzed the structure of the ion RB from the point of view of independently acting magnetic and electric diffusion. We will consider in more detail the issue of the relative contribution of these two types of particle transport.

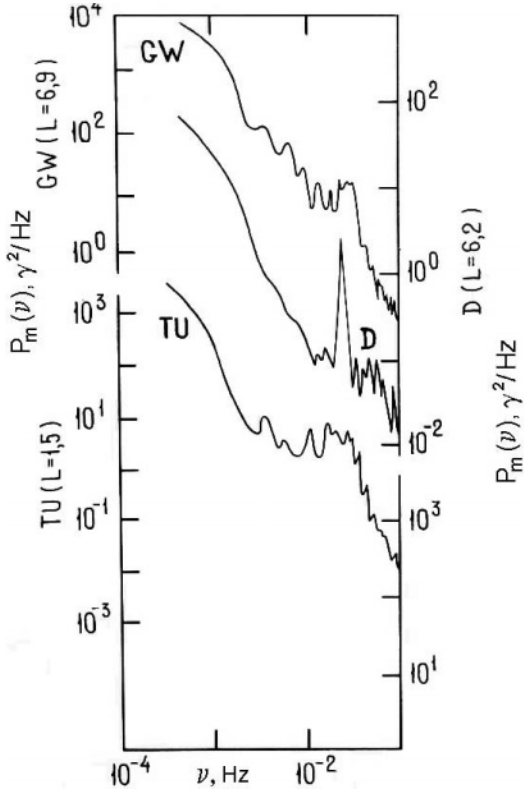

Figure 7. Power spectra of geomagnetic field fluctuations according to data of ground stations Thule (TU), Great Whale (GW) and the Dodge satellite in geostationary orbit.

Fig. 9 shows the experimental data on direct measurements of electric fields in the magnetosphere. These data actually support the model, which assumes electric field attenuation inside the plasmasphere. Basing on this fact we suggested an exponential form of the radial dependence:

$$P_e(\nu, L) = P_{0e}(L)\nu^{-l},$$

where

$$P_{0e}(L) = P_0 \exp(\beta \cdot L^{\alpha}) \qquad (11)$$

Here coefficients P_0, α, and β were matched with direct experimental measurements of the power spectra of electric field fluctuations.

Using relation (11) the values of D_l were calculated for different slope indices $P_e(\nu) \propto \nu^{-l}$: l =0,1 and 2.

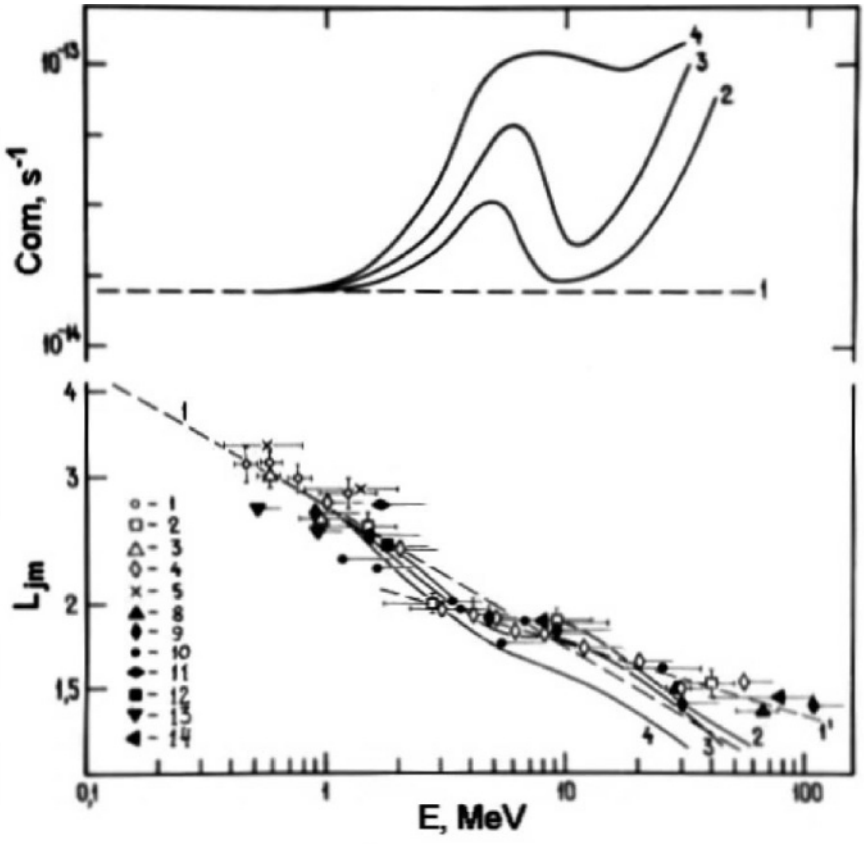

Figure 8. Simulation of the proton L_{jm} for different diffusion coefficients C_{0m} (1,2,3,4) with account for the predominant harmonics (see Fig. 7) of the geomagnetic field fluctuation spectrum.

Comparison of D_e and $D_m = 2 \cdot 10^{-14} \cdot L^{10}$ (the characteristic value for $K_p \leq 2 \div 3$) are shown in the lower part of Fig.9 in the form of a $(E/Q)_{me} = f(L)$ dependence, where E is in MeV.

The calculations presented above, show that magnetic diffusion should be the main transport mechanism in the RB at E/Q exceeding hundreds of keV/Q on $2 < L < 5$. In the inner regions of the RB on $L < 2$ the region of efficient electric diffusion can extend to energies exceeding several MeV/Q. A similar situation occurs on the outer L shells, where the region of efficient impact of electric diffusion can also reach several MeV/Q (for a harder fluctuation spectrum with $l = 1$).

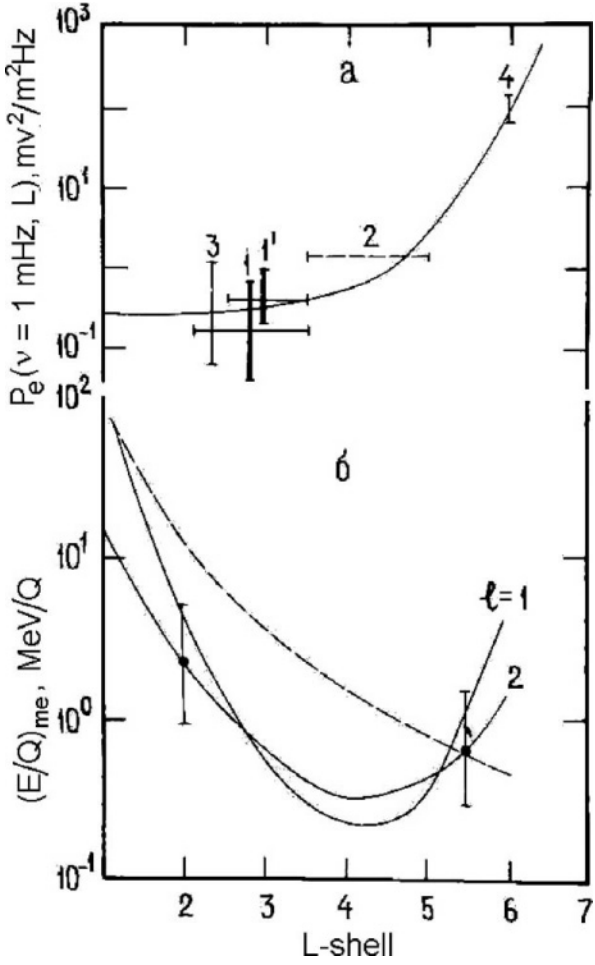

Figure 9. The upper panel shows the model dependence of the amplitude of the power spectrum for electric field fluctuations at ν=1 mHz versus L and comparison with direct experimental data (see text). Below we show the regions of predominant electric and magnetic diffusion for different values of the l parameter of the $P_e(\nu) \propto \nu^{-l}$ power spectrum

The physical meaning of the obtained result is that larger velocities of electric diffusion in the outer regions of the RB are achieved due to a relatively large value of the spectral density of electric field fluctuations in comparison to inner L-shells, therefore, here D_e exceeds D_m. With decreasing distance to the Earth $P_e(\nu, L)$ decreases displaying a plateau in the region of small L-shell values. However, since $D_e \propto L^6$, and

$D_m \propto L^{10}$, in the inner RB zone D_e once again approaches D_m and in the case of a hard fluctuation spectrum can exceed it. This could be the reason for the increase of particle transport velocity in the inner belt on L<2 for energies of several MeV/Q. And, consequently the shift of L_{jm} to L shells lower, than it could be expected for solely magnetic diffusion.

The analysis presented in this section corresponds to the space-energy structure of the ion RB, typical for magnetically quiet times, i.e. the stationary structure. Analysis of the particle dynamics in the RB lies outside the scope of this survey. However, it should be mentioned, that sometimes fast (pulse) injection of energetic ions into the RB is observed. This injection is associated with the influence of single powerful pulses of the magnetic and electric fields (SSC), which have specific form, and are associated with the arrival of a shock. For the first time this effect was observed on the CRRES satellite (Blake, et al.,1992) and was interpreted in the paper of Tverskoy (see Pavlov et al. (1993)). Such effects are very rare: since 1960 not more than 6 such effects have been reported (Loretzen et al., 2002).

4. CONCLUSIONS

1. Radial diffusion is the main transport mechanism for RB ions. Magnetic fluctuations play the main role in the formation of the space-energy distributions of RB particles. Fluctuations of the electric field can be responsible for the diffusion of low-energy particles (below 300 keV) in the outer zone of the RB, and, possibly play an important role in the inner zone. Obviously, the power spectrum of electrostatic field fluctuations is not flat, and is subject to attenuation inside the plasmasphere. Radial diffusion is a slow process. Rapid variations (on the scale of minutes) of the stationary ion RB structure, can be caused by powerful sudden pulses, which have a characteristic shape and are associated with the arrival of shocks from the Sun.
2. There are several sources of RB ions; among them galactic cosmic rays, solar and ionospheric plasma are the most important ones. Ionospheric plasma cannot be regarded an efficient source of particles with energies exceeding hundreds of keV inside the RB. For energetic RB ions (more than hundreds of keV) scaling of the energy distributions in E/Q is observed, where Q is the charge state of ions, typical for the interplanetary medium. This type of scaling differs from that of solar ions, which typically have a E/A representation.

5. ACKNOWLEDGEMENTS

The author would like to warmly thank the organizers of the NATO Advanced Research Workshop on Effects of Space Weather on Technology Infrastructure. The author is extremely grateful to Dr. Ioannis Daglis for providing the opportunity to participate in the ARW and to present this report. I also wish to thank Dr. Ekaterina Tolstaya for translation and editing this manuscript.

6. REFERENCES

Alfven H., Falthammar C-G., Cosmic Electrodynamics, Oxford, 1967.

Andrews M.K., Power density of equatorial electric field at L=2,3, J. Geophys. Res., 85, 1687-1694, 1980.

Baker D.N., Satellite anomalies due to space storms, in Space Storms and Space Weather Hazards, edited by I.A. Daglis, NATO Science series, 285-311, 2001.

Basilova R.N., Grigorov N.L., Kalinkin L.F., Excess fluxes of relativistic particles beneath the radiation belts, Cosmic Research, 11, 627-635, 1973 (in Russian).

Belyaev A.A., Koroteeva E.G., Panasyuk M.I., Modeling of oxygen ion spatial-energetic distributions in the Earth's radiation belts, Cosmic Research, 33, 322-325, 1995 (in Russian).

Biswas S., Durgaprasad N., Skylab measurements of low-energy cosmic rays, Space Sci. Rev., 25, 285-327, 1980.

Blake J.B., Friezen L.M., A technique to determine the charge states of the anomalous low-energy cosmic rays, Proc. 15th Intern. Cosmic Ray Conf., 2, 341-346, 1977.

Blake J.B., Kolasinski W.A., Fillins K.W., Mullen E.G., Injection of electrons and protons with energies of tens of MeV into L<3 on 24 March 1991, Geophys. Res. Letters, 19, 821-824, 1992.

Beutier T.D., Boscher D., France M., Salambo: A three dimensional simulation of the proton radiation belt, J. Geophys. Res. 100, 17181-17195, 1995.

Carpenter,D.L., Stone K., Siren J.C., Magnetospheric electric fields deduced from drifting whistler paths, J. Geophys Res., 77, 2819-2825, 1972.

Caufman D.P., Gurnett D.A., Satellite measurements of high latitude convection electric fields, Space Sci. Rev., 13, 369-392, 1972.

Cornwall J.M., Diffusion processes influenced by conjugate-point wave phenomena, Radio-Science, 3, 740-745, 1968.

Cornwall J.M., Transport and loss processes for magnetospheric helium, J. Geophys Res., 76, 264-267, 1971.

Cummings J.R., Mewaldt R.A., Selesnik R.S., Stone E.C., Blake J.B., Looper M.D., MAST observations of high energy trapped helium nuclei, EOS Trans., AGU, 75, 303, 1994.

Daglis I.A., Space storms, ring current and space-atmosphere coupling, in Space Storms and Space Weather Hazards, edited by I.A. Daglis, NATO Science Series, 1-42, 2001.

Falthammar C-G., Radial diffusion by violating of the third adiabatic invariant, in: Earth's particles and fields, edited by McCormac B.M., Reinhold Book Corp., v.4, pp.157-169, 1966.

Fritz T.A., Spjeldvik W.N., Observations of energetic radiation belt helium ions at the geomagnetic equator during quiet conditions, J. Geophys. Res., 83, 2579-2586, 1978

Fritz T.A., Feunell J.F., Zurbuchen T.H., Perry C., Grande M., Friedel R., Gloeckler G., Hefti S., Chen J., The use of iron charge state changes as a tracer for solar wind entry and energization, Ann. Geophysical, submitted in 2002.

Galperin Yu.I.., Gladyshev V. A.., Kuzmin A.K., Mullarschick T.M., Spatial inhomogenity of magnetosheath proton precipitation along the dayside cusp from the Arcad experiment, J. Geophys. Res., 85, 5105-5112, 1986.

Grigorov N.L., Kondrat'yeva M.A., Panasyuk M.I., Tretyakova Ch.A., Adams J.H., Blake J.B., Schultz M., Mewaldt R.A., Tylka A.J., Geophys. Res. Letters, 18, 1959-1962, 1991.

Gurnett D.A., Frank L.A., Observed relationship between electric fields and auroral particle precipitation, J. Geophys. Res., 78, 145-152, 1973.

Holthworth R.H., Mozer F.S., Direct evaluation of the radial diffusion coefficient near L=6 due to electric field fluctuations, J.Geophys. Res., 84, 2559-2566, 1979.

Hovestadt D., Gloeckler G., Evidences for solar wind origin of energetic heavy ions in the Earth's radiation belts, Geophys. Res. Lett., 5, 1055-1066, 1978.

Ilyin V.D., Ilyin I.V., Kuznetsov S.N., Stochastic instability of charged particles in the geomagnetic trap, Cosmic Research, 31, 75-86, 1986 (in Russian).

Kawashima N., Akai K., Murasato Y., Sasaki S., Electron beam emission from the EXO-5-B (Jikiken) satellite as a powerful diagnostic tool in the magnetosphere, Astrophysics and Space Sci., 106, 117-123, 1984.

Konradi A., Fritz T.A., Su S.Y., Time-averaged fluxes of heavy ions at the geostationary orbit, J. Geophys. Res., 85, 5149-5152, 1980.

Kovtyukh A.S., Dynamics of the main energetic spectra parameters of ionospheric and solar ions in the ring current and plasma sheet, Cosmic Research, 39, 24-33, 1999.

Kremser G., Studemann W., Wilken B., Gloeckler G., Average spatial distributions of energetic O^{+}, O^{2+}, O^{6+} and C^{6+} ions in the magnetosphere observed by AMPTE/CCE, J. Geophys. Res., 12, 847-850, 1989.

Kuznetsov S.N., Measurements of energetic ions at 500 km altitudes in the Earth's inner radiation belt. Izvestiya Akademii Nauk SSSR, 52, 821-823, 1988 .(in Russian).

Kuznetsov S.N., Nymmik R.A., Ryumin S.P., Yushkov B.Yu, Kudela K., Bucik R.,, Energetic charged particle fluxes under the radiation belts, Proc. 27th Int. Cosmic Ray Conf., 2, 329-333, 2002.

Lemaire J.F., From discovery of radiation belts to space weather perspectives, in Space Storms and Space Weather Hazards, edited by I.A. Daglis, NATO Science Series, 79-102, 2001.

Lorentzen I.R., Mazar J.E., Looper M.D., Feunell J.F., Blake J.B., Multisatellite observations of MeV ions injected during storms. J. Geophys. Res, A9, SMP 7-1-7-11, 2002

Mozer F.S., Power spectra of the magnetospheric electric field, J. Geophys. Res., 76, 3651-3658, 1971.

Mozer F.S., Lucht P., The average auroral zone electric field, J. Geophys. Res., 79, 3215-3220, 1974.

Panasyuk M.I., Reyzman S.Ya., Sosnovets E.N., Experimental results of α-particle measurements with energy more than 1 MeV/nucl., Cosmic Research, 15, 887-894, 1977. (in Russian).

Panasyuk M.I., Energetic ions of the solar origin in the radiation belts, Izvestiya Akademii Nauk SSSR, 47, 1850-1857, 1983 (in Russian).

Panasyuk M.I., Sosnovets E.N., Formation of proton radiation belts in the energy range of several MeV, Cosmic Res., 22, 756-762, 1984 (in Russian).

Panasyuk M.I., Experimental test of the transport mechanisms in the Earth's radiation belts, Cosmic Res., 22, 572-587, 1984 (in Russian).

Panasyuk M.I., Cosmic Ray and Radiation Hazards for Space Missions, in Space Storms and Space Weather Hazards, edited by I.A. Daglis, NATO Science series, 251-284, 2001.

Pavlov N.N., Tverskaya L.V., Tverskoy B.A., Chuchkov E.A., Variations of energetic particles of radiation belts during the strong magnetic storm on 24-26 March, 1991. Geomagnetism and Aeronomy, 33, 41-46, 1993 (in Russian).

Plyaskin V., Simulation of particle fluxes in the Earth's vicinity, Phys. Letters, B 516, 231-218, 2001.

Selesnik R.S., Mewaldt R.A., Cummings J.R., Multiply charged anomalous cosmic rays above 15 MeV/nucl, Proc. 25th Int. Cosmic Ray Conf., 2, 269-272, 1997.

Sibeck D., Ross K.S., McIntire R.W., Charge states of substorm particle injection, Geophys Res. Letters, 1283-1286, 1988.

Southwood D.J., Kaye S.M., Drift boundary approximation in simple magnetospheric convection models, J. Geophys. Res., 84, 5773-5779, 1979.

Spjeldvik W.N., Equilibrium structure of equatorial mirroring radiation belt protons, J. Geophys. Res., 9, 2801-2815, 1977.

Spjeldvik W.N., Fritz T.A., Theory for energy states of energetic oxygen ions in the Earth's radiation belts, J. Geophys. Res., 83, 1583-1598, 1978.

Tverskoy B.A., Dynamics of the Earth's Radiation belts, Geomagnetizm and Aeronomy, 4, 436-441, 1964.

Tverskoy B.A., Transport and acceleration of charged particles in the Earth's magnetosphere, Geomagnetizm and Aeronomy, 5, 793-798, 1965 (in Russian).

Tverskoy B.A., Electric fields in the magnetosphere and the origin of trapped particles, Solar Terrestrial Physics, 3, 297-317, 1970.

Van Allen J.A., The geomagnetically trapped corpuscular radiation, J.Geophys. Res., 64, 1683-1689, 1959.

Vernov S.N., Artificial satellite measurements of cosmic radiation. Doklady Akad. Nauk SSSR, 120, 1231-1233, 1958 (in Russian).

Volland H.A., A semiempirical model of large-scale megnetospheric electric field, J. Geophys. Res., 78, 171-178, 1973.

Chapter 5

Outlook on Space Weather Effects on Spacecraft

E.J. Daly
Space Environments and Effects Section,ESTEC, European Space Agency
2200 AG Noordwijk, The Netherlands

Abstract: Spacecraft are becoming more susceptible to space weather hazards for a number of reasons. The types of missions being flown are increasingly demanding and payloads are becoming more sophisticated. In addition, commercial pressures can result in selection of more lightweight spacecraft and less radiation hardened components. Non-availability of radiation-hardened components in some areas can lead to the use of technologies that are sensitive to radiation effects. New types of space weather effects are also emerging. Traditionally the concerns have been with effects such as single event upsets and latch-up, internal and external electrostatic charging, drag effects and some communications effects. Modern systems have to contend with new kinds of problems, for example ion-induced circuit transients, and with increasing complexity associated with other problems such as interference with sensors. To these problems must be added the hazards to man in space, especially in the light of ambitions to progress beyond low Earth orbit. Serious problems persist with capabilities to evaluate space weather hazards to spacecraft. For example, single event upset and internal charging anomalies remain difficult to predict quantitatively. This contribution reviews the trends and problems arising, and proposes actions that are needed to address the problems.

Key words: ESA, space weather, space environments and effects

1. INTRODUCTION

Space missions are becoming more susceptible to space weather hazards for a number of reasons. Spacecraft themselves are becoming more complex, resulting in greater sensitivity to radiation and other environmental effects. The types of missions being flown are also increasingly demanding, both in terms of what they aim to do and where they try to do it. Commercial pressures can result in lighter spacecraft or parts of spacecraft, which implies

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 91-108.

less shielding against radiation. Such pressures can also result in selection of less radiation hardened components. Obtaining radiation-hardened components with the required performance is in any case proving more and more difficult. The nature of effects is also continually changing. For example, systems now have to contend with circuit signal transients induced by ion strikes on analog electronics, while problems related to radiation interference with sensors are increasing.

It is expected that human presence in space will expand. Space agencies around the world have bold ambitions, including putting humans on Mars. Such enterprises entail considerable risks, among which radiation from "space weather events" is a major one.

Serious problems are evident in the techniques used to evaluate quantitatively many of the space weather hazards to spacecraft. As a result, new types of evaluation techniques and supporting development programmes are needed.

Several reviews are available of space weather effects on spacecraft in general and on ESA's spacecraft in particular (e.g. Daly, 2001), and it is not the purpose of this paper to repeat such reviews. It rather seeks to address the ways in which the evolution of mission types and space technologies will change the susceptibilities of space missions to space weather. After a brief overview of space weather effects, future mission trends are presented. The implications for space weather effects are detailed, and the means needed to deal with them discussed. It will be seen that radiation environments and effects receive the most attention. It is the opinion of the author that this is the most important and challenging area for the future and the reader should appreciate that this is a personal view.

2. SPACE WEATHER EFFECTS

Space Weather effects on space systems include:

– Radiation damage to spacecraft electronics, solar cells and materials from Earth's radiation belt particles and solar energetic particles - The Earth's inner (proton) radiation belt is relatively static, arising from cosmic ray atmospheric neutron decay but is affected indirectly via atmospheric changes and can be subject to changes during particularly severe events. The outer (electron) belt by contrast is highly dynamic over short times. Jupiter and Saturn also have intense radiation belts. Since radiation damage is a time-integrated phenomenon, the dynamic behaviour of the environment is not important. The exception is with solar cycle variations; inner belt proton fluxes are higher at solar

minimum when the atmospheric densities are lower and the outer belt electron fluxes are higher on average during the declining phase of the solar activity due to recurrent geomagnetic storms induced by coronal holes;

– Single event effects in spacecraft electronics due to the ionization tracks from galactic cosmic ray or solar energetic particle ions, or due to the ionizing products of nuclear interactions between radiation belt or solar protons and component materials - Since the sources of these particles are strongly time varying, responding to solar and geomagnetic activity, the rates of single event effects such as memory errors can vary significantly. Sudden increases in error rates can cause serious problems for systems;

– Interference to spacecraft imaging and sensing systems - These effects are similar to single event effects. Particles passing through a detector can cause noise in the detector, obscuring the signal being sought. In this case, the particles can also include electrons and the many secondary products of particle interactions with surrounding material;

– Radiation hazards to astronauts - The international space station (ISS) is located close to the Earth, keeping it below the most dangerous parts of the radiation belts and shielded by the Earth's magnetic field from most solar energetic particles and, to some extent, from galactic cosmic rays. Nevertheless, the doses are important and so the environment must be continually monitored. The ISS inclination brings it into contact with the high latitude extensions of the outer radiation belt, where "space-walks" have to take account of enhancements due to geomagnetic storms, as well as exposure to solar energetic particles at these less well-shielded parts of the orbit. Future missions passing through the radiation belts and into interplanetary space, beyond the protection of geomagnetic shielding, will be subjected to more severe radiation hazards and so radiation protection and space weather warning are important elements for these ventures;

– Electrostatic charging from "hot" (~keV electron temperature) plasmas and energetic (~MeV) electrons - In a plasma, because of the higher mobility of electrons, surfaces usually charge negatively. Although in space this behaviour is modified by secondary emission and sunlight induced photoemission, close to the Earth (and close to Jupiter and Saturn) the currents of hot electrons can cause high levels of electrostatic charge to accumulate. Subsequent spontaneous discharging can disrupt electrical systems. The higher energy electrons can give rise to a similar phenomenon inside a spacecraft – charges can build up on internal cable insulation or other dielectric materials and lead to discharges. These phenomena are often closely associated with strong

geomagnetic storms when hot plasma and energetic electrons are injected into near-Earth regions;

- Drag caused by the upper layers of the Earth's atmosphere - Since the upper atmosphere responds to changes in solar radiations and particle inputs, the drag effects observed on satellites are often strongly affected by space weather. Sudden increases in drag can cause premature orbital decay or attitude de-stabilization. Tracking of orbiting objects can also be disrupted during such episodes;
- Interference with electromagnetic signals - The ionosphere responds strongly to changes in solar inputs, geomagnetic storms, and charged particle precipitation. Ionospheric variations can include changes to electron density and therefore to the "total electron content" encountered by a signal, and sources of ionospheric "noise" such as scintillation. As a consequence, space weather seriously disrupts ground-to-ground communications channels that employ the ionosphere, degrades space-based communications, and can cause errors and outages in navigation systems.

During the development of spacecraft, the expected environmental effects are carefully considered. The development process includes definition of the environment, analyses of possible problems caused, and implementation of appropriate measures to avoid or cope with effects. Analyses make use of information on the environment in the form of models and tools that have developed over the years to cope with an evolving set of problems.

3. TRENDS AND FUTURE ISSUES

Future ESA and European space missions will continue across a broad range of domains:

- Science missions in space, using space as a vantage point for astrophysics missions, for solar-terrestrial investigations, for exploration of the solar system and as a place to perform fundamental physics experiments;
- Space applications development programmes such as Earth observation programmes, communications technology development, development of programmes for Global Monitoring for Environment and Security (GMES), and development of navigation systems;
- Commercial space activities, built on the technology and applications development programmes mentioned above, notably including communications, meteorology and navigation systems,

- Manned missions to ISS and beyond, particularly missions to the Moon and Mars;
- Launcher developments.

3.1 Science Programmes

While European science activities will continue in near-earth orbit, there is a trend towards putting astrophysics missions at the L2 Lagrange point for launch, thermal, space environment and communications reasons. Here, the plasma environment is relatively mild, characterised by the deep geomagnetic tail and there are no radiation belt populations. The radiation environment becomes dominated by sporadic solar energetic particle events, in addition to background cosmic rays.

Solar-terrestrial and solar-heliospheric missions are also planned away from the immediate vicinity of Earth, for example the Solar Orbiter mission will approach the Sun to within 21 solar radii. Planetary missions are a strong feature of the future programme, requiring consideration of non-terrestrial atmospheres, magnetospheres and other environments. The BepiColombo mission to Mercury, as with Solar Orbiter, will have to cope with an environment potentially quite different from the near-Earth interplanetary environment. Fundamental physics missions (tests of relativity and detection of gravitational waves for example) also favour locations in deep space although precursor missions could take place close to the Earth.

3.2 Navigation

The principal challenge for Europe in this area is the Galileo programme and the development and maintenance of its constellation of “Galileosat” satellites. Their radiation and plasma environments will be particularly severe, and service quality issues related to interference with propagation of signals through the ionosphere are also an important factor for the system. Clearly, in an eventual commercial environment, there will be a drive to minimise the procurement costs and spacecraft mass.

3.3 Commercial Satellites

The main market for commercial satellites at present is in communications and broadcasting. Here again, procurement costs and spacecraft mass issues are important, as are minimisation of operations effort and long-term reliability. The trends in this sector include more on-board processing and extensive use of commercial off-the-shelf components (COTS). These are usually considerably less radiation hardened or poorly characterised

compared to “traditional” parts and have a greater degree of on-chip complexity.

3.4 Earth Observation Missions

The trend here is towards smaller, lower cost individual missions with a high level of on-board processing or storage. Stability requirements are increasing and the use of electric propulsion, advanced attitude sensing and sensitive payloads lead to increasing environmental susceptibility.

3.5 Manned Missions

Radiation effects are mission-limiting in view of their potentially lethal effects. In the near-term, the international space station will be the main concern. But in the more distant future, returns to the Moon and missions to Mars are expected. ESA’s Aurora programme is intended to prepare for such exploration missions, and the programme includes the necessary preliminary studies and technology activities. While radiation issues can be expected to have a high profile in this programme, other environmental components are also important. Meteoroids and space debris are potentially catastrophic for manned missions, and planetary missions also need good understanding of planetary atmospheres for aero-manoeuvring, entry, decent and landing.

3.6 Platform and Generic Trends

On-board complexity is increasing and advanced platform concepts are being pursued to reduce mass and power, and increase performance. Some technology trends lead to increasing environmental susceptibility, such as: platform wireless and fibre interconnectivity, widespread use of application specific integrated circuits (ASICs), increasing on-chip complexity of components, star-tracker based attitude control systems, electric propulsion, and solar cell technology advances. The rate of change of technology is a problem. Components are quickly introduced, replaced and retired. There is a general acceptance that space mission developments are hampered by the inability to reduce launch costs. As a consequence, there are trends towards smaller, lighter weight spacecraft to get more performance per kilogram.

4. IMPLICATIONS FOR SPACE WEATHER RESEARCH

A significant part of future scientific and technological research on space weather should aim to support the abilities of industries and agencies to execute ambitious programmes. There is a general need to improve “analysis tools” to allow quantitative evaluations of the effects in the various environmental domains. Models of the environment, based on in-flight measurements and on physical principles are needed, along with models of the effects of the environment based on ground-based testing. The links between environmental assessments and testing and ground-based assessments are important and need strengthening. Rigorous methodologies combining testing with application of quantitative assessment tools are needed. An example of what is meant here is the way in which ground-based accelerator testing of electronic components’ susceptibility to single event upset is combined with models of the environment and the single-event upset (SEU) process in sensitive chip geometries. It is also believed that the need for in-orbit experimentation is increasing. This allows testing in environments that cannot be simulated numerically or experimentally on the ground, the proper validation of methodologies and analysis tools, and the gathering of data on environments and effects. In this section the trends in effects of various parts of space weather are outlined taking into account the mission trends presented in Section 3. The research and development directions are indicated in each area. A summary of these space weather components is given in Table 1, along with the main effects of each of them.

4.1 Radiation Environments and Effects

For future space missions, improved models of the radiation environment are needed for each of the environmental sources. Because of the trends in the science programme in particular, better treatment of solar energetic particles is needed, including their time-behaviour and their variation with location in the heliosphere. Closer to Earth, this needs to be accompanied by improved treatment of the modification of solar event and cosmic ray fluxes by the Earth's magnetic field known as geomagnetic shielding. Given the growing importance of applications programmes that have spacecraft located in the radiation belts, better modelling of the energetic electron environment for medium-altitude orbits is needed.

Table 1. The various space weather environments and their effects

Environment	Effects
High Energy Radiation:	
Cosmic Rays	Upsets in electronics;
	Long-term hazards to crew;
	Interference with sensors;
Solar Energetic Particle Events	Radiation damage of various kinds;
	Upsets in electronics;
	Serious prompt hazards to crew;
	Massive interference with sensors;
Radiation Belts	Radiation damage of various kinds;
	Upsets in space electronics;
	Hazards to astronauts;
	Considerable interference with sensors;
	Electrostatic charging and discharges
Near-Earth Plasma Populations:	
Geomagnetic (sub-) storms	Electrostatic charging and discharges;
Ionospheric Effects	Communications disruption;
	Navigation services disruption
Others:	
Atmosphere	Increased drag on spacecraft and debris;
	Attitude perturbation
Meteoroids	Spacecraft damage

Many of the future requirements for environmental data, and assessments of effects arising, are not met by traditional approaches of synthesising space environment data into models. This inevitably involves some averaging or establishment of worst-cases. Data-based analyses are preferable, holding data in a form including time and location information. Then, better statistical information concerning effects end-points such as background, single-event effects and charging effects, which are time-dependent effects, can be derived (Stamper and Hapgood, 2001). Physics-based models, for example of the radiation belts (Boscher et al., 2001), have the ability to provide more extensive information on the environment and in many ways are analogous to satellite data. They can therefore be used to augment such data-driven models.

The following sub-sections highlight some specific issues and requirements.

4.1.1 Solar Particle Environments and Effects Models

The future Science programmes outlined in section 3 require the use of models of the solar particle environment and appropriate models of consequent effects. In addition, the Aurora exploration programme has similar requirements. Current statistical models of solar particle radiation focus on provision of long-term radiation damage estimates. New

requirements include the assessment of the temporal behaviours (durations, peaks, thresholded durations, spectral variations) and sounder treatment of heavy ions in solar particle events. The data-based analysis techniques mentioned in the previous section are applicable to such problems. In addition, the variations of the solar energetic particle environment with position in the heliosphere need to be known for unmanned missions to the inner (<1AU) heliosphere and for manned missions beyond the near-Earth environment. However, the most extensive data sets on solar energetic particles are from spacecraft close to the Earth. Helio-radial variations are therefore very difficult to derive without recourse to models of solar energetic particle acceleration and propagation (Aran et al., 2001), since a significant proportion of the energetic particles are produced in interplanetary shocks. As a result, their variations in space are far more complex than represented by simple engineering rules, such as the commonly used $1/r^2$ scaling (r being the helio-radial location). Finally, the high radiation levels possible necessitate better quality assessment of radiation shielding and radiation effects, whether for automated missions to the inner heliosphere or for manned missions to the Moon, Mars and Lagrange points.

4.1.2 Radiation Effects on Components, Detectors, Solar Arrays and Materials

Severe problems are being encountered in the development of radiation hardened technologies for space. With some export restrictions on US products and the reducing production facilities for rad-hard technologies due to lack of commercial viability, hardened components are becoming increasingly difficult to procure, particularly in the higher performance components needed by users. Together with the increasing radiation sensitivity of payloads on scientific spacecraft, this leads to a requirement to improve assessment methodologies to enable “softer” technologies to be employed. Improvements are needed in tools for predicting the environment, in testing methods and in the gathering of key space environmental and in-orbit technology performance data.

Single event effects in electronics appear to be a growing problem (Harboe-Sørensen, 2002). Modern components are increasingly complex. Memories for example now include control logic that can be susceptible to upsets, “locking” the component, or to destructive latch-up. Manufacturing processes vary considerably making it difficult to ensure that a flight component is from the same manufacturing batch as a tested component. The complexity of modern electronics makes quantitative assessment difficult. For example, in the past one was able to assume that all the

sensitive parts of “bits” on a chip where well approximated by identical rectangular parallelepipeds. This considerably simplified the prediction of single event upset rate since a component could be characterised by a single “path-length distribution”. Modern electronics, including memory devices, have many different logical elements on the chip and these are often not parallelepipeds. In such a situation, many of the assumptions made in the prediction break down. Furthermore, many of the assumptions made in interpreting accelerator test data also break down. Tilting a component with respect to the accelerator ion beam is a common way of trying to mimic ions that produce greater ionisation, but this is only true for flat two-dimensional structures. There has recently been a growth in problems related to “analog SEU”, the single event transient (SET), which is also a complex problem to deal with. This is where an ion strike on linear circuits such as comparators, operational amplifiers or analog-digital converters can result in a transient pulse of variable magnitude on the output. Generally the seriousness of the SET depends on where in the linear circuit the ion strikes, on the settings of the circuit, on the filtering of the signal and on the way in which the signal is used. For example it may be “latched” to indicate a warning condition – with subsequent effects on the spacecraft.

As a result, assessments of single event effects require improved methods for predicting charge generation in complex modern component circuit geometries, and improved testing methods, which allow correlations to be made. In-space data on the behaviour of new component technologies is also very important.

Radiation background in detectors is a related phenomenon, but is often much more complex to analyse because the specific sensitivity of a payload system depends on its application, what background event rates can be coped with and whether there are specific energy (or energy-deposit) threshold in play. Often background can be rejected by software, but such solutions depend on a clear difference in nature of the background events, and good knowledge of them. Analysis depends increasingly on application of Monte-Carlo simulation of the passage of the radiation through the spacecraft and detector and of the interaction with detector elements. Such techniques may also be necessary for electronic effects. An important toolkit for these applications is Geant4, developed and supported by a world-wide collaboration including ESA (Agostinelli et al., 2003). In applying such tools, it is often important to have good environmental input, and methods that synthesize the required input spectra directly from databases of flight measurements are best.

Solar cell assessment methods have been based in the past on a “damage-equivalence” methodology where extensive testing of cell types was used to derive the behaviour in space. With more complex cell types, a more general

approach, utilising direct calculation of the non-ionising dose in the cell materials, and an assessment of cell structure, is necessary. Materials, including optical components, are also become complex and in need of special attention.

4.1.3 Radiation Effects to Crew

For assessment of effects in the future ISS exploitation phase and for missions beyond low Earth orbit, the astronaut *dose equivalent* has to be evaluated. This evaluation relies on data on biological effects of radiation, and international standards defining the biological effectiveness continue to evolve as knowledge improves.

For the environmental part of the problem, particular importance is attached to solar energetic particle events, which are by their nature difficult to predict. Apart from statistical models of the near-Earth environment based on long-term records of solar particle event fluxes, the variability of event characteristics with location in interplanetary space mentioned above has to be dealt with. For real-time protection, means of predicting event occurrence and magnitude based on solar precursors have to be established. Also important is the prediction of shielding effects and consequent secondary radiation production. Finally, equipment for monitoring crew, habitat, and ambient environments are needed.

4.1.4 In-flight Measurements

Continued and improving monitoring of the environment is obviously important for improving knowledge of the environment, but also for providing the resources to respond to emerging requirements and for use in new analysis methods. ESA has successfully promoted the idea of having standard radiation environment monitors (SREMs) on as many missions as possible. These both contribute to the general body of data and provide mission-specific data, which can be used in real-time for protection, or later for evaluation of spacecraft or instrument behaviour. At the time of writing, SREMs are flying on PROBA and INTEGRAL, providing high-quality data (Mohammadzadeh et al., 2003), and the future missions carrying SREMs include Rosetta, Galileo-GSTB-V2, Herschel and Planck. A demonstration version of a scintillator-based miniature radiation monitor is also flying on PROBA. Manned missions have important requirements for monitoring and this may include a tissue equivalent proportional counter (TEPC).

There are a number of other European radiation instruments either ready or under development, and coordination of efforts both within Europe and beyond needs to be established. It is important that proper attention is given

to data merging and quality issues. Experience has shown that different detectors can give very different results making them difficult to use. In addition, detectors often have data gaps or saturate during extreme environments. Such behaviour is often acceptable for science missions but for "applications", it is not. In addition, any detector for applications has to have open development information, with full simulation and calibration history. The International Standards Organization working group on space environment standards has recently highlighted such requirements (Heynderickx, 2003).

4.2 Plasma Environments and Effects

Plasma-induced hazards continue to be a concern (Harris, 2001). While the engineering of spacecraft needs to take account of surface charging and the related material properties and electrical bonding issues (Purvis et al., 1984), design of science missions often requires careful treatment of the electrostatic fields of a spacecraft. The effects of solar array voltages and electric propulsion can cause further complications. Ground testing of system-plasma interactions is virtually impossible so there is a heavy reliance on numerical plasma simulation. Tools exist, such as NASCAP (the NASA Charging Analysis Program, Katz et al., 1979), to predict the electrostatic surface charging of a satellite, but improved tools that can handle finer geometrical details and details of the time behaviour, such as those arising from space weather environmental changes, are needed. A collaboration has recently been established by ESA to go a large way to developing the tools necessary to perform such analyses (Roussel, 2003).

Improved understanding of the environment and its time and spatial variations – rather than crude worst-case specifications – is also needed. Related to this is the need for *operations* support. If satellite-specific forecasting of severe environmental conditions can be made available at a high level of reliability, safeing of system, avoidance of hazardous operations and general operator vigilance can improve the resistance of systems to charging-related anomalies.

As with radiation, there is a clear need to validate any methods or results obtained through numerical or ground experiment simulation in real space conditions. In addition, characterisation of the ambient environment around a spacecraft is required both for gathering data for future missions but also for aiding interpretation of spacecraft behaviour (e.g. anomalies, unexpected behaviour).

4.3 Atmosphere

Improvement of models of the Earth's upper atmosphere and its response to space weather drivers is a subject well covered and coordinated by international actions such as the COSPAR International Reference Atmosphere (Rees, 1986). The importance of drag effects has led to considerable effort being deployed in this area.

Because of the increasing number of planetary missions, models of planetary atmospheres also have to be developed to support design of aero-manoeuvring, entry, decent, landing and operations. Planetary global circulation models and climatic databases are being developed and will continue to be improved. In the case of Mars, dust storms are also an important part of that planet's "weather".

4.4 Ionosphere

Communications and navigation systems in orbit need to take account of ionospheric disruption to their signals. The developing European navigation programmes in particular – including the Galileosat constellation – are paying careful attention to the inclusion of ionospheric corrections in system designs and account is being taken of potential service degradation. As a result of the effects, considerable effort is being devoted to understanding and anticipating ionospheric effects, much of it sponsored by the military sector because of its strategic importance. Within Europe, the European Union sponsored "Cooperation on Science and Technology" (COST) action 271 on "Effects of the Upper Atmosphere on Terrestrial and Earth-Space Communications" (Zolesi and Cander, 2002) provides an important focus for research.

4.4.1 Microparticle Environments and Effects

While not normally considered part of space weather, microparticles (micrometeoroids and small-sized space debris) are nevertheless an important environmental hazard to spacecraft and crew, are variable and are affected by other space weather phenomena such as atmospheric density variations. However, apart from details of the populations and their fluxes, prediction of effects requires knowledge of the way penetration of spacecraft surfaces depends on particle properties – so-called damage equations. These are constantly in need of improvement. As far as knowledge of the environment is concerned, better in-orbit data, including analyses of retuned surfaces are needed (Drolshagen, 2001). Apart from the obvious hazard to manned missions through penetration of crew modules or space suits, many

elements of spacecraft and payloads are sensitive, including pressurise fuel and other tanks, solar arrays, and instrument baffles. One interesting issue recently discovered relates to X-ray astrophysics experiments, where the open nature of grazing incidence X-ray optics can allow propagation of microparticle fragments to the focal plane (Struder, 2001). This issue will be important to consider for future X-ray mission such as XEUS. Impact penetration of light shades and other important enclosures of scientific payloads also have to be carefully assessed. For this good flux models and risk analysis tools are needed, sometimes augmented by dedicated impact testing.

Analyses of surfaces returned from space are very important, as exemplified by the analyses of Hubble solar arrays. Return of structural elements from space is likely to decrease in the future as Shuttle missions become space station oriented and the Shuttle is replaced by a vehicle without return capabilities. Therefore in the longer term, it will be necessary to ensure that surfaces can be returned from space for analysis or that analyses of space hardware impact effects are performed in-situ.

4.4.2 Microsatellites and Facilities for In-Orbit Experimentation

Lack of rapid, frequent and cheap access to space for technology experiments is a serious impediment to technology progress, both for assessing space environmental effects on systems, and in other areas. Microsatellites and nanosatellites offer relatively low-cost access to space for technology flight experiments and are often ideal for experiments related to environmental effects and as platforms for monitoring the environment in important regimes. This has been recognised by those establishing the Living with a Star programme in the US, and while the programme has been broadened to become the International Living with a Star programme, the testbed elements are still essentially a US initiative. Expansion of activities of this type is needed. Cheap and frequent access to space would allow in-flight testing of components, emerging technologies, materials, solar cells, etc. It would allow the environment mitigation methodologies employed on the ground – usually combinations of model application and testing to simulate the environment – to be validated and improved.

4.5 Space Weather

What has been described so far has been a perspective of requirements in each of the space weather sub-domains or relevant technology domains. The space weather paradigm stresses the treatment of the solar-terrestrial system in a unified fashion. From the point of view of space weather *applications*, a

space weather service would ideally make use of a dedicated network of ground-based and space-based observations of solar-terrestrial conditions and, together with mature models, be able to make predictions or interpretations of the environmental conditions. Crucially, space weather applications services include the integration functions and major information technology elements such as data systems and communications. In this context, integrated space weather resources are highly valuable for application to space systems design and operation, as well as for other users.

4.5.1 Space Based Measurements

Studies performed in the past (Glover et al, 2003) have provided a detailed analysis of the parameters that need to be measured, the characteristics of these measurements (parameter ranges, accuracy, acceptable delay, etc.) and the locations where they should be made. Key requirements are:

- Monitoring of the solar surface to identify flares and mass ejection events and to establish inner boundary of interplanetary environment models;
- Monitor the solar wind plasma conditions as a warning for Earth-directed hazardous events;
- Monitor the environment close to the earth in various orbits to establish the plasma and energetic particle environments.

Further in the future, as the scientific understanding of processes close to the solar surface advances, it may be possible to identify important solar features before they give rise to flares or coronal mass ejections.

4.5.2 Ground Based Measurements

Ground based measurements are inexpensive compared to space and should be exploited wherever possible. Valuable ground-based measurements include:

- Geomagnetometry networks;
- Visible and Lyman-alpha solar imaging;
- Solar Magnetography;
- Cosmic Ray Neutron monitoring.

Emerging measurement techniques include:

- Muon telescope observations of cosmic ray modulation;

- Interplanetary scintillation observation of coronal mass ejections;
- High altitude and balloon-based solar observations.

These need to be exploited and monitoring facilities installed where gaps exist in global coverage, particularly in the European sector. For example, the extensive ground based monitoring network of the former Soviet Union is degrading and needs to be replaced in some way. Dedicated small-size solar telescopes (E.g. Lyman–alpha) are lacking in Europe.

4.5.3 Space Weather Economics

While the economic impact of "spectacular" space weather induced problems, such as the loss of a complete satellite or the collapse of a regional power grid, is easy to gauge, identifying the economic impact of non-spectacular space weather events is difficult. If a purely economic rationale is attempted, the costs of establishing a space weather service need to be weighed against the financial benefits that the products can bring. In particular one needs to address the question of whether having better data, predictions and tools for space weather can allow measures to be taken other that those taken at present which are generally characterised by "design for the worst case". In space systems this can, for example, translate into additional mass to protect systems and this has knock-on effects on cost. It is probably not possible to perform such cost-benefit analyses in all user domains while the user communities are still becoming educated to space weather hazards and what can be done about them. Another problem is that much of the data on the cost impacts of space weather effects are commercially confidential. It has also to be recognised that many of the benefits of an application-oriented space weather service are unquantifiable and relate to strategic advantages or the well being of citizens.

The strategy being perused by ESA is to initiate "pilot" service provision projects across the many user domains, and subsequently to investigate their values and costs (Glover et al, 2003). This type of activity, and benefits analyses in general, will continue to be necessary to justify any large-scale system developments. Nevertheless, in the space-effects field there is a clear appreciation of the need to continue to develop space weather resources for future missions.

5. CONCLUSIONS

Space missions are becoming more susceptible to space weather hazards and while on the one hand this necessitates work to understand more fully the phenomena, it is important that efforts are also made to establish the

capability to link in a quantitative way the environment with the resulting effects. This more engineering-oriented aspect is somewhat underplayed in much discussion of space weather. Nevertheless, those aspects are becoming increasingly difficult as the technologies affected and the effects "pathways" become more complex.

ACKNOWLEDGEMENTS

I am very grateful to all members of the ESA Space Environments and Effects section of for their considerable efforts and useful discussions, and particularly to Alain Hilgers and Alexi Glover for their efforts related to ESA space weather initiatives.

REFERENCES

Agostinelli S. and the Geant4 Collaboration, Geant4 – a simulation toolkit, Nuclear Instruments and Methods in Physics Research A 506, 250 –303, 2003.

Aran A., B. Sanahuja, D. Lario and V. Domingo, An operational code for solar energetic proton flux prediction. first approach, proceedings of the ESA Space Weather Workshop: Looking Towards a European Space Weather Programme, Noordwijk, 17-19 December 2001, ESA-WPP-194, 2003.

Daly E.J, ESA space weather activities, in *Space Storms and Space Weather Hazards*, edited by I. A. Daglis, Kluwer Academic Publishers, Dordrecht, 2001.

Boscher D., S. Bourdarie, A. Masclet, S. Barde, R. Friedel, Modeling the Radiation Belt Environment, proceedings of the ESA Space Weather Workshop: Looking Towards a European Space Weather Programme, Noordwijk, 17-19 December 2001, ESA-WPP-194, 2003.

Drolshagen G., Hypervelocity Impact Effects on Spacecraft, Proc. Meteoroids 2001 Conf, Kiruna, 6-10 August 2001, ESA SP-495, , pp. 533-541, 2001.

Glover A., A. Hilgers, E. Daly, R. Marsden, Tomorrow's Space Weather Forecast, ESA Bulletin, number 114, May 2003. See also www.esa.int/spaceweather/

Harboe-Sørensen R., An overview of radiation single event effects testing of advanced memory components and associated problems, ESCCON 2002 – Proceedings of the European Space Components Conference, 24-27 September 2002, Toulouse, France ESA SP-507 ed. Harris R.A., 2002.

Heyderickx D., ISO TC20/SC14/WG4 web-site; http://www.magnet.oma.be/iso/

Harris R.A. (ed), Proceedings of the 7th Spacecraft Charging Technology Conference, ESA-ESTEC, Noordwijk, The Netherlands, 23-27 April 2001, ESA-SP-476, 2001.

Katz I., J.J.Cassidy, M.J.Mandell, G.W.Schnuelle, P.G.Steen and J.C.Roche, The capabilities of the NASA charging analyzer program, in Spacecraft Charging Technology 1978, ed. R.C.Finke and C.P.Pike, NASA CP-2071/AFGL TR-79-0082, ADA045459, p.101, 1979.

Mohammadzadeh A., H. Evans, P. Nieminen, E. Daly ,P. Vuilleumier, P. Bühler, C. Eggel, W. Hajdas, N. Schlumpf, A. Zehnder, J. Schneider and R. Fear, The ESA standard radiation environment monitor programme: first results from PROBA-1 and INTEGRAL, IEEE Trans. Nucl. Sci. NS-50, 6, 2003.

Purvis, C.K., H.B. Garrett, A.C. Whittlesey and N.J. Stevens, Design guidelines for assessing and controlling spacecraft charging effects, NASA Technical Paper 2361, 1984.

Rees D. (ed.), CIRA 1986: part I thermospheric models, Advances in Space Research, Vol. 8, Nos. 5-6, 1988.

Roussel J.F., The SPIS we site http://spis.onecert.fr/spis/index.html, 2003.

Stamper R. and M. Hapgood, Space Environment Database, proceedings of the ESA Space Weather Workshop: Looking Towards a European Space Weather Programme, Noordwijk, 17-19 December 2001, ESA-WPP-194, 2003. See also SEDAT project description, http://www.wdc.rl.ac.uk/sedat/

Strüder L., B. Aschenbach, H. Bräuninger, G. Drolshagen, J. Englhauser, R. Hartmann, G. Hartner, P.Holl, J. Kemmer, N. Meidinger, M.Stübig, and J.Trümper, Evidence for micrometeoroid damage in the pn-CCD camera system aboard XMM-Newton, Astronomy & Astrophysics 375, L5-L8, 2001.

Zolesi B., and Lj.R. Cander, Effects of the Upper Atmosphere on Terrestrial and Earth Space Communications: The new COST271 action of the European scientific community, Advances in Space Research, Vol.29, No 6, pp 1017-1020, 2002. See also http://www.cost271.rl.ac.uk/

Chapter 6

Space Weather Effects on SOHO and its Space Weather Warning Capabilities

P. Brekke[1], M. Chaloupy[3], B .Fleck[1], S. V.Haugan[1], T. van Overbeek[2], H. Schweitzer[2]

1. *RSSD, European Space Agency, GSFC, Greenbelt,MD 20771, USA*
2. *SCI-P, European Space Agency, GSFC, Greenbelt,MD 20771, USA*
3. *Astrium, GSFC, Greenbelt,MD 20771, USA*

Abstract Since its launch on 2 December 1995, the Solar and Heliospheric Observatory(SOHO) has provided an unparalleled breadth and depth of information about the Sun, from its interior, through the hot and dynamic atmosphere, and out to the solar wind. In addition SOHO has several times demonstrated its leading role in the early-warning system for space weather. SOHO is in a halo orbit around L1 Lagrangian point where it views the Sun 24 hours a day. Thus, it is situated outside the Earth's protective magnetosphere which shields other satellites from high energy particles and the solar wind. We present a summary of the observed effects on the instruments and electronics on SOHO throughout the mission. In particular we focus on a number of large particle events during the recent years while the Sun was approaching maximum activity, and how they affected both the scientific data as well as hardware components.

Keywords Sun, space weather, space environment and effects.

1. THE SOHO SPACECRAFT

The SOHO mission is a major element of the International Solar Terrestrial Programme (ISTP), and, together with Cluster, forms the Solar Terrestrial Science Programme (STSP), the first cornerstone in ESA's long-term science programme 'Horizons 2000' (Domingo, Fleck, and Poland, 1995). ESA was responsible for the spacecraft's procurement, integration and testing. It was built in Europe by an industry team lead by Matra Marconi Space (now

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 109-122.

called Astrium). Weighing in at 1,850 kg, the SOHO spacecraft measures about 9.5 m across with its solar panels extended and is 4.3 m high. Figure 1 provides a schematic view of the SOHO spacecraft. NASA provided the launcher, launch services and ground-segment system and is responsible for in-flight operations. Mission operations are conducted from NASA/Goddard Space Flight Center (GSFC).

SOHO was launched by an Atlas II-AS from Cape Canaveral on 2 December 1995 and was inserted into its halo orbit around the L1 Lagrangian point on 14 February 1996, six weeks ahead of schedule. Commissioning of the spacecraft and the scientific payload was completed by the end of March 1996. The launch was so accurate and the orbital manoeuvres were so efficient that enough fuel remains on board to maintain the halo orbit for several decades, many times the lifetime originally foreseen. An extension of the SOHO mission for a period of five years beyond its nominal mission duration (2 years), i.e. until March 2003, was approved in 1997 by ESA's Science Programme Committee (SPC). A second extension of another four years, i.e. until March 2007, was granted by the SPC in 2002. This will allow SOHO to cover a complete 11-year solar cycle.

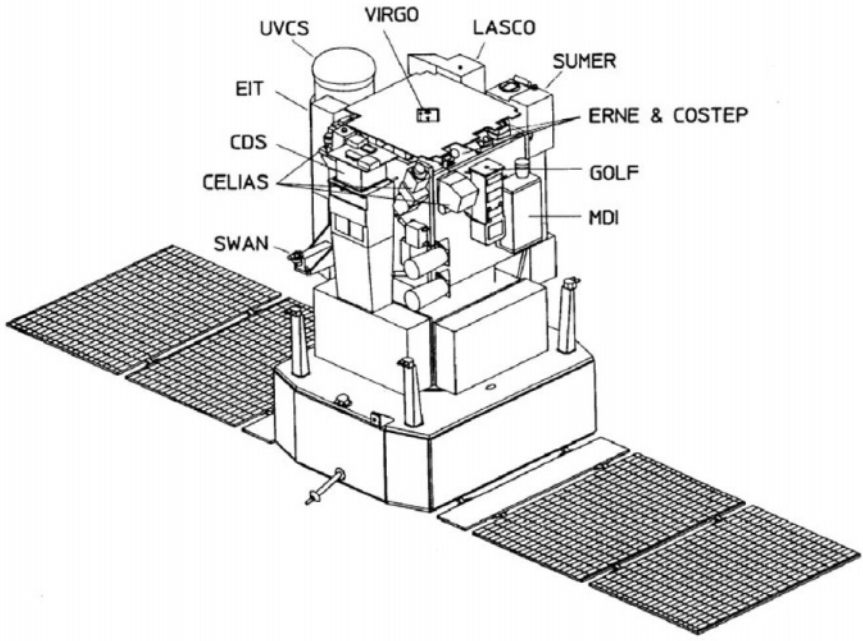

Figure 1. SOHO spacecraft schematic view.

SOHO has a unique mode of operations, with a "live" display of data on the scientists' workstations at the SOHO Experimenters' Operations Facility (EOF) at NASA/Goddard Space Flight Center, where the scientists can command their instruments in real-time, directly from their workstations.

SOHO enjoys a remarkable "market share" in the worldwide solar physics community: over 1500 papers in refereed journals and over 1500 papers in conference proceedings and other publications, representing the work of over 1500 scientists.

2. SOHO - A SPACE WEATHER WATCHDOG

Observations of the solar corona with the Large Angle Spectrometric Coronagraph (LASCO) and the Extreme ultraviolet Imaging Telescope (EIT) instruments on SOHO provide an unprecedented opportunity for continuous real-time monitoring of solar eruptions that affect space weather. LASCO takes images of the solar corona by blocking the light coming directly from the Sun itself with an occulter disk, creating an artificial eclipse within the instrument. LASCO best observes limb CMEs, but its extreme sensitivity even allows unprecedented detection of halo CMEs. EIT provides images of the solar atmosphere at four extreme ultraviolet wavelengths and reveals flares and other associated events in the atmosphere. EIT can usually determine whether CMEs seen by LASCO originated on the near or far side of the Sun, based on the presence or absence of corresponding events on the near side.

LASCO has been collecting an extensive database for establishing the best statistics ever on CMEs and their geomagnetic effects. By June 2003 more than 6000 CMEs have been recorded*. CME's are vast structures of plasma and magnetic fields that are expelled from the Sun. CMEs moving outward from the Sun along the Sun-Earth line can, in principle, be detected when they have expanded to a size that exceeds the diameter of the coronagraphs occulting disk. CME's directed toward or away from the Earth should appear as expanding halo-like brightenings surrounding the occulter. An example of a halo-CME is shown in Figure 2 as recorded by the LASCO C3 detector on 6 June 2000. Although halo CMEs were discovered by the SOLWIND coronagraph two solar cycles ago (Howard et al., 1982) the LASCO experiment is the first to observe a significant number of these events, thanks to its extended field of view and its improved sensitivity compared with earlier coronagraphs.

---

* A complete list of all detected CMEs with LASCO can be found at:
http://lasco-www.nrl.navy.mil/cmelist.html

St.Cyr et al. (2000) reported the properties of all the 841 CMEs observed by the LASCO C2 and C3 white-light coronagraphs from January 1996 through the SOHO mission interruption in June 1998 and compared those properties to previous observations by other instruments. The CME rate for solar minimum conditions was slightly higher than had been reported for previous solar cycles, but both the rate and the distribution of apparent locations of CMEs varied during this period as expected. While the pointing stability provided by the SOHO platform in its L1 orbit and the use of CCD detectors have resulted in superior brightness sensitivity for LASCO over earlier coronagraphs, they have not detected a significant population of fainter CMEs. The general shape of the distribution of apparent sizes for LASCO CMEs is similar to those of earlier reports, but the average and median apparent size of 72° (50°) is significantly larger.

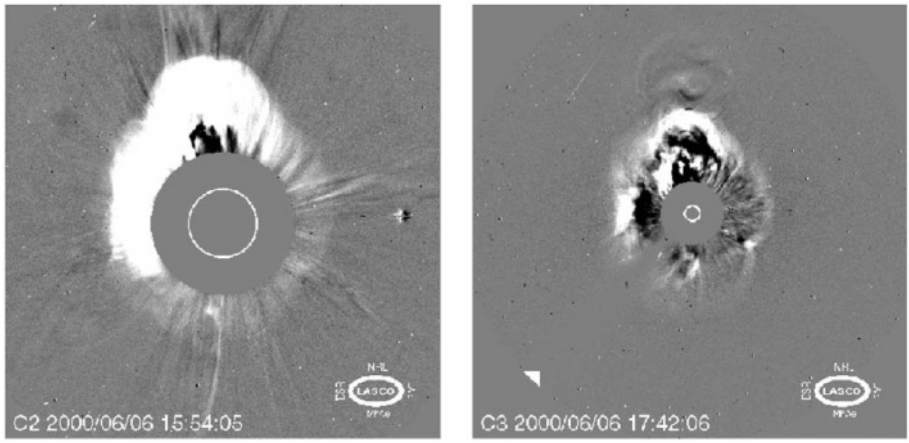

Figure 2. Example of a full halo CME observed by LASCO C2 (left panel) and C3 (right panel) coronagraphs. The field of view of the images are 2-6 and 3.5-30 solar radii.

St.Cyr et al.(2000) also reported on the statistics of halo CMEs. Using full disk EIT images they found that 40 out of 92 of these events might have been directed toward the Earth. A comparison of the timing of those events with the Kp geomagnetic storm index in the days following the CME yielded that 15 out of 21 (71%) of the Kp > 6 storms could be accounted for as SOHO LASCO/EIT frontside halo CMEs. An additional three Kp storms may have been missed during LASCO/EIT data gaps, bringing the possible association rate to 18 out of 21 (86%).

More recently, Zhang et al. (2002) found that 22 out of 27 (81%) major geomagnetic storms occurring between 1996-2000 were identified with front side halo (FSH) CMEs. Of these 16 (59%) were associated with unique FSH

CMEs while 6 (22%) where related to multiple FSH CMEs. They also find that while these geoeffective CMEs are either full-halo CMEs (67%) or partial-halo CMEs (30%), there is no preference for them to be fast CMEs or to be associated with major flares and erupting filaments. Again, this illustrates that SOHO has been providing new valuable information to better understand CMEs as well as being the only monitoring system for Earth directed CMEs until more ideal missions are launched (e.g. STEREO)

2.1 SOHO's role at the Space Environment Center

The Space Weather Operations Center at the Space Environment Center (SEC) in Boulder uses SOHO images daily. The forecast operations have become to rely on SOHO on a routine basis as a key input to solar observing and geomagnetic forecasting. LASCO is the only direct observation of coronal mass ejections. Prior to LASCO they had to rely on activity they knew to be well associated with CMEs, but none of these associations are 100% reliable. They use direction, size, and velocity information in LASCO images to help determine the arrival time and effectiveness of the disturbance.

EIT also plays an important role at SEC to pin down the source of any eruption. In addition EIT is a very good source for identifying erupting prominences and to identify coronal hole locations. Coronal holes have become an increasingly important part of the geomagnetic forecasting process. In fact at this point in the solar cycle coronal hole activity has become the predominant driver of geomagnetic activity.

Finally, forecasters use the MDI data on SOHO in order to track sunspot growth and decay and the magnetograms are used to track magnetic field strengths and complexity, a valuable input for flare forecasting.

2.2 Automated detection of CMEs

The visual detection of CMEs in the flood of incoming LASCO data is a labour-intensive task. Until today it is essentially the human eye that detects a CME occurrence and a scientist that collects all the CME parameters. An automated detection system called "Computer Aided CME Tracking (CACTus) has been developed for the LASCO images (Berghmans et al., 2002). The software detects bright ridges in [height, time] maps using the Hough transform and creates a list of events with principle angle, angular width and velocity estimation for each CME. In contrast to lists assembled by human operators, these CME detections by software can be faster and possibly also more objective. The first version CACTus has been evaluated

and it obtained a success rate of about 75%. This number is expected to improve in later versions. The software also detected some CMEs that were not reported in the official human created catalogs.

2.3 Solar wind shock spotter

The CELIAS/MTOF/PM instrument on SOHO measures the solar wind speed, density and temperature. A group at the University of Maryland recently implemented a "Shockspotter" program to identify interplanetary shocks in near-real time using proton monitor data. The program is based on semi-empirical algorithms using only solar wind proton data (since no magnetometer data is available on SOHO). Shock candidates are classified into 4 distinct zones, with confidence levels ranging from about 40% to 99%. Results have been used to study the frequency distribution of interplanetary shocks over the solar cycle.

The Shockspotter program is now part of the proton monitor real time data page at http://umtof.umd.edu/pm. The program can alert users (via eMail, upon request) whenever a shock front passes the SOHO spacecraft approximately 30-60 minutes prior to the arrival at Earth. A catalog of interplanetary shocks is also maintained at http:// umtof.umd.edu/pm/figs.html. The Maryland CELIAS group has also developed Web pages that show the solar energetic particle flux deduced from proton monitor background levels (http://umtof.umd.edu/pm/flare) and the solar soft X-ray flux from SEM measurements (http://umtof.umd.edu/sem/).

3. SPACE WEATHER EFFECTS ON SOHO

SOHO is designed to withstand the effects of the varying flux of high energy particles encountered in its L1 Halo orbit. These effects can be separated at least into three classes. First we will discuss the effects on the spacecraft (service module and solar panels) and then the effects on the different scientific instruments. A brief summary of efforts to prevent interruptions to the daily operation of the spacecraft is also discussed.

3.1 Effects on the spacecraft

3.1.1 Radiation Hazards:

During its lifetime the spacecraft components receive an integrated radiation dose that degrade their performance and can cause the following failures:

a) Solar Arrays

The only permanent effect so far is the degradation of the solar arrays due to high energetic protons from solar eruptions. This degradation is due to "displacement damage": energetic particles interact with the solar cell lattice producing defects which enhance electron and hole recombination thus reducing the solar cell's output voltage and current. The actual degradation of the solar array is given in Figure 3. The degradation due to proton events is evident with significant drops during the July 14, 2000 and November 4 & 23, 2001 events. The degradation after 80 months in space is 13.8%. This is an annual average degradation of 2.01%, well within the 4% per year requirement. SOHO can operate down to 70% sensitivity without taking any energy saving action.

b) Sensitivity of the Fine Pointing Sun Sensor

The Fine Pointing Sun Sensor (FPSS), together with the Star Sensor Unit (SSU), is part of SOHOs Attitude and Orbit Control System (AOCS).
Similar to the solar array, the FPSS sensitivity is gradually decreasing due to the impinging radiation over several years. The present performance is still sufficient, but in the long run, we might eventually require a new calibration of the output level of the FPSS. This is a simple on-board parameter change.

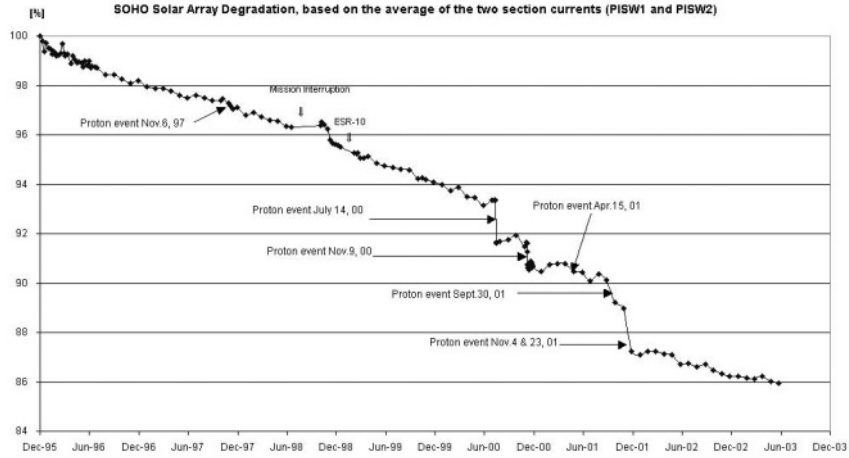

Figure 3. SOHO solar array degradation since the launch in 1995. The drop in sensitivity due to individual proton events is evident.

3.1.2. Radiation Induced Background

Radiation impinging on detectors or associated electronics can produce an increase of the background noise. The Star Sensor Unit consist of an optical system with thermal sensors for calibration of the focal length of the optics and a CCD detector (377 x 283 pixels), mounted on a Peltier cooler with thermal control for the CCD temperature (– 40°C) and for the electronics of the detector drivers and data pre-processing. The background noise of the Star Sensor Unit so far is very stable since the beginning of the mission

3.1.3. Single Event Upsets (SEUs)

Cosmic rays or heavy ion impact can provoke single event upsets, which may disrupt the operation of sensitive electronics.

a) Electronic units self switch-off

A fair number of self switch-off events occurred, which are attributed to Single Event Upsets (SEUs). Three of them caused transitions to the spacecraft safe mode (Emergency Sun Reacquisition – ESR), causing major disruptions of science operations. Five times the battery discharge regulators switched themselves off and there were about 7 occurrences, where instrument boxes were switched off or required rebooting. Many of the self switch-offs are probably caused by false triggering of internal

protection circuits, which are designed to protect against over-voltage or over-current. In all cases, no permanent damage occurred and the systems could be re-activated successfully.

b) Solid State Recorder

A major temporary radiation effect is the SEUs in the Solid State Recorder (SSR), resulting in bit flips in the memory. The EDAC (Error Detection And Correction) detects and corrects these single errors (in the same word). Double errors are detected but not corrected.

Single errors are very common for the SSR 2Gbit memory:

- at solar min: 1 SEUs/minute
- at solar max: 0.5 SEUs /minute
- during proton events: up to 76 SEUs /minute (July 14, 2000 event)

So far there has only been 1 double error since launch, which was corrected as soon as the effected memory location was overwritten with new data. A plot of the SEU's/ minute/2GB over the entire mission is given in Figure 4.

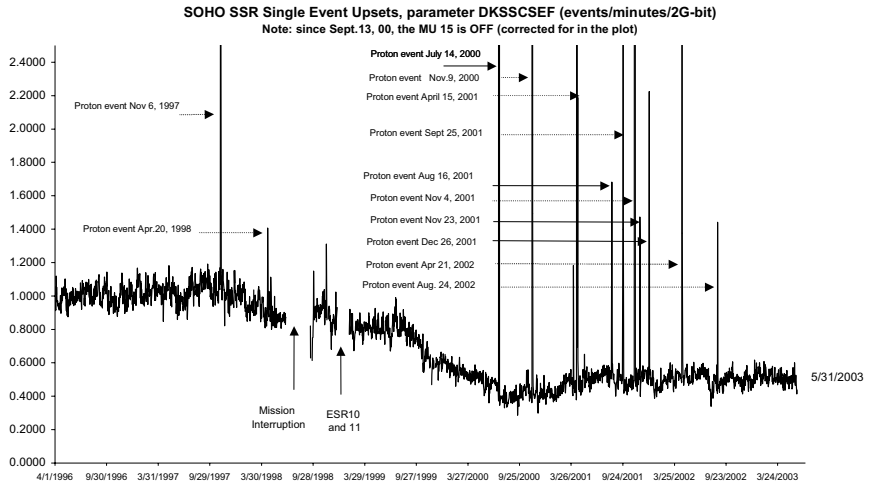

Figure 4. Number of SEU's per minute per 2GB over the entire mission. The solar cycle effect can be clealry seen with more SEU's during solar minimum (weaker solar magnetic field). Overlayed are the proton events during the SOHO mission.

c) Star Sensor Unit

Another temporary radiation effect is observed on Star Sensor Unit (SSU). When particles hit the CCD (Charge Coupled Device) of the SSU, they generate electrons, which charge up the pixels just like the regular photons, producing bright star-like signatures.
The SOHO star tracker tracks five stars in small tracking windows. If a particle hits the tracking window it can result in a wrong assessment of the tracked star's barycenter and/or magnitude.

The SSU interprets this as a movement of the star it was tracking thus providing wrong information to the attitude control software, resulting in turn in wrongful attitude correction orders to the wheels. Furthermore, the Star Tracker itself is moving its tracking window to the new wrong barycenter, and sometimes loses the true star in doing so.

The star tracker had lost the guide star 54 times the first 3 years.
Most of these resulted in loss of nominal attitude (fall back into Roll Maneuver Weels mode/gyro mode), with the consequence of reduced science during the special operation to recover to nominal configuration.

3.1.4. Improvements to Onboard Software

The onboard software was designed to be upgraded and a series of improvements have taken place throughout the mission. Two improvements to increase the robustness to SEUs have been implemented:

3.1.4.1. Star tracker and AOCS software changes:

The star tracker's internal software was modified:

- to filter out transient changes in the star barycenter (Position Jump Filter).
- to filter out transient changes in the star's magnitude

The result is that for both transient changes no false event report is sent to the AOCS software.

The AOCS overall task is to provide the spacecraft with the requisite pointing performance during the various spacecraft activities. The AOCS software was modified to delay the effect of false event reports of the SSU to the attitude control computer (Staircase Filter). This filter was first implemented as a patch, but was later included in the gyroless software.

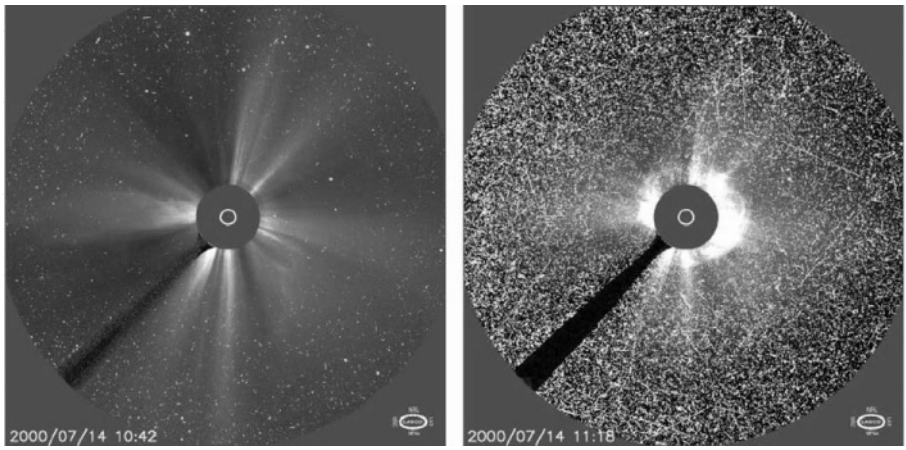

Figure 5. Images taken by the LASCO C3 coronagraph during the July 2000 solar energetic particle event showing severe effects on the detector from radiation background. Note that even though the images appear to be totally swamped during a proton storm, we are scaling the images to show the subtle coronal changes so that the particles are enhanced. They really don't saturate the detectors.

Thanks to this, SOHO was able to remain in normal mode during the Bastille Day (July 14, 2000; 3 star swaps) and the Nov 9, 2000 event (also 3 star swaps). There have been 11 star swaps in all, since October 1999.

Since the new gyroless software with the star swap feature was uploaded, we had not a single loss of nominal attitude (i.e. fall back into the new gyroless Coarse Roll Pointing mode). SEUs can still cause the stars to be flagged "invalid" for a while, but they have always returned to valid on their own. With these new upgrades SOHO is now extremely stable!

3.2 Effects on the Scientific Instruments and Operation

As with spacecraft electronics and detectors, several instruments are also subject to effects from energetic particle events. For some *in situ* sensor instruments, the particles are the main subject matter; for some, the particles are mostly a nuisance causing image degradation, but some instruments have health and safety concerns, due to e.g. high voltages on their detectors, the potential for arcing and permanent damage.

3.2.1 Image Degradation

As can be seen in Figures 5, 6 and 7, the image degradation experienced during energetic particle evens can be quite severe. Not only does the (relatively short-lived) degradation render images nearly useless for scientific analysis and space weather purposes – they also cause them to be

much less compressible by the on-board software, in the case of EIT and LASCO. With a limited amount of telemetry and on-board storage, this results in the instrument getting "backed up", with a shifting of scheduled observations to a later time. While not necessarily critical under regular circumstances, certain joint observing programs rely on a closely coordinated timeline between a number of instruments both on board SOHO and on other spacecraft, as well as ground based observations. This can only be corrected by intervention from the ground, using near-real-time (NRT) commanding to flush queues, skip observations, or upload new plans.

3.2.2 Health and Safety Effects

With several types of instruments operating detectors that have high voltage "image intensifiers" of different types, energetic particle showers are not purely an inconvenience. Although no incident has yet damaged any of the SOHO instruments, precautions have been put in place to ensure that the likelihood of damage is being kept as low as possible. Since, in general, the image-intensified instruments' data during particle events are not very useful anyhow, the loss of science data is not of concern; health and safety takes priority for those that feel a "better safe than sorry" approach is appropriate. The instruments that do take precautions of various kinds are: Coronal Diagnostic Spectrometer - CDS (continuous detector readouts to prevent charge build-up), Ultraviolet Coronagraph Spectrometer - UVCS and Solar Ultraviolet Measurements of Emitted Radiation - SUMER (high voltages turned down).

3.2.3 Operational Implications

The main operational "warning system" is the spacecraft solid state recorder. Since the SSR SEU counter is being monitored on the ground while the spacecraft is in contact (to prevent the SEU counter from overflowing), the impacts to normal operations are minimal when there is no particle event. If the SEU counter needs to be reset more often than once per hour, the Science Operations Coordinators are contacted, alerting instrument teams about the situation according to their own criteria.

For times when the spacecraft is not in contact, the warning system is based on NOAA GOES data from the web. Of course, with no spacecraft contact, nothing can be done about the instruments, so the status is only checked some time in advance of station passes with commanding ability. In addition, a 24/7 system based on automatic paging of the SOCs is in place, using NOAA GOES data from the web.

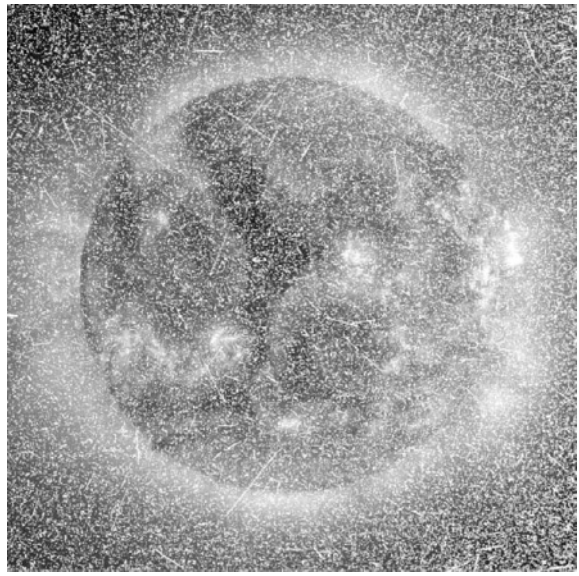

Figure 6. The Extreme Ultraviolet Imaging Telescope (EIT) observing during a proton event.

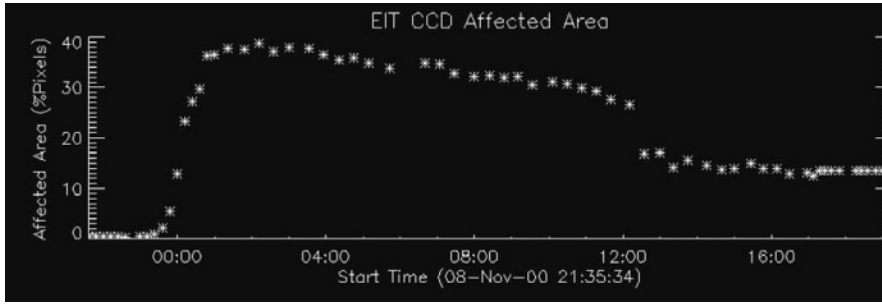

Figure 7. During a strong proton storm about 40% of the EIT detector is affected. The effect of the protons can be seen for a few.

3.2.4 Long-term Effects

No serious long-term adverse effects have yet been noted, although the high energy particle environment does contribute to the gradual degradation of instruments. In particular, contaminant "doping" of refractive optics changes the absorption coefficients (impacting the optics temperatures), and the indices of refraction (focus changes). In addition, parts of the gradual sensitivity losses experience by many instruments can be attributed to contamination of detector electronics.

4. MISSION STATUS AND FUTURE PLANS

Although long past the design lifetime of 2 years, SOHO is doing remarkably well. Fuel reserves of 123 kg should last 10 more years according to conservative estimates, and the solar array degradation is at only 10%, with a remaining margin of 25% before conservation measures must be applied. The gradual degradation of instruments and multi-layer insulation due to EUV exposure and high-energy particles is as expected, and not a cause of concern. Barring unexpected events, there seems to be no technical reason why SOHO and its instruments should not be able to complete observations of a full solar cycle. With several years until heirs to the throne can be expected, we hope that SOHO will continue its hegemony in the field of solar and heliospheric observations for years to come.

5. ACKNOWLEDGEMENTS

We would like to thank Christopher Balch (SEC) and Fred Ipavich (Univ. of Maryland) for very useful input and comments to this paper.

6. REFERENCES

Berghmans, D., Clette, F., Cugnon, P., Gabryl, J.-R., Hochedez, J.-F., Van der Linden, R., A.M., Verwichte, E., 2002, JASPT, 64, 757.

Domingo, V., Fleck, B., and Poland, A.I. 1995, Solar Physics, 162, 1.

Howard, R. A., Michels, D. J., Sheeley, N. R., Jr., and Koomen, M. J. ApJL, 263, 101.

St.Cyr, C., Howard, R. A., Sheeley, N. R. 2000, J. of Geophys. Res., 105, 185.

Zhang, J., Dere, K. P., Howard, R. A., Bothmer, V. 2002, ApJ, 582, 520.

Chapter 7

Prevention of Spacecraft Anomalies – The Role of Space Climate and Space Weather Models

Janet L. Barth
Flight Data Systems and Radiation Effects Branch, National Aeronautics and Space Administration, GSFC, Code 561, Greenbelt, MD, 20771, United States of America

Abstract Space-based systems are developing into critical infrastructure to support the quality of life on Earth. Mission requirements along with rapidly evolving technologies have outpaced efforts to accommodate detrimental space environment impacts on systems. This chapter describes approaches to accommodate space climate and space weather impacts on systems and notes areas where gaps in model development limit our ability to prevent spacecraft anomalies.

Keywords NASA, ESA, space weather, space climate, anomaly prevention

1. INTRODUCTION

The Sun emits time-varying magnetic fields, plasmas, and energetic particles. This solar variability drives changes in the interplanetary environment which then interacts with the Earth's magnetic field and outermost atmosphere to produce changes in the near-Earth space environment. The space environment and its solar-induced changes interact with spacecraft and instrument components and can cause anomalies resulting in loss of data, degradation of capability, service outages, and, in extreme cases, the loss of spacecraft. The most effective time to prevent spacecraft anomalies is during the pre-launch phases when risk can be minimized through technology selection and system design. For most missions, some level of "residual risk" must be assumed due to cost constraints, increasing complexity of space systems, unknowns in the space environment, and/or unknowns in space environment effects mechanisms. Possible consequences of the residual risk on spacecraft health and safety

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 123-145.

and on degradation of service must be evaluated and mitigated by writing operational guidelines for spacecraft operators and instructing the operators on how to use them effectively. The need for space weather models to manage residual risk during launch and operational phases is clear. However, space "climate"[1] models are equally important because of their crucial role in reducing risk in pre-launch phases of missions. In the case of both space climate and space weather models, model development lags behind the increase in the complexity of space systems and our dependence on space-based assets.

Even during the early 1960s, when space systems were very simple, spacecraft electronics were found to be unreliable in space environments. Problems from differential charging from the solar wind and from noisy data transmission to the Earth from soft fails were noted. These problems were largely dealt with by building redundancy into systems. However, the production of enhanced radiation levels from the explosion of nuclear devices at altitudes above 200 kilometers (Starfish and others (Barth et al., 2003)) and the ensuing problem of shortened spacecraft lifetimes emphasized the need for a uniform, quantitative description of the trapped particle environment. Later, as other effects induced by space environments were better understood, efforts to model the space environment resulted in models of all components of the environment (Barth et al., 2003, Lemaire, 2001).

Revolutionary changes have occurred in space-based systems since the development of the commonly used models of the space environment. First, humanity is increasingly reliant on space-based assets. In addition to the research functions that are performed in space in the areas of space science, earth science, human exploration of space, and aeronautics and transportation; critical services are also space-based, including navigation, telecommunications, defense, space environment monitoring, and terrestrial weather monitoring. Second, the performance demands of reconfigurable systems, constellations of small spacecraft, large deployable structures, imagers, and on-board computing increase the complexity of spacecraft and payloads and may require the use of rapidly evolving, complex technologies. Finally, space agencies and industry are developing missions that must operate in challenging space environments. For example, earth science missions that seek to understand complex global change processes require global coverage that cannot be achieved in Low Earth Orbits (LEOs).

---

[1] Dave Schwartz (htpp//www.Weather.com) defines weather as "the historical record and description of average daily and seasonal weather events that help describe a region. Statistics are usually drawn over several decades". This definition is easily adapted to space climate used by inserting "space" before weather.

However, placing spacecraft in the higher altitude regions of Medium Earth Orbits (MEOs) and geostationary orbits (GEO), exposes them to much higher radiation. Europe's global positioning satellite system and NASA's Living With A Star (LWS) Program also plan multiple spacecraft in high radiation regions of the magnetosphere.

Our increasing dependence on space-based systems demands that we increase their reliability, ideally achieving "all weather" space systems. This requires that we address the effects of space environment through design accommodations and operational countermeasures. However, most of the current space environment models are inadequate to effectively prevent anomalies, especially on technically complex systems in challenging environments.

2. THE CAUSES OF SPACECRAFT ANOMALIES

To understand where best to focus efforts to improve our ability to prevent spacecraft anomalies, it is useful to examine compilations of spacecraft anomalies and to understand the space environment effects that cause them. The components of the space environment that can pose hazards to normal spacecraft operations include micrometeoroid and orbital debris which cause impact damage and increased contamination; the neutral thermosphere which causes surface erosion due to atomic oxygen, satellite drag, and spacecraft glow; hot plasmas which induce charge on surface of spacecraft; relativistic electrons which cause deep dielectric charging; and particle radiation environments which cause surface material degradation (in synergy with atomic oxygen and ultraviolet radiation), microelectronics and sensor degradation, and single event effects. A review of the radiation environments can be seen in Panasyuk (2001).

Mazur (Mazur, 2002) presented the results of an Aerospace Corporation study (Koons, 1999) that analyzed the causes of spacecraft anomalies. Figure 1, from that study, shows the number of anomalies as a function of the space environment effects that caused them. "ESD" is damage from electrostatic discharges (spacecraft surface charging and deep dielectric charging), "SEU" is single event upsets or bit-flips, "Radiation Damage" is total ionizing dose and non-ionizing dose, and "Other" represents other and unknown causes. Below is a description of these common causes of anomalies.

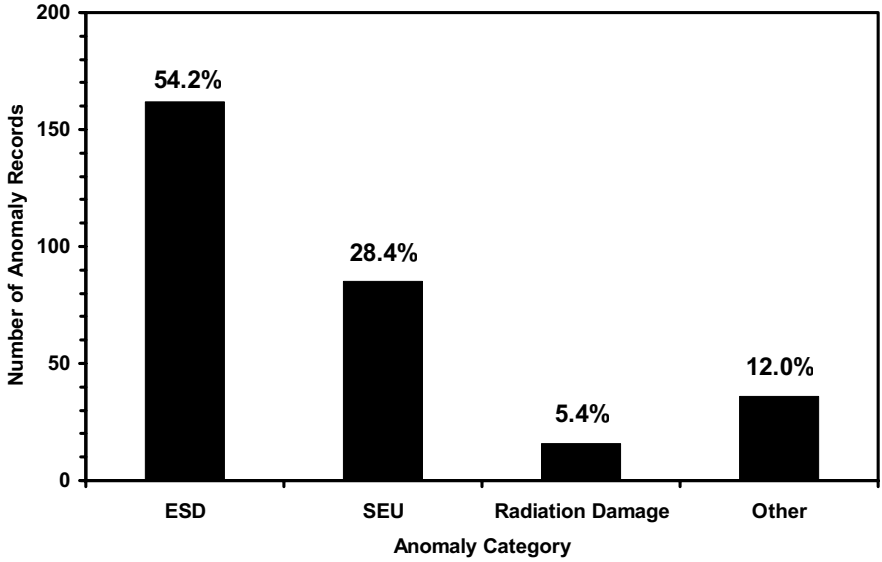

Figure 1. Spacecraft anomalies as a function of the space environment effect, where ESD is electrostatic discharge, SEU is single event upsets, Radiation Damage is total ionizing or non-ionizing dose, and Other represents other causes or unknown sources, from Koons et al., Aerospace Technical Report, 1999.

2.1 Spacecraft Charging

Spacecraft surface charging and deep dielectric charging result in discharges that can cause background interference on instruments and detectors, biasing of instrument readings, physical damage to materials, upsets and physical damage to electronics, increased current collection, reattraction of contaminants, and ion sputtering which leads to acceleration of erosion of materials. Plasmas are responsible for surface charging; particularly in planetary radiation belts where storm induced fluctuations occur. Deep dielectric charging results from higher energy electrons penetrating and collecting in non-conducting materials until the material's dielectric breakdown is reached and a discharge occurs. As with plasmas, storm induced increases in high-energy electron levels are known to increase the risk of deep dielectric charging problems. For an overview of spacecraft charging and induced anomalies, the reader is referred to Baker (Baker 2001).

2.2 Single Event Upsets and Single Event Effects

Single event effects (SEEs) occur as a result of charge being generated along the path of a primary or secondary ionizing particle, collected on circuit nodes, and disrupting normal circuit operation. Both the total collected charge and the rate of charge collection can be important in triggering the effect. SEEs affect memories, power devices, control logic devices, etc. Although increased levels of protons and heavy ions from solar particle events can increase the level of SEEs on systems, daily exposure to background levels of protons and ions in interplanetary space and in planetary radiation belts is a significant source of SEEs.

Single event upsets (SEUs) in memories is the most common and best known SEE, however, other effects on newer technologies can be more disruptive to spacecraft operations. Less known non-destructive effects are single event transients (SET), single event functional interrupts (SEFI), and multiple bit upsets (MBU). MBUs can result in uncorrectable errors in data systems. SEFIs occur in high-density memories when control regions of a device are hit by a particle possibly resulting in system lockup or reset. SETs are a well known problem in some detector technologies generally limiting their use to low radiation regions, however, SETs also cause voltage dropouts in logic devices which can result in system resets. For example, the increase in the heavy ion population during the November 2001 solar particle event caused an SET on a linear bipolar device on NASA's Microwave Anisotropy Probe (MAP). As a result, MAP's processor was reset and the spacecraft went into a safehold condition.

SEEs can also be destructive resulting in permanent loss of the functionality of a component. Single event latchup (SEL), single event gate rupture (SEGR), and single event burnout (SEB) are examples of permanent failures from single particle strikes and can cause the loss of a system or a spacecraft.

SEEs must be mitigated through component selection or system design to avoid temporary or permanent loss of spacecraft service. The SEE vulnerability of newer technologies used in spacecraft and instrument systems is increasing because it is difficult to use processing techniques to make devices immune to SEEs. "Hardness by design" is gaining in use to mitigate SEEs, however, the penalty in required overhead is severe. Regardless of the technique used to mitigate SEEs in designs, the overhead required in the system is increased by inaccuracy in space climate models, and operational countermeasures are compromised by inaccurate space weather forecasts.

2.3 Radiation Damage

Cumulative radiation damage is caused by two mechanisms, total ionizing dose (TID) and total non-ionizing dose, otherwise known as displacement damage dose (DDD). TID degrades the performance of surface materials, such as, lens coatings and thermal control materials, and of electronics. It is possible to avoid TID effects through the selection of radiation-hardened components. Unfortunately, these components are increasingly unavailable because the space market share for microelectronics is less than 0.5% of the total market share (down from 40%). Also, many radiation-hardened components do not meet mission requirements because they are based on older generation technologies. It has become common practice to use commercial off the shelf (COTS) devices; however, their radiation response can be difficult to characterize due to large variation of radiation response within a device lot and the difficulty of testing imposed by packaging and hybridization. Large safety margins are used to accommodate the uncertainty, which when combined with inaccurate space climate models, often results in "over designing" systems. Electrons and protons in interplanetary space and trapped in planetary radiation belts cause TID. Because TID affects components from the surface to deep inside a spacecraft or instrument, particles across broad energy range (eVs through MeVs) are a concern.

DDD degrades the performance of solar cells, detectors (e.g., charge coupled devices), optocouplers, and optical lenses. It is more difficult to harden against DDD, therefore, the use of shielding and planning for "graceful" degradation is used to mitigate its effects. As with TID, particles in a broad energy range affect systems. When using heavy shielding to protect detectors, inaccuracies in the estimates of the levels of high-energy particles (>100 MeV) result in large error bars on damage estimates.

2.4 Other

Other causes of anomalies could include damage from micrometeoroid and orbital debris or degradation of materials from combined surface effects or operator error. Often the causes of anomalies cannot be determined due to the lack of information on either the space environment at the time and location of the anomaly or the specific effect or the system in which it occurred.

3. PREVENTION OF ANOMALIES

Reports of spacecraft anomalies in the space weather community focus on the space weather phenomena that cause them but they rarely discuss the mechanism of failures or the "lessons learned" that can be applied to design methodologies or to operational countermeasures to prevent anomalies in future systems. The causes of spacecraft anomalies given in Figure 1 are effects that can be minimized in pre-launch phases of missions by defining the expected space environment over the lifetime of the mission, understanding the effects that it has on the components used on the spacecraft, and defining specific environment accommodations.

The accommodation of space environment effects is a complex process that involves both physics and engineering disciplines. To ensure mission success, engineers, scientists, and program mangers rely on engineering judgment as guided by analysis of component response to the space environments. The success of such analysis depends on several factors. Accurate climate models of the space environment that represent variations for all conditions of the solar cycle are crucial for evaluating the extent to which environment threats may compromise mission goals. Measurements of component responses to laboratory simulations of the space environment provide critical data for bounding on-orbit device performance. Equally important, however, is a detailed model of the interaction and transport of environment sources through observatory models and device structures. Such models not only serve as a bridge for understanding laboratory data to prediction of on-orbit performance, they also provide guidance as to the test methods and laboratory measurements needed for such predictions. Of necessity, these models make simplifying assumptions that must be reexamined as mature technologies evolve, as new technologies are introduced, and as advances in desktop computing make more powerful and realistic analyses feasible.

Figure 2 shows the sources of uncertainties in simulating the space environment and effects on spacecraft and instrument components. The uncertainties translate directly into design margins that must be applied to estimates of space environment hazards to minimize the risk of mission failure. While the focus of this chapter is on requirements for space environment models, it is important to point out that it is not the only source of uncertainty in simulation of space environment and effects. Concurrent efforts are underway to develop tools such as GEANT4 (Allison et al., 2000; Truscott et al., 2000), MCNPX (Walters, 2003), workbenches, and integrated mission design centers.

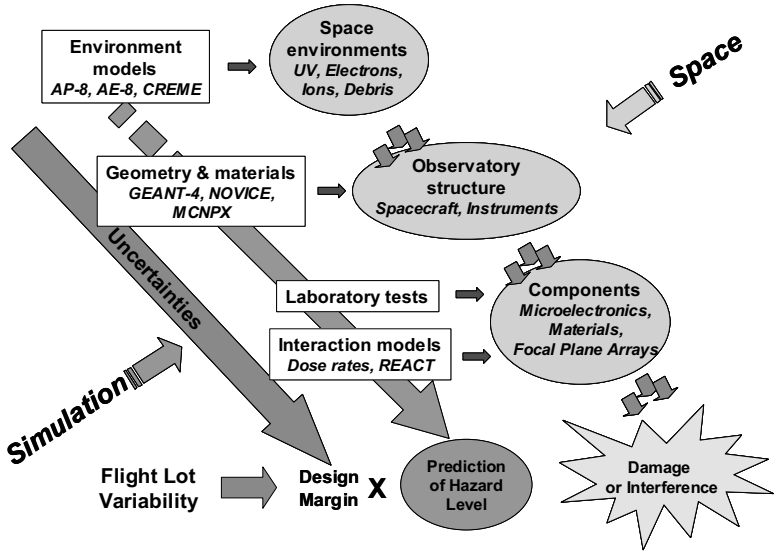

Figure 2. Sources of uncertainties in simulation of the space radiation environment. Simulation is required to predict performance in space. Errors in simulations result in under and over predictions of the hazard level.

The most effective way to assure spacecraft reliability is to use preventative measures throughout the mission life cycle. The challenge is to maintain the balance between meeting mission requirements, cost, and reliability. Residual risk for a mission is assumed when it is recognized that 100% reliability is not possible due to cost constraints and mission requirements drivers. The level of this residual risk must be assessed before launch and operations so that a risk management plan can be implemented early in the program. Where possible, spacecraft vulnerabilities should be identified so that operators can be trained on effective use of space weather forecast models.

The phases of mission development that must take into account space environment effects can be divided into concept, planning, design, launch, operations, and anomaly resolution. Figure 3 clearly shows the important role that space climate and space weather models play in minimizing risk for space-based systems.

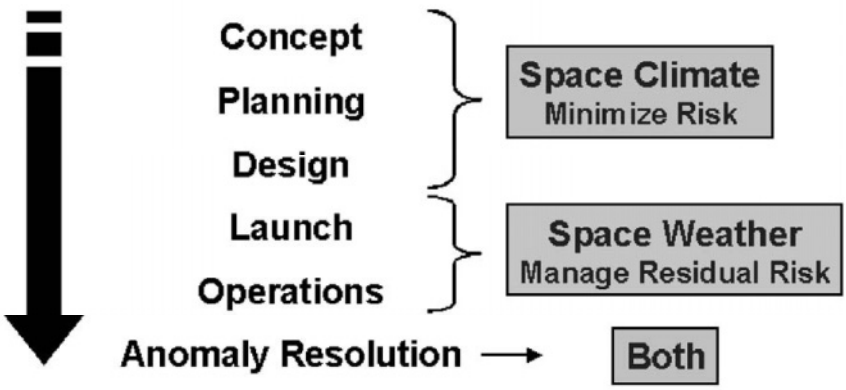

Figure 3. Phases of mission development which require space environment models. Space climate models are critical for minimizing risk of mission failure.

3.1 Space Environment Information for Pre-launch Phases

Issues that are addressed during the mission concept phase include observation requirements, observation vantage points, and development and validation of primary technologies. Required capabilities at this stage are integrated mission design tools, which include space climate models that can simulate the space environment throughout the solar cycle. Spatial resolution is also required so that trades between vantage points can be considered. Worst-case space environments are also needed to assess the survivability and function of the primary technologies.

Issues that are addressed during the mission-planning phase are observation requirements, mission success criteria, architecture trade studies, and risk acceptance criteria. Most accommodations for space environment effects are implemented during mission design including component selection and testing, subsystem design, shielding requirements, grounding, error detection and correction, and estimates of observation loss. Time distributions of levels of activity are needed to estimate lost observation time from instrument interference and data corruption. Worst-case levels of the space environment are also required for determining the survivability of components and the level of required error mitigation. To guide decisions on the acceptable level of risk, confidence levels for the space climate models are required and the capability of forecasting models for specific environments of concern should be assessed. One of the most critical features of the space climate models is that they cover an energy range that

is adequate for addressing degradation or interference from the surface (e.g., thermal control materials) to heavily shielded systems (e.g., detectors).

3.2 Launch and Operations

Good engineering practice is not a guarantee of a spacecraft that is 100% free from vulnerabilities from the space environment. As discussed above, this is due to cost constraints, increasing complexity of space systems and technologies, unknowns in the space environment, and unknowns in space environment effects mechanisms. As a result, spacecraft are often vulnerable to increases in space environment levels, i.e., space storms. Therefore, launch and operation phases require models that can forecast space storms to protect the space-based asset by shutting down systems or avoiding risky operations, such as, maneuvers, system reconfiguration, data download, or re-entry. The need to forecast quiet times is as important as forecasting storms to give operators "windows" during which these risky operations can be performed. Spacecraft operation facilities find it useful to be able to schedule extra personnel when space storms are expected. Forecasts must be specific to the region, the particle population, and even the energy range.

3.3 Anomaly Resolution

Regardless of the service provided by a space-based system, it is critical to be able to restore the system to normal operations quickly after an anomaly. Often this is accomplished before resolution of the anomaly. However, as soon as possible, the anomaly must be resolved in order to prevent possible permanent damage to the system. Once the anomaly is resolved, the risk is reevaluated, and operational countermeasures AND design guidelines are updated. It is not unusual for anomalies to be unresolved. Health and safety monitoring on the spacecraft may be inadequate to pinpoint the system component that was sensitive to the in the space environment hazard. Frequently the space environment hazard is inadequately defined in terms of spatial resolution or energy and particle resolution. Science spacecraft often have data that are valuable for anomaly resolution; however, timely access to that data is generally an impediment.

The third type of model for "nowcasting" the environment is used to resolve anomalies so risk can be reassessed for both the operating systems and for other systems that are in development. As with forecasts, nowcasts must be specific to the region, the particle population, and the energy range. If anomaly resolution is critical to the mission and must be performed in near

real-time, monitors that are in close proximity to the system may be required.

4. MODEL DEVELOPMENT ACTIVITIES

Before discussing the status of model development, a review of organizations supporting model development is useful. The discussion is not intended to be inclusive of all researchers, but instead focuses on major agency support.

The European Space Agency (ESA) recognized the need to define priorities for new space climate models in the early 1990s and initiated a series of studies to improve the models of the radiation belts. The goals of the Trapped Radiation ENvironment Development (TREND) studies were to first analyze existing models for shortfalls and to later develop new models of the radiation belts (Heynderickx, 2003). Using data from SAMPEX, UARS, and CRRES, the TREND studies have resulted in some improvements in the radiation belt models which will be listed in the next section. Later, the United States (U.S.) National Aeronautics and Space Agency (NASA) started the Space Environment and Effects (SEE) Program (Kauffman, 2003) to develop space climate models, environment interaction models, and databases to be used for spacecraft design. The SEE program has sponsored space climate modeling development efforts for solar protons and heavy ions and trapped protons.

The need for space weather forecasting capability was outlined in the U.S. National Space Weather Program for a broad user base (NSWP, 2000). The original study identified the need for space weather forecasting for spacecraft operations, and in 2003, it was recognized that improved space climate models are also required to reduce the risk of on-orbit failures through design accommodations. Since 1965 the U.S. National Oceanic and Atmospheric Administration's (NOAA) Space Environment center (SEC) has been the official U. S. course of space weather alerts, warnings, and forecasts. Activities include monitoring, data management, providing space environment information, research and research transition to operations, including models, and education. The origin of the international component can be traced to the early 1910s. Currently, the International Space Environment Service (ISES) includes several world wide regional warning centers with NOAA acting as a hub ("World Warning Agency"). An overview of NOAA's role and activities is given in Joselyn (Joselyn, 2001).

Recently, the European Space Agency conducted two parallel space weather feasibility studies to assess the requirements for space weather service in Europe. ESA has now begun a space weather applications pilot

project to expand the results of the studies and to develop the European space weather community (Daly, 2001, Daly, 2003).

In 2001 NASA's Living With a Star (LWS) Program (LWS, 2003) was initiated with the goal to develop the scientific understanding to address the aspects of the connected Sun-Earth system that affect life and society. One target area is to improve knowledge of space environments for spacecraft applications. Space missions are being developed which will help to fill the need for environment measurements, and a Targeted Research and Technology Program has been defined to address the need for improved modeling capability for both space climate and space weather.

In 2002 the international science community established the International Living With a Star Program (ILWS). The mission of the program is to stimulate, strengthen, and coordinate space research to understand the governing processes of the connected Sun-Earth System as an integrated entity. The objectives are to stimulate and facilitate 1) study of the Sun-Earth connected system and the effects which influence life and society, 2) collaboration among potential partners in solar-terrestrial space missions, 3) synergistic coordination of international research in solar-terrestrial studies, including all relevant data sources as well as theory and modeling, and 4) effective and user driven access to all data, results, and value-added products (Sibeck, 2003).

5. SPACE CLIMATE PREDICTION

This section will describe the current capabilities of the space climate models commonly used in the pre-launch phases of mission development. Areas where models improvements are required will be highlighted.

5.1 Galactic Cosmic Rays

The need for understanding the galactic cosmic ray (GCR) environment for astronaut protection was identified early in space programs resulting in a program to measure (IMP-8 spacecraft) and model the variations in GCR levels. Later when GCRs were identified as the cause of SEEs on spacecraft, the microelectronics community benefited from this work. The model most commonly used for mission planning and spacecraft design is embedded in the CREME96[2] (Tylka et al., 1997) workbench tool that calculates SEU rates in devices. This GCR model predicts energy-flux spectra for all of the ions

[2] Funded by the U.S. Naval Research Laboratory and by the NASA Space Environments and Effects Program

from Hydrogen through Uranium for energies from 1 to 10,000 MeV/n. The energy-flux spectra are converted to linear energy transfer (LET) spectra which is a crucial metric to understand the level of the space environment hazard to microelectronics. The current GCR models, including the CREME96, are acceptable because they are estimated to predict the GCR levels over the solar cycle to within ±15-25%.

5.2 Solar Heavy Ions

The CREME96 tool also contains a model of the solar heavy ion environment based on analysis Dietrich's analysis of the solar heavy ion data from the University of Chicago instrument on the IMP-8 satellite (Tylka et al., 1997). The dataset was especially important for modeling the energy-flux spectra at higher energies. The analysis of 100 solar heavy ion events in Dietrich's database showed that the October 1989 solar particle event could be used to represent an upper bound on the maximum solar heavy ion environment. The CREME96 model gives "worst-case" flux-energy spectra for H through U ions based on that event. Later Dyer et al. (Dyer et al., 2000, Dyer et al., 2002) presented data on the LET spectra of solar particle events occurring between 1998 and 2001. While there are very low enhancements of high LET fluxes for many of the events, three events approach or equal the CREME96 worst-day model. At low LET, where protons dominate and usually lead to single event effects by nuclear interactions, two events slightly exceed the model.

The CREME96 solar heavy ion model represents a large improvement over previous models but, unlike the GCR models, fails to meet the requirements of pre-launch phases of missions. It is not always practical or possible to design microelectronics systems that are 100% free from destructive and non-destructive SEEs in worst-case solar heavy ion environments. Increasingly, mission planners require space climate models that are based on confidence levels which guide risk acceptance decisions.

In recognition of the need to understand the statistical variation of the solar heavy ion event intensities, the NASA Space Environments and Effects program funded Xapsos et al. (Kauffman, 2003) to study solar heavy ion datasets and to develop a statistical model. The Xapsos model of the solar heavy ion environment will provide a statistically based upper limit on the event fluxes for systems that must operate through events and will allow assessment of the levels of risk for other systems by providing a distribution of the flux levels as a function of confidence levels. The major obstacle to modeling solar heavy ions is the lack of measurements. The infrequency of the events requires several solar cycles of data to get a database large enough to do statistical analyses.

5.3 Solar Protons

King (King, 1974) published the first statistical model for solar proton events using Poisson distributions. He concluded from his analysis of proton data from the 20th solar cycle that solar proton events could be classified into “ordinary” and “anomalously large”. This was based on the fact that only one anomalously large event occurred in the 20th solar cycle - the August 1972 event. That event alone accounted for 84% of the total proton fluence in the solar cycle at energies E > 30 MeV. However, when Feynman et al. (Feynman et al., 1993) added cycle 19 and 21 data to the solar proton event database, they were able to conclude that individual solar proton events actually form a continuum of event severity from the smallest to the largest, blurring the distinction between ordinary and anomalously large events. Their work resulted the JPL model for solar proton events (Feynman et al., 1993).

Many large events similar to the August 1972 event occurred in cycle 22 increasing concern about the validity of the solar proton models. With the goal of improving the ability to address practical aspects of spacecraft reliability, a team led by Xapsos began compiling solar proton data for solar cycles 20, 21 and 22 and using statistical techniques to derive probability distributions of cumulative solar proton fluences. Xapsos et al. (Xapsos et al., 1999) applied extreme value theory to determine probability of encountering a single large event over the course of a mission. They also used compound Poisson process theory to describe the probability of encountering various fluence levels during a mission. The work of the Xapsos team confirmed the Feynman conclusion that a “typical event” cannot be defined.

The Xapsos team have also worked on models that define the peaks of solar proton events for E > 10 MeV. To accomplish this, they applied Maximum Entropy Principle (MEP) to select the least biased event probability distribution. The MEP, used for earthquake predictions, is valuable for analyzing incomplete datasets. They validated the results with Lunar Rock Records dating back to ancient times. The Xapsos team continued their work by establishing worst-case solar proton spectra for solar events (Xapsos, 1999). When comparing their model with the CREME96 (Tylka et al., 1997) solar proton model which was based on the October 1989 solar particle event, they found that, statistically, the CREME96 model is closer to a 90% worst-case event model. Xapsos et al. have combined the model elements into the Emission of Solar Protons (ESP) model (Xapsos et al., 2003).

The JPL and ESP models are commonly used to predict solar proton levels for total ionizing dose, displacement damage, and single event effects.

Both models would benefit from a larger database of solar proton event measurements and from measurements of protons at energies > 100 MeV. Large uncertainties in the solar proton environment at high energies translate into large design margins for heavily shielded applications, especially, instrument detectors operating outside of the radiation belts where solar proton induced damage dominates.

5.4 Trapped Particles

The trapped particle models that are most often used at this time are the AP-8 (Sawyer and Vette, 1976) for protons and the AE-8 (Vette, 1991) for electrons. The AP-8 model, released in 1976, was the culmination of a long-term effort to include all of the previous models under one common approach and to include all of the data after 1970. After 1977, the modeling budget was significantly reduced so a similar effort to consolidate the electron models into the AE-8 model was not completed until 1983. The formal documentation of that model was released in 1991. The AP-8 and AE-8 models include data from 43 satellites, 55 sets of data from principal investigator instruments, and 1,630 channel-months of data.

By the 1970s, scientific interest had shifted from trapped particles to the plasma regime to determine the physical mechanisms of particle energization and transport. As a result, the number of new datasets available for trapped radiation environment modeling was drastically reduced. It was not until the measurement of storm belts by the CRRES mission in 1991 that concerns were renewed about the ability to model the trapped radiation belts to sufficient accuracy for using modern microelectronics in space. Analysis of the CRRES instrument and experiment data showed that not only is the environment extremely dynamic but also that electronic parts respond to the short-term changes. The AP-8 and AE-8 radiation belt models, with their 4-6 year averages, are not adequate for application to design mitigation problems related to effects on shorter time scales. Also, the frequency of occurrence of the atypical events that could form storm belts is unknown, therefore, applying AP-8 and AE-8 or like models to setting design and operational rad-hard requirements creates uncertainties that are impossible to quantify.

The U.S. Air Force Research Laboratory (AFRL) CRRES Program, the ESA TREND Program, and the NASA SEE Program have sponsored efforts to improve in the radiation belt models including:

ESA TREND Program - an alternate interpolation method for AP-8 and an additional L increment at the low L values to give better resolution at steep gradients (Daly et al., 1996)

AFRL - the CRRESPRO model (Gussenhoven et al., 1993) to give estimates of the trapped proton levels before and after the March 1991, simulating quiet and active conditions in the magnetosphere

AFRL – the CRRESELE model (Brautigam et al., 1992) to give estimates for eight conditions of magnetic activity, six ranges of activity as determined by the AP15 magnetic index and for the average and worst case conditions that CRRES measured

AFRL – the CRRESRAD model (Gussenhoven et al., 1992) to give estimates of dose based on the CRRES 4-domed dosimeter for quiet and active conditions

AFRL – the APEXRAD model (Gussenhoven et al., 1997) from a CRRES-like dosimeter on the APEX spacecraft to extend the CRRESRAD model to low altitudes and high latitudes.

NASA SEE Program – the LOWTRAP model (Huston and Pfitzer, 1998) to predict proton fluxes below 850 km as a function of the solar radio flux proxy for atmospheric heating

ESA TREND Program – the SAMPEX/PET model (Heynderickx et al., 1999) to predict proton flux levels as a function of solar activity effects

ESA TREND Program – A_p15 dependent models of the outer electron belt to understand storm-time behavior of trapped electrons in the outer part of the radiation belts using Salammbô (Bourdaire et al., 1995) and data from Meteosat-3/SEM-2, CRRES/MEA, and STRV-1b/REM (Heynderickx, 2003)

ESA TREND Program – proton flux anisotropy in the altitude range of the MIR and ISS stations including secular, solar-cycle, and seasonal variations (Heynderickx, 2003)

6. SPACE WEATHER FORECASTING AND NOWCASTING

Regardless of the space environment in question, most space weather forecasting models do not provide adequate information or are accurate enough to be of practical use for operators of space systems. Operational actions cannot be taken every time a forecast of increased solar activity is issued. To be effective tools, warnings need to have spatial and spectral resolution and provide information about the level of severity. They also need to be specific about the ion composition of particle events. For space assets operating in the Earth's magnetosphere, current forecasting is particularly ineffective. Forecasting capability focuses on storms, however, it is equally important to know when the space environment will be "quiet" so

that critical operations, such as, reprogramming, maneuvers, or re-entry, can be performed.

Anomaly resolution and subsequent modifications to design guidelines and operational countermeasures are not effective without identifying the specific cause of the anomaly. This requires local knowledge of all relevant environments and information on the expected technology response. In some cases, the current monitoring capability provided by the NOAA GOES and TIROS spacecraft have been very useful to resolve anomalies, particularly for spacecraft outside of the magnetosphere, in geostationary, or in orbits similar to TIROS. Comments on capabilities for specific environment components are given in the sections below.

6.1 Galactic Cosmic Rays

The variations in the GCR levels occur slowly in comparison with the other space environment populations, and CREME96 model predicts the levels adequately for spacecraft design and operations needs. Therefore, there is no need for a GCR forecasting capability.

6.2 Solar Heavy Ions

Solar heavy ions pose a significant threat to spacecraft systems through their ability to cause SEEs on spacecraft microelectronics which can result in loss of spacecraft service. At this time there is little capability to monitor or forecast solar heavy ions. Science instruments on ACE and WIND make heavy ion measurements, however, because the data are not available within a reasonable time, this does not comprise a monitoring capability that can be used for forecasting or nowcasting (anomaly resolution). For warnings or anomaly resolution for microelectronics, direct measurement of the linear energy transfer (LET) profile of the event is crucial for understanding if systems are vulnerable to the event. The MAP anomaly caused by an SET on a linear bipolar device was resolved because the spacecraft is outside of the magnetosphere at L2, and LET data were available from the CREDO environment monitor (Dyer et al., 2002) on a U.S. Naval Research Laboratory space environment testbed, the Microelectronics and Photonics Testbed (MPTB).

6.3 Solar Protons

The NOAA series of GOES spacecraft carry proton monitors that have been very useful for nowcasting solar proton levels. It has been shown that, for

levels of nominal spacecraft shielding, increases in SEUs are correlated to > 100 MeV solar protons (Poivey et al., 2003). Therefore, the addition of the > 100 MeV channel to the NOAA Space Environment Center space environment nowcasts has been particularly useful for analysis of solar proton effects on spacecraft microelectronics.

Forecasts of increases of solar proton levels are still not accurate enough to be of practical use for spacecraft operations. Although it is generally, useful to know when a problem might arise, the forecasts of non-specific storms are too frequent to allow preventative shutdowns. For solar protons, forecasts of the expected maximum energy would help to reduce false alarms.

6.4 Trapped Particles

As with the space climate models, trapped particle forecasting capabilities are poor. The "geoeffectiveness" of solar events can't be forecast and existing monitoring is not adequate to cover geospace regions inside of geostationary and below high-latitude inclinations. Since the CRRES mission, there have been some improvements in "post-diction" of events. In 1993 Li et al. (Li, 1993) used a simplified model of the Storm Sudden Commencement (SSC) compression of the magnetosphere to show that electron belts like those measured by CRRES can be created in tens of seconds when the interplanetary shock wave from the storm interacts with the magnetosphere. Later Hudson et al. (Hudson, 1996) showed that this shock acceleration theory could also be used to explain the sudden formation of proton storm belts. Bourdaire et al. (Bourdaire et al., 1995) are developing a 4-D diffusion code to calculate the transport of particles throughout the inner magnetosphere. Case studies have been validated using CRRES and STRV-1b measurements.

7. CAPABILITY GAPS

In spite of recent developments in space climate and space weather models and forecast capabilities, serious shortcomings remain, including:

Climate Models

- Solar heavy ions: larger database required for statistical analysis
- Solar protons: larger database required for statistical analysis and higher energies needed

Trapped particles: no statistical information to predict extreme climates or climate distribution as a function of confidence level; inaccurate predictions at extremes of energy spectra; unknown accuracy in many regions, particularly at GEO and MEOs; lack of long term (yearly) and short term (hourly) time resolution; lack of understanding of variations in "slot region"; lack of understanding of duration of slot region storm populations

Weather Models

Solar heavy ions: no forecast or nowcast capability

Solar protons: inaccurate forecast capability, forecasts do not have energy resolution

Trapped particles: little forecast capability, nowcasts not available for all regions

The Living With a Star Program was established to address research aspects of space weather. The Targeted Research and Technology Element of LWS has funded grants to support improvements in the models (LWS, 2003). Examples of current efforts on the radiation belt models climate models are:

Understand the fundamental plasma interactions and particle transport processes responsible for the extreme conditions that pose the most serious threat to space systems

Address the deficiencies in the AE-8 models by understanding variance from long-term average models and worst case levels

Develop time-dependent maps of energetic particles fluxes in inner magnetosphere

Establish Center for Space Radiation Modeling (CIRM), data acquisition and management, construction, validation, dissemination

Develop quantitative of the geomagnetic field that is valid in the entire geospace region

Understand long-term dynamics of the trapped radiation slot region

Understand variability in the Low Earth Orbit plasma environment

Determine the conditions in the solar wind and within the magnetosphere that are responsible for the variability of relativistic electrons

Understand the dynamics of energetic electron fluxes in the inner magnetosphere, produce electron models coupled to the solar wind variables

LWS is also supporting numerous efforts to improve space weather modeling capability by funding research that addresses physical processes from the interior of the sun to the ionosphere. The reader is referred to the

LWS website for a list of those efforts (LWS, 2003). ISES members also support numerous research projects to understand space weather processes.

In spite of the infusion of opportunity provided recently by space agencies, major roadblocks to developing effective space climate and space weather models remain:

Other than the GCRs, long-term baseline measurements throughout Sun-Earth space do not exist for understanding trends and for model validation. In particular, the radiation belt missions in the LWS program are inadequate to provide the required solar cycle and regional coverage.

Transition of research models to validated space climate and space weather models is not adequately addressed. Space programs are reluctant to increase the risk of mission failures by endorsing the use of new, unproven environment models to guide mission designs and operations.

The length of time required to "authorize" space environment models is too long. More support for these activities from agencies is required, and there needs to be stronger international cooperation.

The assessment of user requirements needs to be more formal and coordinated between agencies.

More opportunities for researchers and users to interact effectively are required. NOAA Space Weather Week and the NATO Advanced Research Workshops should continue to bring together researchers and users for open discussions of requirements. ILWS and LWS should include users on their panels and task groups.

8. SUMMARY

The protection of space assets requires attention to the effects of space environments through all phases of mission design, development, and operation. Space weather models only address post-launch phases when it is difficult to effectively prevent anomalies. Reducing the risk of anomalies in pre-launch phases requires space climate models which have not received as much attention by the international community. Neither space climate nor space weather models meet current or future needs of spacecraft designers or operators. The lack of resources is not the only obstacle to the development and implementation of effective space environment models. Serious thought needs to be given to requirements definition, model transition from research to applications, and model standardization. The most important need is for increased communication between research, application, and user communities.

9. ACKNOWLEDGEMENTS

Many ideas presented here are the result of discussions with members of the Flight Data Systems and Radiation Effects Branch at GSFC, especially Dr. Robert Reed who has been instrumental in pursuing computational physics methodologies for application to problems related to spacecraft and instrument reliability. The author wishes to thank Dr. Daglis for his efforts to provide a forum for researchers and users to interact and for supporting the documentation of that interaction.

10. REFERENCES

Allison, J., J. Apostolakis, G. Cosmo, P. Nieminen, M.G. Pia, for the Geant4 Collaboration, Geant4 status and results, in *Proc. of CHEP-2000 Conference*, pp. 81-84, Padua, 2000.

Baker, D.N., Satellite anomalies due to space storms, Chapter 19 in *Space Storms and Space Weather Hazards*, edited by I. A. Daglis, Kluwer Academic Publishers, Dordrecht, 2001.

Barth, J.L., C.S. Dyer, and E.G. Stassinopoulos, Space, Atmospheric, and Terrestrial Radiation Environments, IEEE Trans. on Nucl. Sci., vol. 50, no.3, pp. 466-482, June 2003.

Bourdaire, S., D. Boucher, and T. Beutier, Modeling the charged particle transport in the Earth's internal magnetosphere, in Workshop of the GdR Plasmae, Tournon, France, Apr. 10-12, 1995.

Brautigam, D.H., M.S. Gussenhoven, and E.G. Mullen, Quasi-static model of outer zone electrons, IEEE Trans. on Nucl. Sci., vol. 39, no. 6, pp. 1797-1803, December 1992.

Daly, E.J., ESA space weather activities, Chapter 19 in *Space Storms and Space Weather Hazards*, edited by I. A. Daglis, Kluwer Academic Publishers, Dordrecht, 2001.

Daly, E.J., ESA Space weather programme, available at http://www.estec.esa.nl/wmwww/wma/spweather

Daly, E.J., J. Lemaire, D. Heynderickx, and D.J. Rodgers, Problems with models of the radiation belts, IEEE Trans. on Nucl. Sci., vol. 43, no. 2, pp. 403-415, April 1996.

Dyer, C.S., C. Sanderson, R. Mugford, C. Watson, and C. Peerless, Radiation environment of the Microelectronics and Photonics Test Bed as measured by CREDO-3, IEEE Trans. on Nucl. Sci., vol. 47, no.3, pp. 481-485, June 2000.

Dyer, C.S., K. Hunter, S. Clucas, D. Rodgers, A. Campbell, and S. Buchner, Observation of solar particle events from CREDO and MPTB during the current solar maximum, IEEE Trans. on Nucl. Sci., vol. 49, no. 6, pp. 2771-2775, December 2002.

Feynman, J., G. Spitale, J. Wang, and S. Gabriel, Interplanetary fluence model: JPL 1991, J. Geophys. Res., vol. 98, pp 13281-13294, 1993.

Gussenhoven, M.S., E.G. Mullen, M.D. Violet, C.Hein, J. Bass, and D. Madden, CRRES high energy proton flux maps, IEEE Trans. on Nucl. Sci., vol. 40, no. 6, pp. 1450-1457, December 1993.

Gussenhoven, M.S., E.G. Mullen, M. Sperry, and K.J. Kerns, The effect of the March 1991 storm on accumulated dose for selected orbits: CRRES dose models, IEEE Trans. on Nucl. Sci., vol. 39, no. 6, pp 1765-1772, December 1992.

Gussenhoven, M.S., E.G. Mullen, J.T. Bell, D. Madden, and E. Holeman, APEXRAD: low altitude orbit dose as a function of inclination, magnetic activity and solar cycle, IEEE Trans. on Nucl. Sci., vol. 44, no. 6, pp. 2161-2168, December 1997.

Heynderickx, D., Trapped radiation environment model development, available at http://www.magnet.oma.be/home/trend/trend.html

Heynderickx, D, M. Kruglanski, V. Pierrard, J. Lemaire, M. D. Looper, and J.B. Blake, A low altitude trapped proton model for solar minimum conditions based on SAMPEX/PET data, IEEE Trans. on Nucl. Sci., vol. 46, no. 6, pp. 1475-1480, December 1999.

Hudson, M.K., A.D. Kotelnikov, X.Li, I. Roth, M. Temerin, J. Wygant, J.B. Blake, and M.S. Gussenhoven, Simulation of proton radiation belt formation during the March 24, 1991 SSC, Geophys. Res. Lett., Vol. 22, No. 3, pp. 291-294, February 1, 1995.

Huston, S.L. and K. A. Pfitzer, A new model for the low altitude trapped proton environment, IEEE Trans. on Nucl. Sci., vol. 45, no. 6, pp. 2972-2978, December 1998.

Joselyn, J., State of the Art in space weather services and forecasting, Chapter 17 in Space Storms and Space Weather Hazards, edit by I.A. Daglis, Kluwer Academic Publishers, 2001.

Kauffman, W., NASA Space environment and effects program, available at http://see.msfc.nasa.gov

King, J.H., Solar proton fluences for 1977-1983 space missions, J. Spacecraft and Rockets, vol. 11, no. 6, pp. 401-408, June 1974.

Koons, H.C., J. E. Mazur, R.S. Selesnick, J.B. Blake, J.F. Fennell, J.L. Roeder, and P.C. Anderson, The impact of the space environment on space systems, Aerospace Technical Report TR-99(1670)-1, 1999.

Lemaire, J.F., From the discovery of the radiation belts to space weather perspectives, Chapter 3 in Space Storms and Space Weather Hazards, edited by I.A. Daglis, Kluwer Academic Publishers, 2001.

Li, X., I. Roth, M. Temerin, J. R. Wygant, M. K. Hudson, and J. B. Blake, Simulation of the prompt energization and transport of radiation belt particles during the March 24, 1991 SSC, *Geophys. Res. Lett.*, Vol. 20, p. 2423-2426, 1993.

LWS, Living With a Star Program, available at http://lws.gsfc.nasa.gov

LWS, Living With a Star Program, Targeted Research and Technology Element, available at http://lws.gsfc.nasa.gov/tmda.htm

Mazur, J., 2002 IEEE NSREC Short Course, Radiation effects - from particles to payloads, Section II: The radiation environment outside and inside a spacecraft, July 15, 2002, Phoenix, Arizona.

National Space Weather Program Council, National space weather program, The implementation plan, 2nd Edition, FCM-P31-2000, Washington, DC, JULY 2000.

Panasyuk, M.I., Cosmic ray and radiation belt hazards for space missions, Chapter 10 in Space Storms and Space Weather Hazards, edited by I.A. Daglis, Kluwer Academic Publishers, 2001.

Poivey, C., J.L. Barth, K.A. LaBel, G.B. Gee, and H. Safron, In-flight observations of long-term single-event effect (SEE) performance on Orbview-2 solid state recorders (SSR), Nuclear and Space Radiation Effects Conference, Radiation Effects Data Workshop, July 2003.

Sawyer D.M. and Vette J.I., AP8 trapped proton environment for solar maximum and solar minimum', NSSDC 76-06, 1976.

Sibeck, D., The international living with a star program, available at http://ilws.gsfc.nasa.gov

Truscott P., F. Lei, C. Ferguson, R. Gurriaran, P. Nieminen, E. Daly, J. Apostolakis, S. Giani, M.G. Pia, L. Urban, M. Maire, Development of a spacecraft radiation shielding and effects toolkit based on Geant4, CHEP-2000 Proc., 2000.

Tylka, A.J., F. Dietrich, and P. R. Boberg, CREME96: A revision of the cosmic ray effects on microelectronics code, IEEE Trans. on Nucl. Sci., vol. 44, no. 6, pp. 2150-2160, December 1997.

Tylka, A.J., F. Dietrich, and P. R. Boberg, Probability distributions of high-energy solar-heavy-ion fluxes from IMP-8: 1973-1996, IEEE Trans. on Nucl. Sci., vol. 44, no. 6, pp. 2140-2149, December 1997.

Vette, J.I., The AE-8 trapped electron model, NSSDC/WDC-A-R&S 91-24, 1991.

Walters, L., Monte Carlo N-Particle, http://mcnpx.lanl.gov, 2003.

Xapsos, M.A., J. L. Barth, E. G. Stassinopoulos, G. P. Summers, E. A. Burke, G. B. Gee, Model for prediction of solar proton events, Proceedings from the Space Radiation Environment Workshop, Farnborough, UK, 1-3 November 1999.

Xapsos, M.A., G. P. Summers, J. L. Barth, E. G. Stassinopoulos, and E. A. Burke, Probability model for worst case solar proton event fluences, IEEE Trans. on Nucl, Sci., vol. 45, no. 6, pp. 1481-1485, December 1999.

Xapsos, M.A., J. L. Barth, E. G. Stassinopoulos, E. A. Burke, and G. B. Gee, Model for emission of solar protons (ESP) – cumulative and worst case event fluences, NASA-Marshall Space Flight Center SEE Program, http://see.msfc.nasa.gov.

Chapter 8

The relation of high- and low-orbit satellite anomalies to different geophysical parameters

A. Belov[1], L. Dorman[2], N. Iucci[3], O. Kryakunova[4], N. Ptitsyna[5]

1. *IZMIRAN, Russian Academy of Science, Moscow, Russia*
2. *Cosmic Ray Center, affiliated with Tel Aviv University and Israel Space Agency, Israel*
3. *Dipartimento di Fisica "E. Amaldi", 'Roma-Tre' University, Rome, Italy*
4. *Institute of Ionosphere, Kazakhstan*
5. *SPb FIZMIRAN, Russian Academy of Science, St. Petersburg, Russia*

Abstract Satellite anomaly data in the period 1971-1994 were analyzed in the search of possible influence of different space environmental parameters. The database was created by combining, beyond the malfunction information, various characteristics of space weather: geomagnetic activity indices, fluxes and fluences of electrons and protons at different energy, high energy cosmic ray variations, solar wind characteristics and other solar, interplanetary and geophysical data. Satellites were divided on several groups according to the orbital characteristics (altitude and inclination). It was found, that the relation of satellite malfunctions to the environmental parameters is different for various orbits. This fact should be taken into account for the developing of malfunction frequency models.

Keywords Space weather, satellite anomalies, geomagnetic activity, proton enhancements, relativistic electrons

1. INTRODUCTION

Satellites usually spend several years in the space under the influence of variable electromagnetic fields, plasma and different radiations. Since satellites are not protected by the atmosphere and to some extent also by the magnetosphere, they are much more exposed to the cosmic radiation than ground level devices. This is one of the reasons why space weather changes can be hazardous for satellites. It is known that high energy cosmic rays (CR) of galactic and solar origin lead to single-event upsets (SEU) in

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 147-163.

microelectronic devices; low energy electrons create electric charge on the satellite surface and can cause solar battery degradation; and high energy electrons may create volume charging inside the satellite and damages to the operating electronic (see e.g., Adams et al., 1981; Gussenhoven et al., 1985; Wilkinson et al., 1991; Shea et al., 1992; Wrenn, 1995; Baker et al., 1998). Geomagnetic storms are also dangerous for satellites, not only because of electromagnetic field variations, but also because of their influence on the charged particle access to a particular orbit and on particle precipitation from radiation belts (Lanzerotti, 1979; Wilkinson, 1994). During magnetic storms the density of the upper atmospheric layers increases; this may cause changes in the orbit for the low-orbit satellites, and even the loss of their orientation. The fuller listing of satellite failures with space weather association can be found in the literature (see e.g., Stephen, 1993; Fredrickson, 1996; Koskinen et al., 1999; Feynman and Gabriel, 2000). Anyway, it is clear from all results, that the influence of space weather on satellites is complicated and variable. Besides, a degree of this influence and possible damages depend significantly on the satellite location and characteristics (e. g. Vampola, 1994; Hastings, 1995). Two basic methods can be used to estimate the probability of satellite anomalies: 1) a search for relation between anomalies and global characteristics of space weather to create the models suitable for all satellites or for groups of satellites; 2) an environmental monitoring on separate satellites and the estimation of danger based on these local observations. Both approaches seem to be insufficient. On one side, computations cannot replace measurements in situ. On the other side, single satellite observations cannot furnish a complete picture of the environmental variations. It would be convenient to combine the two approaches.

In this paper we analyze the relationship between satellite malfunctions and different geo- and helio-physical parameters. For this work we use either environmental characteristics out of magnetosphere, or global characteristics of magnetosphere, as the planetary indices of geomagnetic activity. We will also take into account the main specific features of satellites represented by their orbital characteristics.

2. DATA AND METHODS

Data on satellite anomalies and different characteristics of space weather were combined into a special database (Belov et al, 2003). The main part of satellite malfunction data was taken from NGDC satellite anomaly database (Wilkinson, 1994). A substantial contribution was also given by “Kosmos” satellite data (circular orbit at 800 km altitude and 74º inclination). The

majority of anomalies in 1994 were taken from NASA report (Thomas, 1995). The satellite characteristics have been taken from different Internet sources (http://spacescience.nasa.gov/missions/, http://www.skyrocket.de/, http://www.astronautix.com, http://hea-www.harvard.edu). Our database is formed by a total of ~300 satellites and ~6000 anomalies. Since the information on satellite anomalies before 1971 and after 1994 was rather fragmented, we limited our study for the period 1971-1994. However, within this interval the information on malfunctions is not uniform (see Figure 1).

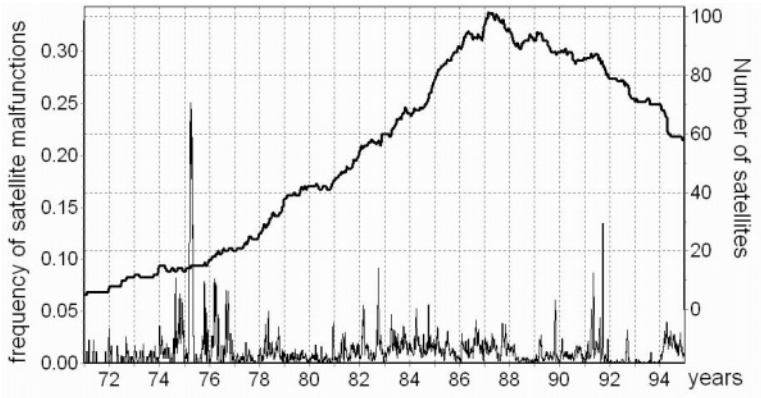

Figure 1. Number of satellites giving information about malfunctions, and frequency of satellite malfunctions ($day^{-1} \cdot satellite^{-1}$) in the period 1971-1994.

The high frequency of anomalies in 1974-1976 is caused by very small number of satellites operated in this period. The maximum number of satellites occurred in 1987; this feature and the following decrease does not look real. Unfortunately, the data incompleteness (the majority of satellite owners prefer not to give information on malfunctions) poses additional problems to the analysis.

All satellites were divided into different groups according to the altitude and inclination of their orbits. In Figure 2 each orbit is presented by a different mark. Sometimes, one point represents many satellites with very close orbits, as the majority geostationary satellites (>100), and 49 "Kosmos" spacecrafts. Since there are no satellites within the wide range 1500-15000 km, the altitude division of satellites was rather easy. It was more difficult to divide satellites according to the orbital inclination. On one hand, it was important to separate orbits fully inside the magnetosphere from those only partly inside. On the other hand, we admit that the "Shuttle" spacecrafts are too specific to be combined with the other satellites. Finally we chose 58º as the inclination boundary.

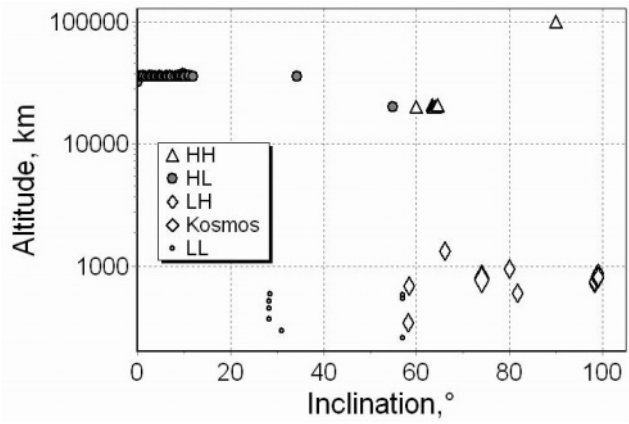

Figure 2. Altitude-inclination distribution by the of the satellite orbits.

As a result, four groups were obtained, with essentially different physical conditions on the orbits: HL (high altitude-low inclination), HH (high altitude-high inclination), LH (low altitude-high inclination) and LL (low altitude-low inclination). HL group contains all GEO satellites and is the most abundant. LH group is approximately the half of HL in number, and it is formed with important share of "Kosmos" satellites. HH group comprises only 14 spacecrafts, but they displayed more than 1000 anomalies. We have here mainly MEO satellites, but the main difference of this group from LH is not in the altitude but in the orbit inclination. LL group (mainly piloted spacecrafts with the special price of malfunctions) is also important. Unfortunately, this group is too small to be discussed here. Sometimes we combined all low and all high orbital satellites together. Satellite malfunctions, unlike the satellites, were not divided on the groups and were not filtered.

We calculated the mean frequency of malfunction (i.e. the number of malfunctions per day and per satellite) for all satellites and for every separate group (only satellites having more than one malfunction were considered). We analyze only daily mean data. This defines the possibilities and peculiarities of our research. We cannot study short-time features (for example, the local time effect) because our data are not so detailed. They have to be better correlated with large-scale effects and with global rather than local conditions. Naturally, this kind of approach limits the possibilities of analysis. Nevertheless, it has some advantages: daily mean data are more reliable, diverse and available and less dependent on occasional factors.

2. RELATION OF SATELLITE ANOMALIES TO DIFFERENT SPACE WEATHER PARAMETERS

2.1 Two examples

Satellite malfunctions are very irregularly distributed by the time. Some days there is no anomalies, in other days, tens malfunctions are given by several satellites. One famous period (e.g. Allen et al., 1989) with high frequency of malfunctions was October 19-26, 1989 (see Figure 3).

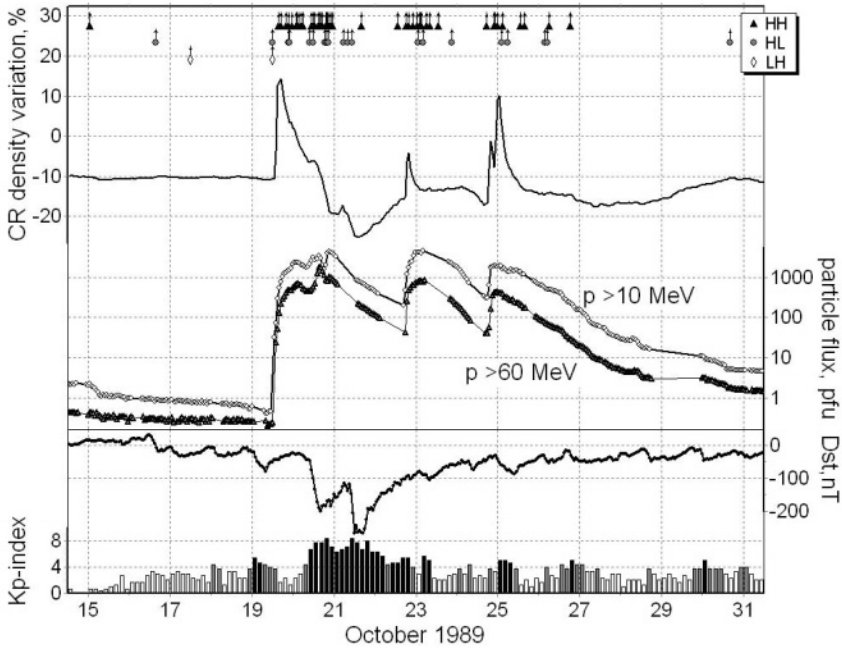

Figure 3. Period with large number of satellite malfunctions in October 1989. Upper panel – cosmic ray activity near the Earth: variations of cosmic ray density, obtained from neutron monitor network; solar proton fluxes (>10 MeV and >60 MeV) recorded by IMP-8. Lower panel – geomagnetic activity: Kp- and Dst-indices. Vertical arrows with points on the upper panel indicate the malfunction in different satellite groups.

In this period we observed several proton events, three Ground Level Enhancements (GLE) of solar cosmic rays (on October 19, 22 and 24), big Forbush-effects, strong geomagnetic storms, including a severe (maximal Kp =8+ and minimal Dst-variation –268 nT) storm on October 20-21. CR variations, derived from neutron monitor data by the global survey method (Belov et al., 1999a), correspond to 10 GV rigidity of galactic CR and to ≈3

GV during GLE. Malfunctions look to be coincided immediately with the maximum of proton enhancements. This connection becomes more evident if we consider the satellite groups. Only one from 73 malfunctions occurred at low altitude, 19 anomalies were recorded at geostationary orbits, and the majority occurred in the HH group, which is maximally exposed to the solar cosmic ray effect. There is usually much smaller number of satellites in this group than in HL (GEO) group, and in this period it was 5.5 times smaller. Thus, on 20 October the malfunction frequency in high altitude-high inclination group was higher than in GEO group by a factor 30.

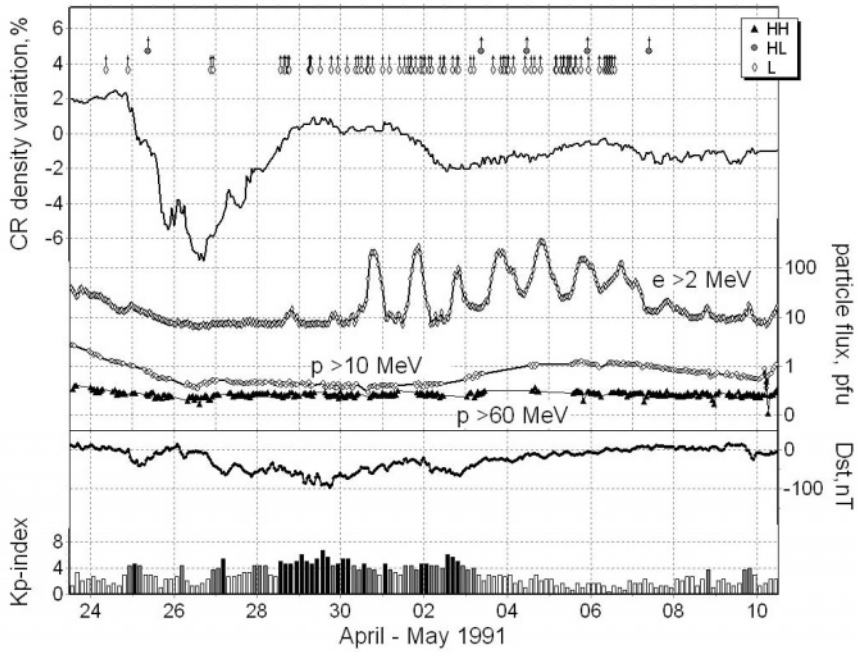

Figure 4. Period with large number of satellite anomalies in 1991. Upper panel – cosmic ray activity near the Earth: variations of cosmic ray density, obtained from neutron monitor network, solar protons of >10 MeV and >60 MeV (IMP-8) and electron fluxes of >2 MeV (GOES). Lower panel – geomagnetic activity: Kp- and Dst-indices. Vertical arrows with points on the upper panel indicate the anomaly in different satellites groups.

Another sample of high frequency satellite malfunctions is presented in Figure 4 for the period April-May, 1991. Here we see a strong magnetic storm (maximal Kp =7- and minimal Dst-variation –97 nT) and big Forbush-effect. There was no significant proton increases, but the flux of relativistic electrons retained on high level during a week. The main amount of

malfunctions happened during the magnetic storm and high electron flux period.

The main feature of this period is that malfunctions were entirely absent in HH group, which played the main role in previous example, and majority of anomalies was observed in low orbital satellites with share of GEO group.

These two examples illustrate a relation between satellite malfunctions and space weather, but this relation is different for different satellite groups. Low correlation has been found between anomaly frequencies at high and low altitudes not only in these examples. Considering the events from our whole database we found out that satellite malfunctions appeared usually on different days at high and low altitudes. Through the period 1975-1994 there was 948 days with ≥2 malfunctions at high and 154 days – at low altitudes. Only 11 days from these subsets coincide. Correlation coefficient between malfunctions at different altitudes over the 1975-1994 was found <0.01. It was close to 0 for any long enough period (3 years or more). The only exclusion were 1992-1994 years, when correlation coefficient increased up to 0.19, that is very probably associated with the increased electron fluxes. Low correlation between anomalies in different satellite groups is the evidence either of the effect of different factors on different groups, or of different character of the same factor influence.

2.2 Seasonal dependence

It is well known that satellite malfunctions have a seasonal dependence, and this is considered as one of the evidences of the relation of anomalies to environmental parameters (e.g. Allen, 1990). In Figure 5 the annual behavior of the malfunction frequency averaged over the period 1975-1994 is presented. To reduce the effect of short-term variations the 27-day running means have been computed. The main feature of this dependence is a semi-annual variation with maxima close to equinoxes. This seasonal behavior is characteristic for the geomagnetic activity indices. We processed Ap-index of geomagnetic activity for the same period and by the same method as the malfunction frequency (see Figure 6).

The comparison between Figures 5 and 6 indicates the similarity of seasonal variations in geomagnetic activity and in satellite malfunctions. Both of them are much higher during spring and autumn than during summer and winter. In these 20 years the spring geomagnetic activity dominated the autumnal one, and the same feature is repeated in the behavior of satellite anomaly frequency. It should be noted that seasonal dependence in the satellite anomalies is better pronounced than in Ap-index. The effect is so big, that it appear to be reasonable that significant amount (or possible, the majority) of satellite anomalies are associated with the environmental

changes. Seasonal dependence was calculated separately for different satellite groups. It was mostly pronounced in HL (GEO) group. In the LH (low altitude – high inclination) group this dependence is approximately in 3 times less, and it is almost absent in HH group.

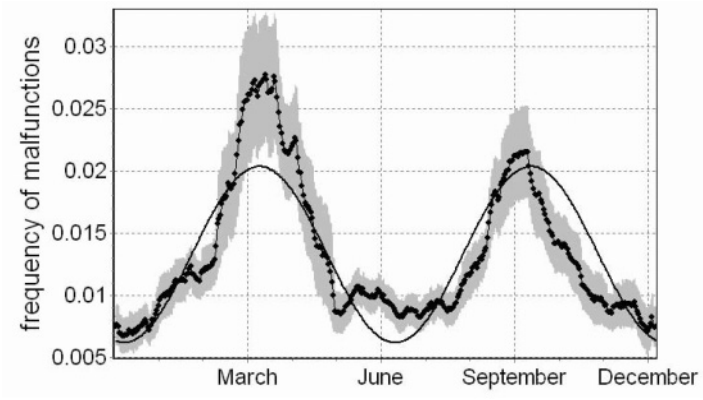

Figure 5. Seasonal dependence of satellite malfunction frequency averaged over the period 1975-1994. The curve with points is the 27-day running mean of frequency; the grey band corresponds to the 95 % confidence interval. The sinusoidal curve is a semidiurnal wave with maxima in equinoxes best fitting the frequency data.

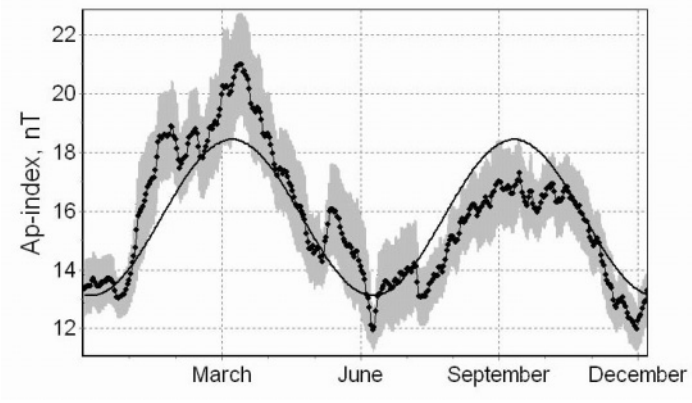

Figure 6. The same as in Figure 5 for the Ap-index of geomagnetic activity.

2.3 Space-weather environmental parameters

The behavior of the daily mean frequency of satellite malfunctions was compared with different characteristics of solar, interplanetary, geomagnetic and cosmic ray activity connected with space weather conditions on the satellite orbits.

Solar activity. As total characteristics of solar activity we used daily sunspot numbers and radio flux at 10.7 cm, as provided by NOAA (ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA). Since the relation of satellite malfunctions with daily solar characteristics was found to be very weak, we used in addition the running means of sunspot numbers, averaged over one year and one solar rotation.

Geomagnetic activity. Ap-, AE- and Dst-indices of geomagnetic activity (ftp://ftp.ngdc.noaa.gov) have been used. We used daily means, and the extreme values of indices: daily maximum value of the 3-hourly Ap and AE indices and minimum hourly value of Dst-index.

Interplanetary medium. Daily mean and maximal hourly solar wind speed and daily interplanetary magnetic field (IMF) intensity were taken from OMNI database (http://nssdc.gsfc.nasa.gov/omniweb/ow.html) as characteristics of near Earth interplanetary space. Some indices were based on the B_z-component of IMF, determined in GSM-coordinates: daily mean and minimal hourly B_z-component and daily sum of negative B_z values. Besides, for all days with sufficient IMF and solar wind velocity data, we estimated the energy transferred into the Earth's magnetosphere, according to Akasofu (1981).

Protons and electrons. Daily fluences of protons of different energies (>1, >10 and >100 MeV) and electrons of >2 MeV energy, calculated from the GOES (ftp://ftp.ngdc.noaa.gov) measurements, have been used as main cosmic ray characteristics. Unfortunately, these data are available only from January 1987 (protons) and from June 1987 (electrons). Together with GOES data, we used the proton fluxes of >10 MeV and >60 MeV, measured by IMP-8 and available in OMNI base for the whole period. To analyze some separate periods we used also electron fluxes >2 MeV from GOES satellites. These data were not included in the model calculations because they are probably contaminated by proton fluxes.

Ground level cosmic rays. We used cosmic ray activity (CRA) indices (Belov et al., 1999b), which characterize the behavior of CR of 10 GV rigidity. They are calculated on the basis of hourly means of CR density and parameters of the first harmonic of CR anisotropy derived from the neutron monitor network by the Global Survey Method. Two different indices were compared with the satellite anomalies: CRA-index based on the CR density and anisotropy, and a simplified index, accounted by density variations only.

CRA indices are strongly connected with the interplanetary and geomagnetic disturbances, and sometimes the greatest proton enhancements are visible in their time behavior.

2.4 Satellite anomalies and SSC

The frequency of malfunctions should vary significantly under sharp changes of the environmental conditions. Space weather strongly changes at the moment of sudden commencements of magnetic storm (SSC), when the interplanetary shock and the solar wind disturbance behind of shock start to interact with the magnetosphere. By analyzing the average behavior of satellite malfunctions by the epoch method, we found that the malfunction frequency increases after SSC.

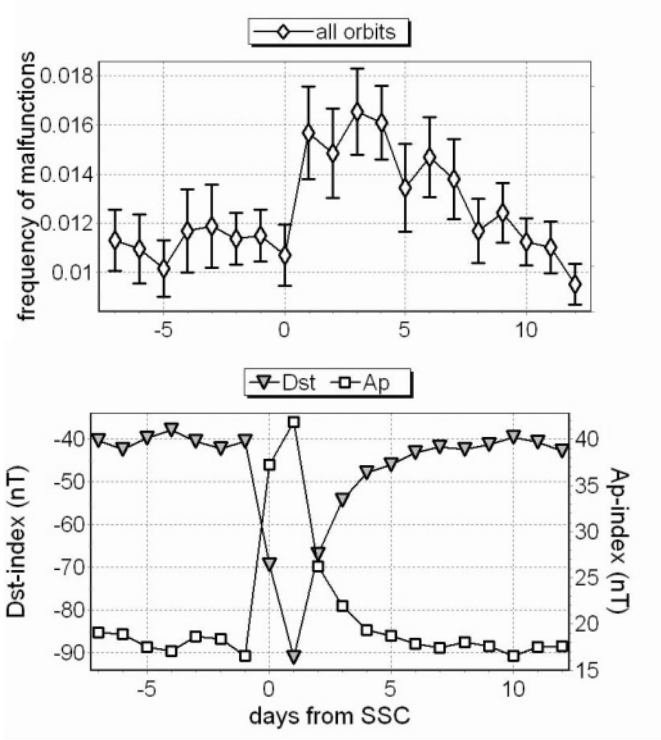

Figure 7. Average behavior of the satellite malfunction frequency, Ap- and Dst-indices in periods of sudden storm commencements. The average was done by epoch method (0 – day of SSC) for 388 magnetic storms with maximal Ap-index >50 nT during 1975-1994.

The largest malfunction frequency increase is observed after SSC in HL-group at geostationary orbits, and it grows with the increase of the magnetic

storm power. In Figure 7 the average variations of the malfunction frequency are presented for magnetic storms with maximal Ap-index >50 nT. For less powerful storms the variations were smaller, for more powerful – bigger. This relation between the magnitude of the effect and the storm power could indicate a direct influence of geomagnetic activity on the malfunction probability. However, one can see from Figure 7 that the frequency increase starts after the magnetic storm onset and lasts much longer. Hence, satellite malfunctions seem to be not always directly related to geomagnetic activity, but depends on some other factors.

2.5 Relation of anomalies to protons and electrons

Space weather changes rapidly not only during SSC, but also during proton increases. Analysis done by epoch method showed that the

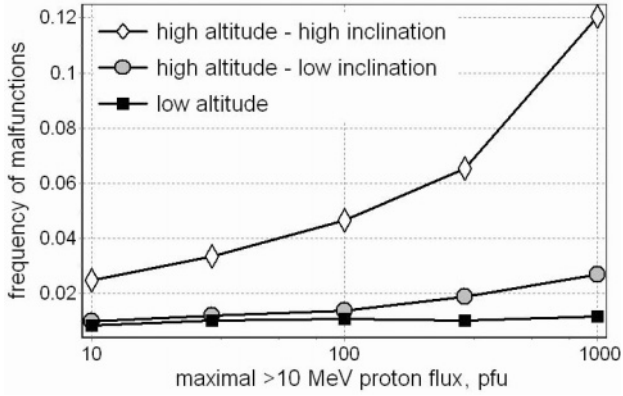

Figure 8. Mean normalized anomaly frequency in the first two days of proton enhancement at different orbits in dependence on maximal >10 MeV proton flux.

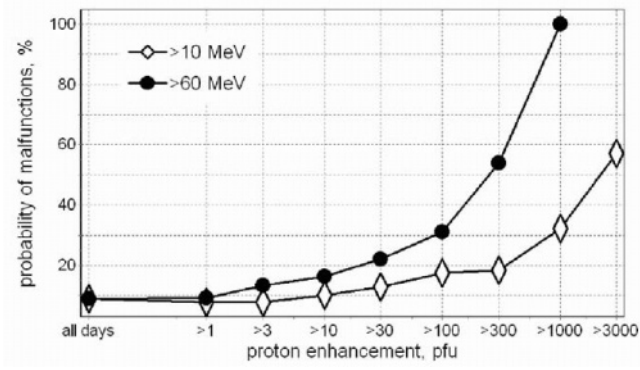

Figure 9. Averaged probability of satellite anomalies in high altitude – high inclination group for days with different maximal proton fluxes of > 10 and >60 MeV (IMP8 data).

frequency of anomalies at high altitudes was significantly larger in the first day of the proton event and also the next day. Moreover, the frequency of anomalies increased with increasing the proton flux. This effect is especially pronounced in the HH group (see Figure 8).

A further evidence of linkage between anomalies registered in HH-group of satellites and proton enhancements is given in Figure 9. As an average, one anomaly per 10 days is registered for satellites in this group, but the anomaly probability increases with the proton flux increasing, and goes to 100% for the days of very big proton events

Electron enhancements, in contrast to the sharp proton enhancements, very often start in a gradual manner. For their study we used a different version of the epoch method, in which the day of each anomaly was chosen as zero-day. Figure 10 shows that the mean fluence of relativistic electrons was maximal in the day when the anomaly was registered. It is important that the electron fluence arises significantly some days before the malfunction. The electron flux variations appeared to be crucial for HL (GEO) and LH groups, but not for the HH group of satellites.

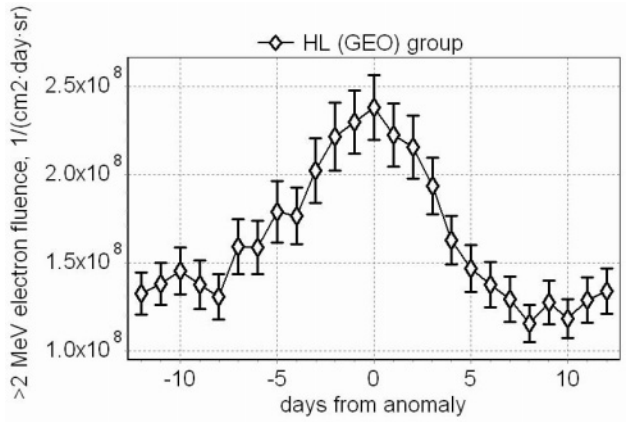

Figure 10. Electron fluences in 1987-94 averaged by the epoch method. 0-day is the satellite anomaly day.

Fig. 11 shows a difference in proton and electron result on anomalies for different orbits. In the anomaly day the mean proton fluence is much higher in HH group than in other groups. LH group is mainly electron-dependent, and HL group may be considered as mixed one.

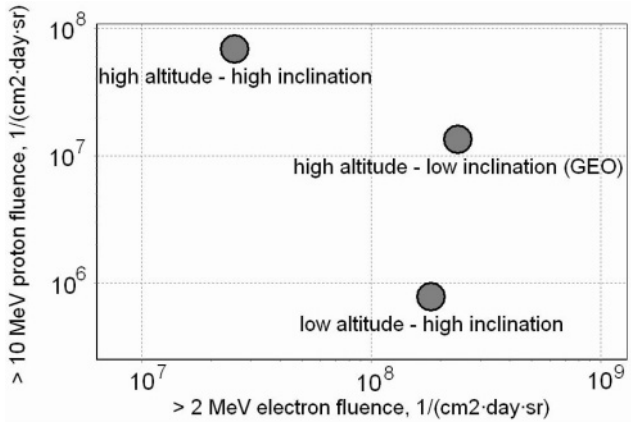

Figure 11. Mean proton and electron fluences in the anomaly day (1987-1994).

2.6 Modeling of satellite anomaly frequency

We examined a relation between different space weather parameters (>30 in total) together with their combinations, and satellite anomalies at different orbits in 1987-1994. This period of time was chosen, because of the presence of electron fluence data, that is very important for the modeling.

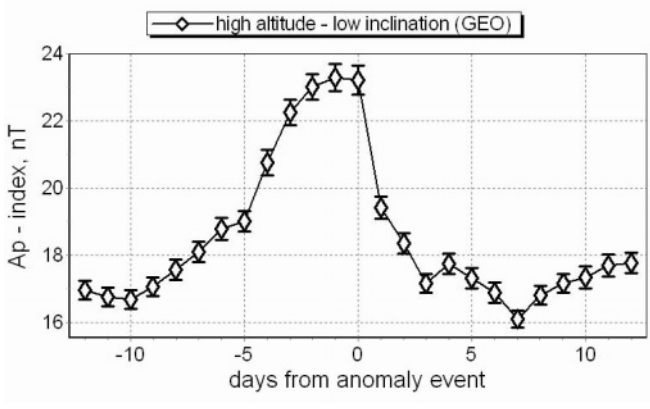

Figure 12. Averaged Ap-index in 1975-94 obtained by the epoch method. 0-day is the anomaly day.

An increase of the environmental index some days before the satellite anomaly, as shown in Figure 10, is a characteristic behaviour not only for electrons, but also geomagnetic activity (see Figure 12), solar wind speed

and some other indices as well. It leads to a conclusion, that there are special recurring conditions lasting several days in space environment, which may be considered as a factor contributing to initiate satellite anomalies. Independently on the reasons or nature of this kind of anomaly precursors, it is possible and necessary to use them for modeling and forecasting. Thus, in our modeling we used space weather parameters for the anomaly day, and for several preceding days. The simple linear regression simulation was used. However, links of registered anomalies with electron, and especially with proton indices appeared to be non-linear (see Figure 8). Therefore, we applied, as an exception, power law dependence for the proton and electron fluxes and fluences.

Models of anomaly frequencies were obtained in three steps due to big number of space weather characteristics. Firstly, for each index simple regression analysis was performed and those that demonstrated the higher correlation with frequency of anomalies were chosen. Such indices for HL group of satellites, for example, were solar wind velocity, >2 MeV electron fluence, geomagnetic activity indices Ap and Dst, flux of proton with energy >60 MeV. Then 3-5 indices that show the best correlation were combined in a many-parameter model. The best indices from the first step not always retained the best at the next stage. For instance, solar wind velocity can show better correlation than electron fluence when they are correlated separately, but in a model comprising both parameters, solar wind index become of the second importance. On the last phase we added remaining indices one by one and if the model was improved significantly, these characteristics were kept in the model. Some peculiarities of obtained models, simulating the frequency of satellite malfunctions by means of 5-8 different indices, are presented in Table 1. Index sequence and letter size in the names of parameters reflect the contribution of this index to the model.

Table 1. Models of the satellite malfunction frequency.

Group	HL	HH	LH
Parameters of model	**e2** p100, p60d sf, Ap, Vsw Bz, da10	p60d, p100 Eak SSN365, Bzns	**e2** Dst,AE,sf,CRA,Bz Vsw

Explanations to the Table: e2 and p100 – >2 MeV electron and >100 MeV proton fluences (GOES); p60d – daily mean flux of >60 MeV protons (IMP8); Ap and AE –indices of geomagnetic activity, Eak – estimation of energy incoming to the magnetosphere (Akasofu, 1981); Vsw – solar wind velocity; Bz – daily mean B_z-component of IMF, Bzns – sum of negative values Bz- component; SSN365 – yearly running averaged sunspot number;

CRA and da10 - cosmic ray activity indices, obtained from neutron monitor network data. Seasonal factor sf (semi-annual variation with maxima on equinoxes) was used as one of the independent parameters.

Correlation coefficients k between observed and simulated values of anomaly frequencies are k= 0.24 for LH satellites, k= 0.39 for HL and k= 0.7 for HH – satellite groups. The model examples for frequency of anomalies f in different satellite groups are given by the following expressions in 10^{-4} $day^{-1}satellite^{-1}$ units:

$$f_{HL} = -54 + 1.4\cdot 10^{-9}\langle e2\rangle_4^{1.2} + 0.83\langle Ap\rangle_5 + 0.19\langle Vsw\rangle_2 - 0.15\langle Bz\rangle_3 + 1.1(p100)^{0.35} + 1.6(p60d)^{0.75} + 20sf + 1.5da10;$$

$$f_{LH} = -16 + 2.2\cdot 10^{-7} e2 + 0.29\langle AE\rangle_2 - 26\langle Bz\rangle_7 + 0.83\langle Dst\rangle_6 + 45\cdot sf + 8.9\langle CRA\rangle_4 + 0.23\langle Vsw\rangle_2;$$

$$f_{HH} = -85 + 6.5(p100)^{0.35} + 2.6(p60d)^{4.4} - 0.53\langle Bzns\rangle_4 + 14\langle Eak\rangle_4 + 0.09\cdot SSN365.$$

where $\langle a\rangle_n$ is the a parameter averaged by the day of anomaly together with n-1 preceding days. The units used in this expression are: nT for Ap, Bz and Bzns, km/s for Vsw, % for da10, electrons·day^{-1}·cm^{-2}·sr^{-1} for e2, protons·$day^{-1}cm^{-2}$·sr^{-1} for p100, and in protons·sec^{-1}·cm^{-2}·sr^{-1} for p60d. SSN365 and normalized Eak are in dimensionless units, dimension of the coefficients and parameters in the equations is omitted.

These equations are presented here as model illustration. They should not be considered as accurate or the only possible description. This is a basis for more advanced models. The coefficients and even kind of model clear to be strongly dependent on chosen time period and satellite set. They will be dependent on many occasional factors as, for example, data gaps. More stable are the sets of parameters involved in the models for different satellite groups and presented in Table 1.

3. CONCLUSIONS

The obtained models describe the relation of the occurrence of satellite anomalies to the space weather parameters in rather complex way and they differ significantly for different satellite groups. They combine the cosmic ray and geomagnetic activity indices, solar wind characteristics and some other parameters. The characteristics of the obtained models allow them to be used for the satellite anomalies forecasting. However, it is difficult to apply these models to the majority of present satellites. The models are

obtained for the anomalies registered 10 years ago or even earlier. Nowadays, satellites, and especially their electronic parts, are completely different. However, some features of the obtained models, as their multi parameter nature, accounting for the global characteristics of space weather and difference for various orbits, should be still valid. Models of satellite anomaly frequency should be improved. First of all it is necessary to increase the malfunction database; during our analysis we were permanently aware that our anomaly database was not representative enough. Moreover, the models can be improved by combining global and local (registered in situ or calculated for the location of the satellite) parameters; by taking into account in more details the satellite position at the time of the anomaly occurrence (local time, latitude); and considering the individual characteristics of satellites (mass, lifetime and others) and the type of anomaly.

4. ACKNOWLEDGEMENTS

This work was supported by INTAS grant N00-0810 and in part by INTAS N2000-752 and RFBR 01-02-17580 grants. This research was based on the collaboration of the following groups: IZMIRAN (V. Chizhenkov, E. Eroshenko, L. Gromova, D. Ivanus, A. Levitin, V. Yanke); Roma.Tre University (G. Villoresi, M. Parisi); SPb-IZMIRAN (M. Tyasto, E. Vernova, O. Danilova), Israel Cosmic Ray Center (L. Pustilnik, I. Zukerman); Irkutsk ISZF (V. Dvornikov, V. Sdobnov); Institute of Ionosphere of National Academy of Kazakhstan (E. Dryn, N. Nikolaevsky), Kiev, MAO (B. Shakhov, Yu. Fedorov).

5. REFERENCES

Adams J.H., Silberberg R., Tsao C.H., Cosmic ray effect on microlectronics in the near-Earth partcle environment, NRL Memo. Rep., p.4506, 1981.

Akasofu S., Energy coupling between the solar wind and the magnetosphere, Space Science Reviews, 28, 2, 121, 1981.

Allen J., Frank L., Sauer H., Reiff P., Effects of the March 1989 Solar Activity, EOS: Transactions, American Geophysical Union, Vol. 70, No. 46, 1479-1488, 1989.

Baker D.N., Allen J.H., Kanekal S.G., Reeves G.D., Disturbed space environment may have been related to pager satellite failure, Trans. AGU 79, p.477, 1998.

Belov A.V., E.A. Eroshenko and V.G. Yanke, Cosmic ray effects caused by great disturbances of the interplanetary medium in 1990-1996, Proc. 26th Int. Cosmic Ray Conf., 6, 431, 1999a.

Belov A.V., E.A. Eroshenko E.A., V. G. Yanke V. G., Indices of the cosmic ray activity as reflection of situation in interplanetary medium, ESA Space Weather Workshop papers, ESTEC, p. 325-328, 1999b.

Belov A., Chizhenkov V., Dorman L., Eroshenko E., Gromova L., Iucci N., Levitin A., Parisi M., Ptitsyna N., Tyasto M., Vernova E., Villoresi G., Yanke V., The Relationship Between Space Weather Conditions and Spacecraft Operational Anomalies, (in press), 2003

Fredrickson A.R., Upsets related to spacecraft charging, IEEE Trans. Nucl. Sci. NS-43, 2, 426, 1996.

Hastings D.E., A review of plasma interactions with spacecraft in low Earth orbit. Journal of geophysical Research. 100, A8, 14457-14483, 1995

Koskinen et al., Space Weather and interaction with spacecraft; SPEE Final Report ISBN 951-697-509-7 Finnish Meteorological Institute, Helsinki, 1999

Lanzerotti, L. J., Geomagnetic influences on man-made systems, J. Atm. Terr. Phys., 41, 787-796, 1979.

Shea M.A., Smart D.F., Allen J.H., Wilkinson D.C., Spacecraft problems in association with episode of intense solar activity and related terrestrial phenomena during March 1991, IEEE Trans. Nuc. Sci. vol. 39 N 6, 1992

Stephen J.H., Hazard to electronic in space, NATO ASI ser. E, 245, 407, 1993.

Thomas W.B., Orbital Anomalies in Goddard Spacecraft for Calendar Year 1994, NASA technical paper 9636, 1995.

Vampola A.L., Analysis of environmentally induced spacecraft anomalies. J. Spacecraft and Rockets, 31, 154-159, 1994.

Wilkinson, D.C., National Oceanic and Atmospheric Administration's Spacecraft Anomaly Data Base and Examples of Solar Activity Affecting Spacecraft, J. Spacecraft and Rockets, 31 No 2, p. 160, 1994.

Wrenn, G. L., Conclusive evidence for internal dielectric charging anomalies on geosynchronous communications spacecraft, J. Spacecraft and Rockets, 32, pp 514-520, 1995.

Chapter 9

SIMULATION OF SPACE RADIATION EFFECTS IN MICROELECTRONIC PARTS

Alexander Y.Nikiforov, Alexander I. Chumakov
Spacialized Electronic Systems
115409, Moscow, Russia

Abstract This chapter describes the main space radiation effects in integrated circuits (total dose and single event effects) and experimental approach of their simulation in microelectronic parts. Simulation experimental techniques are presented to obtain the main radiation sensitivity parameters of IC.

Keywords Space radiation, radiation effects, radiation hardness, simulator, integrated circuit, upset, latchup.

1. INTRODUCTION

Space weather and technologies are the essential constituent of humanity. We can not imagine our life without data obtained from spacecraft for weather forecast, communication, science, *etc*. Unfortunately, there are limitations of spacecraft lifetime mainly due to space radiation effects in microelectronic parts - integrated circuits (IC) and semiconductor devices.

The main reason of space electronic equipment failures and upsets deals with two dominant radiation effects: total dose and single event effects (SEE). Consistent degradation of IC electrical parameters because of positive charge trapping in dielectric layers determines IC failures due to total dose effects. It is obvious that the influence of these failures becomes more considerable as a spacecraft lifetime increases. SEE include upsets and failures in ICs caused by direct ionization of heavy ions cosmic rays those have transferred into IC element sensitive volume. As interaction of single nuclear particle is a random process, SEEs are observed rather in irregular intervals and their frequency depends on nuclear particle intensity. Maximum SEE frequency is obtained for extreme large solar flares.

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 165-184.

2. BASIC RADIATION EFFECTS

2.1 General

The wide variety of space radiation (electrons, protons, heavy charged particles, *etc*.) affect spacecraft microelectronic parts and cause a limited number of dominant radiation effects (Ma and Dressendorfer, 1989, Holmes-Siedle and Adams, 1993, Nikiforov et.al., 1994) (Fig.1), namely:

a) total ionizing dose effects (charge trapping in dielectric layers and surface interface);

b) single event effects (local ionization along a track of a single nuclear particle);

c) displacement effects (radiation defects in semiconductor lattice).

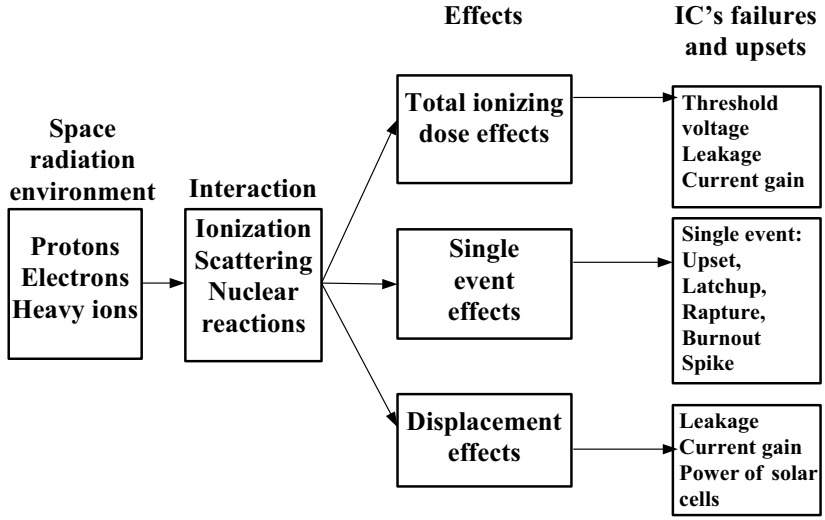

Figure 1. The dominant radiation effects in IC under space radiation

2.2 Displacement effects

The creation of radiation induced defect in semiconductor takes place if nuclear particle transferred energy to the atomic lattice is more than the displacement threshold energy E_d. Displaced atom called primary recoil atom (PRA) also creates additional structural damage if its energy is more than E_d. The stable permanent radiation defects in silicon are complexes, which include connections of primary radiation defects (vacancy, interstitial) with themselves or impurity atoms. Permanent radiation defects cause deep energy levels in a forbidden zone and can act as a recombination centers (reducing minority carriers lifetime), as trapping centers (reducing

concentration of main carriers) and as scattering centers (reducing mobility of free carriers). There are bipolar transistors, solar cells, GaAs devices, optocouplers which are the most sensitive devices to displacement effects.

Displacement effects are determined by proton irradiation in space environment and depend on total non-ionizing energy losses k_{str} (Fig.2) (Summers, 1987).

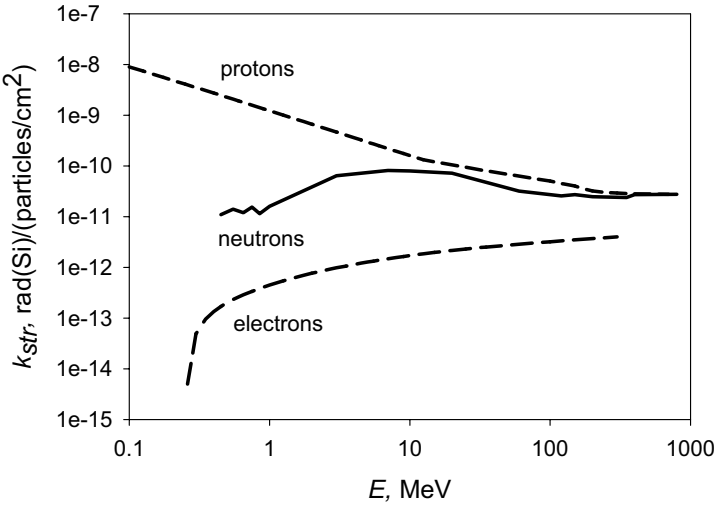

Figure 2. Non-ionizing energy losses of electrons, protons and neutrons in Si (Summers, 1987).

As a rule neutron irradiation is chosen for estimation of IC sensitivity to displacement effects. The typical neutron fluence for IC failures is about $10^{12}…10^{14}$ neutron/cm^2. If the failure threshold neutron fluence is known, then proton fluence is determined from the following equation:

$$\int \Phi_{in\_p}(E_p) \cdot k_{str\_p}(E_p) dE_p = \int \Phi_n(E_n) \cdot k_{str\_n}(E_n) dE_n \qquad (1)$$

A rough estimation can be obtained as follows:

$$D_p \approx k_p \Phi_n; \qquad (2)$$

where $k_p \cong 10^{-7}$ rad·cm^2/neutron. Thus, IC radiation hardness to space proton irradiation will be about 10^6 rad(Si) for $\Phi_n = 10^{13}$ neutron/cm^2.

The most radiation sensitive parameter is lifetime of minority (excess) charge carriers. Radiation-induced lifetime degradation can be estimated as:

$$1/\tau_\Phi = 1/\tau_o + K_\tau \Phi, \qquad (3)$$

where τ_Φ , τ_o are lifetimes after and before irradiation with a fluence of nuclear particles Φ; K_τ is a lifetime damage constant (radiation constant). In general it is lifetime degradation that determines the radiation hardness of bipolar integrated circuits. The value of K_τ depends on type and energy of nuclear particle, semiconductor material, temperature and injection level.

Efficiency of majority carrier concentration induced by radiation degradation is characterized by carrier removal rate R_o:

$$R_o = -(dn/d\Phi), \tag{4}$$

The typical value of *Ro* is equal to (4...10) 1/cm under fast neutron irradiation, (0.1...10) 1/cm under electron irradiation and (10 ...100) 1/cm under proton irradiation. Effects of carrier removal determine the increase of semiconductor resistivity:

$$\rho_\Phi = \rho_o \cdot exp(-R_o \Phi / n_o); \tag{5}$$

where ρ_Φ , ρ_o are resistivities after and before irradiation, n_o is an initial concentration of free charge carriers.

2.3 Total ionizing dose effects

Total ionizing dose effects are induced by space protons and electrons and cause parametrical and functional failures of ICs. The most dose-sensitive devices are MOS and CCD. Typical failure thresholds are about 10…100 krad (Si). The main reason of these failures is concerned with electron-hole pair generation in dielectric layers. Positive charge trapping in gate or field oxides and generation of surface interface traps causes threshold voltage shifts in MOS devices, leakage current, *etc*. The inherent feature of total dose effects in space is low radiation intensity (down to 0.001 rad/s) together with temperature variations.

The basic picture of radiation effects in MOS devices has been formed at the moment (Ma and Dressendorfer, 1989). The ionizing radiation causes the generation of the electron-hole pairsin oxide. Some fraction of this pairs, which depends on the oxide field, irradiation type and temperature, recombines immediately after the creation. The electrons and holes, which avoided the initial recombination, are separated by electric field and transport through the oxide layer according to the oxide field direction. Electrons are much more mobile than holes in SiO_2 and they leave oxide in a very short times (about picoseconds). Holes move through oxide via hopping mechanism towards the interfaces. A hole can be trapped near the interface by the defect precursor (usually – the oxygen vacancy) and forms positively charged defect. It is E'-center, that has been shown to be the dominant

defect, which is responsible for radiation–induced positive charge creation. Trapped holes can be compensated ("annealed") by electrons from the oxide or silicon substrate. These processes extend from seconds to years and has a great importance for IC behavior prediction under a low dose rate.

Another process, which affects IC's radiation behavior, is interface traps generation. In contrary with the positive charge, interface traps are stable and do not anneal under room temperature. Moreover, the postirradiation interface traps build-up is often observed.

The long-term operation of MOS device under low dose rate irradiation is characterized by the several groups of basic effects: oxide fixed charge annealing; $Si - SiO_2$ interface traps build-up; latent charge relaxation; degradation of parasitic CMOS IC's structures; reliability parameters aggravation (especially device aging).

The radiation damage in silicon dioxide layers consists of three components: the build-up of trapped charge in oxide, an increase of the interface traps number and an increase of bulk oxide traps number. As a rule two kinds of radiation-induced charging have been observed: oxide trapped charge (ΔQ_{ot}) and interface trapped charge (ΔQ_{it}).

2.4 Single event effects

Single event effects (SEE) identify any observable effect in microcircuit that can attribute to local ionization of IC sensitive volume by primary or secondary nuclear particle. There are a lot of various SEEs (upset, latchup, burnout, rapture, etc.), but the most important are upset and latchup.

A single event upset (SEU) is a bit flip in a digital element that has been caused either by direct ionization from a traversing particle or by ionization produced by charged particles and recoiling nuclei emitted from a nuclear reaction. The most SEU sensitive ICs are digital very large scale integration (VLSI) circuits, especially, random access memories (RAM).

A single event latchup (SEL) is a low-impedance condition for parasitic p-n-p-n structure switching on induced by local electron-hole generation. Single event latchup (SEL) is one of the dominant failure effects of CMOS ICs under irradiation by nuclear particles with high energy. SEL can trigger large power supply currents, which can damage electronic parts.

SEE depends on an energy absorption in IC sensitive volume from a single nuclear particle. The main SEE sensitivity parameters are LET threshold and cross section of effect for ion radiation and proton threshold energy and cross section of the effect for proton radiation. Typical value of LET threshold is about 1…10 MeV·cm^2/mg and of saturation cross section is about 1…100 μm^2/bit for ion irradiation. The same values for proton irradiation are 15…30 MeV and 10^{-15}…10^{-12} cm^2/bit.

The usual procedure of SEE rate prediction consists of three steps: estimation of radiation environment; determination of IC SEE sensitivity parameters and SEE rate prediction. For heavy ion SEE rates, there are three primary methodologies (Tylka et al., 1997, Petersen, 1997, Pickel, 1996):

- Step function method;
- Effective Flux method;
- Right Rectangular Parallelepiped (RPP or IRPP method).

The Step Function method is based on the suggestion that $\sigma(LET)$ is step function with only two parameters: LET_o and σ_{is}. It is the easiest way in usage but has rather bad accuracy.

The ion flux in the environment $\Phi(LET)$ can be transformed to an effective flux $\Phi_e(LET)$ for an assumed cutoff angle in Effective Flux method. SEU rate R is determined from the equation (Petersen, 1997):

$$R = \int \Phi_e(LET_o)\, d\sigma(LET); \tag{6}$$

The most popular methods are PRR or IRPP . RPP method is based on model of sensitive volume as a right rectangular parallelepiped and relies on chord-length distributions to calculate the number of ion interactions that can induce upset (Tylka et al., 1997, Petersen, 1997):

$$R = 0.25 \cdot A \cdot N \cdot \int I(LET) \cdot f[E_o/(\rho \cdot LET)] \cdot dLET \tag{7}$$

where A is a surface area of the device, $I(LET)$ is an integral LET spectrum, f is differential path length distribution. IRPP method is similar to RPP one but it takes into account the fact that all sensitive volumes are not identical. Implementation of IRPP model requires simplifying assumptions and considerations for model parameters that are concerned with SEU cross section dependence on LET.

Usually, heavy-ion SEU rates are calculated in CREME96 software with the IRPP method (Tylka et al., 1997). To specify the cross-section one can supply Weibull function with four parameters:

$$\sigma_i(LET) = \sigma_{is} \cdot \{1 - exp[-((LET-LET_o)/W)^s]\}; \text{ for } LET > LET_o; \tag{8}$$

where W and s are approximation coefficients.

The proton SEUs are induced by secondary nuclear particles and their prediction methods are based on energy dependence of the proton upset cross section (Petersen, 1997, Pickel, 1996)

$$R = \int \varphi(E_p)\ \sigma_p(E_p) dE_p; \tag{9}$$

where R is a SEU rate; $\varphi(E_p)$ is an energetic proton flux; $\sigma_p(E_p)$ is a proton upset cross section; E_p is a proton energy. As a rule, the dependence of proton-induced SEU cross section energy is obtained from proton accelerator tests for different energies.

Two different approaches to proton-upset cross section calculations are developed now. The first one is concerned with nuclear reaction analysis in microvolume of IC elements (Miroshkin and Tverskoy, 1995). The semiemperical approach, based on Bendel model, is used in the second one (Bendel and Peterson, 1983). Each model needs only two parameters. As a rule, a prediction of SEU cross section dependence on proton energy needs a threshold energy (E_o) and a sensitive volume value (V_{sp}) in digital models.

Bendel approximation is based on the following expression:

$$\sigma_p = \sigma_{ps} \cdot [1 - \exp(-0.18 \cdot \sqrt{\sqrt{(18/A_p)}(E_p - A_p)})]^4; \quad (10)$$

where E_p is a proton energy; A_p, σ_{ps} are approximation coefficients; σ_{ps} is a proton-induced SEU saturation cross section; A_p is a threshold parameter.

3. TOTAL DOSE AND STRUCTURAL DAMAGE EFFECTS EXPERIMENTAL SIMULATION

3.1 General

Generally, structural damage effects of IC in space are caused by proton irradiation and they are important for failure simulation of bipolar ICs.

High-intensity α-source with about 1 Ci activity and 1 cm in diameter can be used for structural damage simulation. It should be mentioned, that defects arising under the impact of protons and α-particles are practically the same. Nevertheless, one should take into account the influence of total dose effects induced by the α-particle irradiation.

Total dose effects are induced by protons and electrons of space radiation environment and cause IC parametrical and functional failures. The inherent features of total dose effects in space are low radiation intensity (down to 0.001 R/s) together with the temperature variations.

3.2 Ground testing

The IC space radiation hardness estimation procedure is usually based on either IC ground testing on particle accelerators, nuclear reactors, Co^{60} source, *etc*. or their computer simulation. The typical structure of IC testing under charge particles from accelerator is shown in Fig.3. The main

problems of such experimental approach are inherent electromagnetic noises, large communication distance, induced radio-activity of device under test (DUT) materials (for particle energy more than 10…15 MeV), *etc*.

Most of IC failures under space irradiation are concerned with total dose effects. Total dose tests use the isotopic sources: Co^{60}, Cs^{137}, Sr^{90}. It is obvious that isotopic tests are simpler and cheaper than accelerator research but they are still not cheap enough.

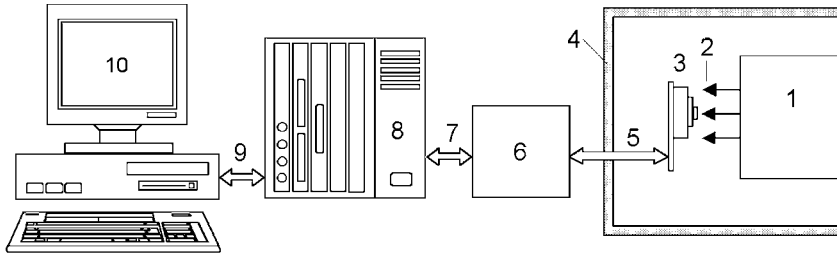

Figure 3. A typical structure of IC testing under electron or proton beam: 1 – accelerator and its particle beam (2), 3 - DUT, 4 - special shielding box, 5 - communication channel with buffer unit (6), 7 – communication with experimental room, 8 – specialized testing equipment, 9 - communication with PC (10).

3.3 X-ray simulator technique

The best quality in getting a low-cost test result can be obtained in experimental simulation tests with the usage of low-energy laboratory simulation sources which can provide the equivalence of IC radiation behaviors based on the basic radiation effects equivalence in nature and in laboratory.

The possibility of radiation simulator application is based on the similarity of main physical processes in semiconductor structures causing IC upsets and failures under the space environment and under simulation influence. As it has been already mentioned, the wide variety of space radiation types (electrons, protons, heavy charged particles, etc.) induces a limited number of dominant radiation effects only. This allows us to substitute each radiation source by an appropriate simulator, obtaining the same IC failure. It is important to choose the minimum necessary set of simulators, which would be sufficient to induce all basic space radiation effects (Chumakov et al., 1999).

The most convenient source for IC total dose tests is X-ray tester. The basic X-ray tester setup is an X-ray source with the average photon energy of 10 keV and with 40…60 keV maximum energy that provides dose rates in the range of 10 to 1000 rad(Si)/s.

IC total dose functional response may depend essentially on dose rate, the used functional test and postirradiation annealing time. The various sensitivity of intrinsic and peripheral elements to total dose make it very important to evaluate the individual upset levels of IC's different elements such as memory cells, bitline amplifiers, X- and Y- decoders, output drivers inside RAM. This information can be obtained from the local X-ray irradiation experiments using the special shielding (masks) (Chumakov and Yanenko, 1996). For example, the map of functional upsets of CMOS RAM after local x-ray irradiation is shown in Fig.4. The first upsets were concerned with memory cells failures and the "imprinting" effect was observed after local X-ray irradiation with total dose of 20 krad(Si).

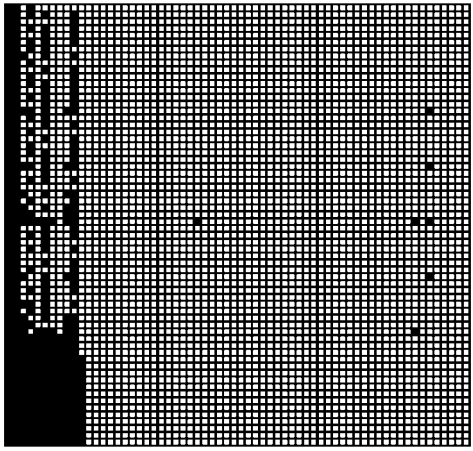

Figure 4. 4K x 1 CMOS SRAM error (black dots) bit map for the local X-ray irradiation (bottom left corner) for total dose D = 32 krad(Si).

In this case memory cell became a read-only-memory cell stored the bit, which had been written before irradiation. Then one can see failures of bitline amplifiers after total dose about 32 krad. After additional irradiation (D = 85 krad) SRAM functional upset due to failures of address decoder elements occurred.

3.4 Low intensity technique

The direct evaluation of IC's radiation hardness in low dose rate space radiation environment (about 0.1...0.001 rad(Si)/s) is practically impossible in laboratory conditions because of enormous time required, so various prediction procedures are under development. There are two main approaches to IC's low dose rate radiation response prediction nowadays.

The first one is based on accelerated total dose test procedure. The most popular accelerated total dose test procedure is specified by MIL-STD883 -

1019.5 and includes the moderate dose rate irradiation followed by 100 ^{o}C 168 hours annealing (Fleetwood et al., 1991). This method does not deal with any specific property of the device-under-test, it is rather simple in practice and is declared to be usable for hardness assurance for all CMOS IC in low and moderate dose rate environment. From the other hand, the procedure is excessively conservative respecting to the low dose rate environment that may lead to hardness underestimation which may be crucial for commercial IC's. Besides, the unified accelerated test conditions (annealing temperature and duration) do not guarantee the complete relaxation of radiation-induced charge within the specified time period, especially taking into consideration the latent processes.

The second approach to IC's low dose rate response prediction includes the evaluation of MOS or bipolar electrical model degradation parameters and specifying of irradiation mode to simulate low dose irradiation effects. In order to simulate low dose rate irradiation conditions the well-known cycled multi-step "irradiation-annealing" technique is used, which is based on irradiation and annealing step-by-step procedure (Sogoyan et al., 1999, Belyakov et al., 2000). Sometimes the technique “radiation-time relaxation” is applied as well (Chumakov and Yanenko, 1996).

The parameters of multi-step "irradiation-annealing" technique (step duration and annealing temperature) can be determined according to models of the device radiation degradation (Sogoyn et al., 1999). So, in order to simulate CMOS IC's behavior under irradiation with time-dependent dose rate $P_2(t)$ and at temperature $T_2(t)$ it is necessary to determine the annealing time and temperature. The annealing process at step i is to be discontinued at the moment t_{ai} (Fig.5) when the shift of n-channel MOSFET threshold voltage under laboratory irradiation is equal to the threshold voltage shift of the same device irradiated to the same total dose but with a lower dose rate P_2. The accelerated test conditions depend on the oxide charge relaxation and the interface traps build-up activation energies.

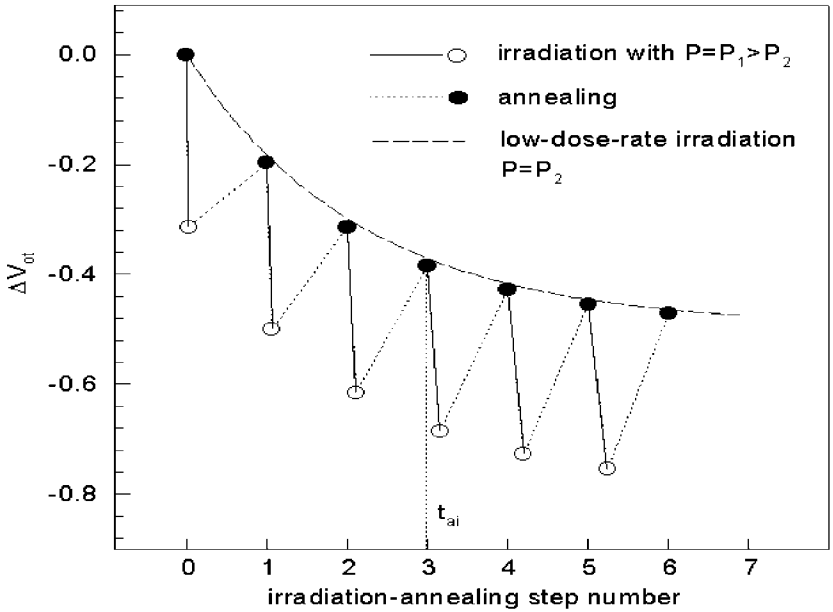

Figure 5. The multi-step "irradiation-annealing" technique to predict low dose rate effects

4. SINGLE EVENT EFFECTS

4.1 Ion-induced cross section

Usually, parameters for Weibull function (13) are obtained from ion accelerator tests for different linear energy transfers (LET) (LET typical range is from 1 MeV·mg/cm^2 to 50 MeV·mg/cm^2)) and for Bendel approximation (15) - from proton accelerator tests for different proton energy (proton energy typical range from 20 MeV to 300 MeV). The typical structure of IC experimental setup under charged particles influence has not got an essential difference as compared to the setup shown in Fig.5.

Ion-beam and/or proton beam experiments are the direct procedure for determining SEE sensitivity parameters. The typical technique for these experiments is to irradiate IC with constant energy and to determine the number of SEE – N_{see}. (functional errors, latchup events, hard errors etc.). Cross section of SEE can be calculated from those tests as follows:

$$\sigma_{see} = N_{see} / \Phi(E_{cp}) \tag{11}$$

where $\Phi(E_{cp})$ is a total fluence of charged particles with energy E_{cp}. Cross section is determined for proton energy and for ion LET. Parameters for Weibull function are obtained from the ion accelerator tests for different

linear energy transfers (minimum 4 experimental data points) and for Bendel function – from the proton tests (minimum 2 experimental data points).

The additional difficulties in accelerator tests should be mentioned – the necessity to remove IC package cups, to locate DUT in vacuum chamber, to make functional testing in real time and under the radiation influence, etc. It should be pointed out that accelerator testing is rather expensive (about $50 000 per part type (Dodd, 1996)).

So it is preferable to predict SEE sensitivity applying the minimum experimental data. The satisfactory results for ion-induced SEUs are obtained using two independent SEU parameters of IC (LET_o and σ_{is}) and supposing that s = 1.5 and W depends on LET_o in peripheral area of sensitive volume of IC elements (Chumakov and Tverskoy, 2001). Only ion-induced saturation cross section σ_{is} and threshold LET LET_o are used in this case. The possible approximation for $W(LET_o, \sigma_{is})$ is as follows:

$$W = k_w \cdot LET_o^{\chi} \cdot [1 + k_p \cdot \sigma_{is1}^{\beta}]; \tag{12}$$

where σ_{is1} is an ion-induced SEU saturation cross section per one bit; k_w, χ, k_p and β are approximation coefficients. Only the ion-induced cross section and threshold LET_o are used in this case. The satisfactory agreement between the two-parameter function and usual Weibull curves is obtained (Fig.6).

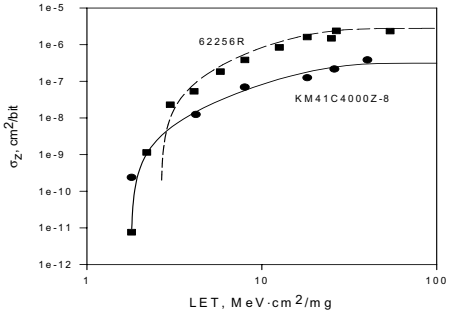

Figure 6. Ion-induced SEU cross sections vs. LET based on σ_{is} and σ_{ps}: symbols are data from (Calvel et al., 1996); curves – approximation

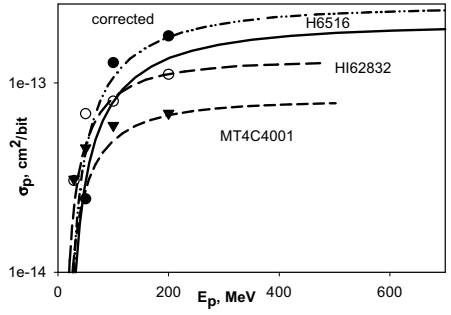

Figure 7. Proton-induced SEU cross sections vs. proton energy based on σ_{is} and σ_{ps}: symbols are data from (Calvel et al., 1996); curves - approximation

Therefore two-parameter curves can be used to evaluate the ion-induced SEU cross section dependence on LET.

4.2 Proton-induced cross section

Proton-upset cross section calculations can be obtained from ion-induced SEU experimental data using nuclear reaction analysis in microvolume of IC

elements (Chumakov, 2002). Two sensitive parameters (a sensitive volume V_{sp} and a threshold energy E_o) are used in this case. IC element threshold energy E_o (a critical charge) can be estimated from LET_o:

$$E_o = LET_o \cdot d_{si}; \tag{13}$$

where d_{si} is an effective thickness of the IC sensitive volume and its value is about 2…10 μm.

A sensitive volume (RPP approach) from secondary nuclei can be estimated as:

$$V_{sp} \cong \sigma_{is}\, d_{sp}; \tag{14}$$

where d_{sp} is a thickness of sensitive volume. Taking into account the funneling effects one can consider that $d_{si} = (1...6) \cdot d_{sp}$.

It is clear that there is a connection between physical and Bendel parameters:

$$A_o \approx 1.09\ (E_o + 1.78)^{1.27}. \tag{15}$$

The connection between σ_{ps} and V_{sp} can be obtained from the following equation (Chumakov and Tverskoy, 2001):

$$\sigma_{ps} = k_{sp}\, V_{sp}^{(1+Eo/E1)}\ exp(-E_o/E_2); \tag{16}$$

where k_{sp}, E_1 and E_2 are approximation coefficients. A satisfactory agreement of approximation curves with digital calculation curves was obtained for k_{sp} = 2.5 10^{-14} cm^2; E_1 = 18.3 MeV and E_2 = 1.43 MeV.

One can see there is the correlation between two sensitive parameters for ion-induced SEU (LET_o and σ_{is}) and proton-induced SEU (A_p and σ_{ps}). Thus, one can use only two experimental data to estimate SEU parameters sensitivity of IC. The best approach is based on usage of two values of saturation cross section (σ_{ps} and σ_{is}). Therefore, only two experimentally determined values of σ_{ps} and σ_{is} are required to estimate the ion- and proton-induced SEU cross section dependences (Fig.6 and Fig.7). Heavy ion saturation cross section for most of ICs can be obtained using Cf^{252} tests.

4.3 Experimental simulation technique

It is possible to use the following simulators to obtain experimental SEE data:

a) low intensity α-source for investigations of IC SEUs due to alpha and other charged particles with relatively low linear energy transfer (LET);

b) Cf^{252} isotopic source for investigations of IC SEUs and SELs due to heavy charged particles;

c) focused picosecond laser source for both SEUs and SELs investigation with the possibility to irradiate the chosen IC's elements;

d) non-focused nanosecond laser source for latchup sensitivity evaluation.

The Cf^{252} source is a simple way to screen SEE effects. The relative high spectrum produced by heavy nuclear particles – products of radioactive spontaneous fission fragments (LET is about 42 MeV·cm^2/mg) can give a short way to estimate SEE sensitivity of IC.

There are limitations inherent to Cf^{252} simulation method:

Cf^{252} emits a lot of α-particles with energy about 6 MeV;

The penetration depth of fission fragments are not large enough (about 10…12 μm);

LET can be changed in very limited range;

fission fragments have large energy range 40…120 MeV;

there is a distribution in ion number of fission fragments.

Because of α-particles, which can cause SEE as well, at the first stage it is necessary to carry out IC testing using low intensity α-source. Because of small penetration length of fission fragments the test result may be false if DUT contains surface protective layers over sensitive volumes. Limited range of LET does not allow to cover cross section versus LET. Energy and ion number distribution complicates getting the experimental results. Overall, this simulator is good for screening purposes and allows to estimate ion-induced saturation cross section.

Picosecond focused laser source (with laser spot radius r_p of less than 5 μm) seems to be the most universal laboratory tool for SEE investigation, since one can use it to simulate a wide range of particle LET and penetration depths varying the laser beam spot diameter and the wavelength (Buecher, et al., 1988). Unfortunately, laser radiation does not ionize dielectric structure, thus SER can not be investigated in laser testing.

There are certain restrictions and difficulties in applying of focused beam laser simulators to SEE investigation:

different distributions of the generated free carriers under single ion and focused laser beam impacts;

optical effects (reflection, diffraction, absorption in polysilicon layers, etc.) leading to discrepancies between the incident laser energy and LET;

simulation adequacy suffers due to metal strip shadowing.

Nevertheless, laser simulators are very useful tools in IC response investigation in space radiation environment.

The typical structure of IC testing under focused laser beam is shown in Fig.8 (Chumakov et al., 1997).

The difference of SEE threshold energy for two wavelengths is determined by the dependence of laser radiation absorption factor vs. wavelength: with the decrease of wavelength from 1.06μm to 0.53μm the absorption factor in Si increases from ~14 to ~7000 cm^{-1}. Thus, the SEE sensitivity for $\lambda = 1.06$ μm is characterized by LET_{o_th} while for $\mu = 0.53$μm - by total absorbed energy (or charge) threshold E_{o_th}.

The SEE thresholds correspondence at various wavelengths and the dependence of laser energy on LET and absorbed energy in Si (expressions 1, 2) gives us a possibility to determine the latchup effective charge collection length L_{th} (Chumakov et al., 1997):

$$L_{tsee}=E_{o\_th}/LET_{o\_th}) \approx 0.5\cdot(E_{0.53}/E_{1.06}) \cdot(1/\alpha_{1.06}). \quad (17)$$

For example, the typical value of L_{seu} for SEU is about 5 μm and L_{seu} ≈10…20 μm for SEL.

Laser tests limitations can be essentially reduced if so called "local" laser irradiation (with laser spot radius r_p of 10…100 μm) is used instead of "focused" one (Chumakov et al., 2002). However, new problems arise, namely the difference in charge collection processes under local and focused laser irradiation. In fact, charge generated by focused laser beam is collected by drift (funneling mechanism) in sensitive volume, while diffusion mechanism of charge collection can be essential under local laser irradiation. Moreover, in latter case charge collection by p-n structures adjacent to sensitive volume affects the upset threshold.

The correlation between "local" and "focused" laser irradiation cases the upset threshold comparative tests and software simulation research have been performed for various laser pulse duration, laser spot radius and structural parameters of IC elements.

Two types of laser simulators (“PICO-2E” and “RADON-5E”) were used as radiation sources. The first one generated picosecond pulses with duration $T_p \approx 10$ ps and wavelength $\lambda = 1.055$ μm. “RADON-5E” was used to generate nanosecond laser pulses with $T_p = 12$ ns and $\lambda = 1.079$ μm.

Upset threshold laser energy as a function of spot diameter in CMOS RAM 537RU6 (4K x 1) is shown in Fig.9 for nanosecond laser irradiation. Curve 1 corresponds to the case when the centre of laser spot is located at the most sensitive area of the memory cell storing “1”. Curve 2 is similar to previous one but for the opposite logic state “0” of memory cell. Curve 3 was obtained for logic state "0" of memory cell (same as for curve 2) and for location of laser spot in the centre of memory cell.

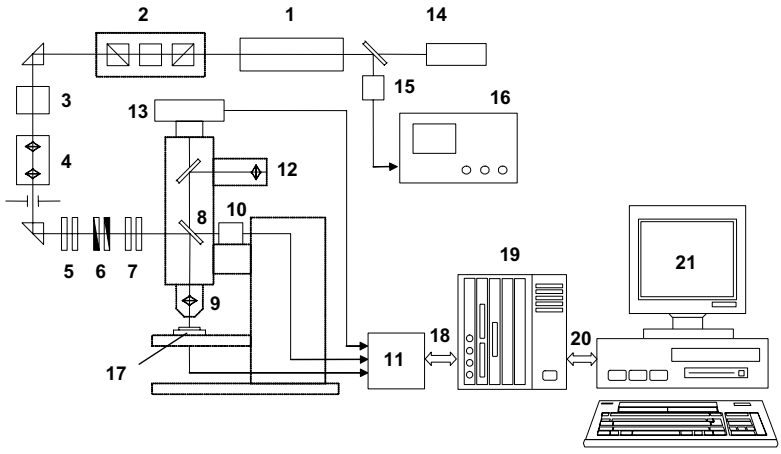

Figure 8. Focused picosecond laser simulating system for SEE in IC: 1 – picosecond laser, 2 – electro-optical modulator, 3 - second harmonic generator, 4 – unit for Gaussian beam profile, 5…7 – beam attenuation unit (set of neutral filters was used for rough (5) and precise (6) intensity attenuation, 7 - spectral filters), 8…13 focusing unit (9 – micro-objective, 10 - laser energy detector, 11 – buffer unit, 12 - ocular 13-Couple Charge Device (CCD) camera), 14 - laser for beam visualization, 15 – laser beam detector, 16 - fast oscilloscope, 17 – DUT, 18 – communication bus with specialized testing equipment (19), 20 - PC communication bus(21).

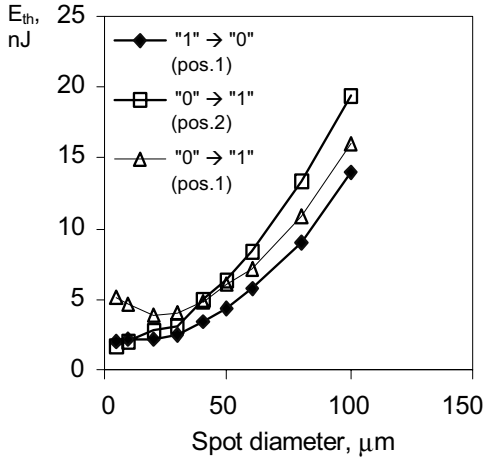

Figure 9. Threshold laser energy as a function of laser spot diameter for 537RU6.

One can see that curves 1 and 2 are practically constant for small laser spot diameters (less than 25 μm). These data correspond to the case when spot area is less than SEU saturation cross section. For the curve 3 the

decrease of laser energy for smaller spot diameters is observed, because the laser spot is located relatively far away from the most sensitive area. In spite of these differences, threshold energies are practically the same for relatively large laser spot diameters. The qualitatively similar results were obtained for local picosecond laser tests.

Thus, non-focused nanosecond laser simulator is very useful in SEE research as well. The main advantage of the non-focused nanosecond laser simulator usage with local radiation is the possibility to estimate SEE threshold without laser scanning of IC surface. However, it is necessary to correct the experimental results taking into account the radius of laser spot and time duration of laser pulse.

5. CONCLUSIONS

The dominant radiation effects under space environment resulting in IC damage and upsets are total dose (surface ionization) and single event effects. Usual approach in predicting of the IC radiation hardness is rather expensive and has limitations for VLSI devices. Laboratory radiation simulator setups are developed for prediction of IC response to space radiation, taking into account total dose and single event effects as well as structural damage. Experimental simulation methods for estimating of IC space radiation hardness are based on possibility of simulators to cause the same dominant radiation effects in sensitive volume of IC elements as compared to the real space radiation. The minimum necessary set of simulators, which is sufficient for investigation of IC failures and upsets under space radiation influence, is proposed. For silicon devices we suggest to simulate the total dose effects - with an X-ray tester; the single event effects (upsets and latchup) - with a combination of Cf^{252} and low-intensity α-sources, picosecond focused and nanosecond nonfocused lasers.

6. ACKNOWLEDGEMENTS

We are very grateful to A. Artamonov, A. Egorov, O. Kalashikov, O. Mavritsky, A. Sogoyan and A. Yanenko at SPELS for their considerable efforts related to the development of space radiation simulators and techniques described here.

7. REFERENCES

Belyakov V.V., Chumakov A.I., Nikiforov A.Y., Pershenkov V.S., Skorobogatov P.K., Sogoyan A.V. IC's Radiation Effects Modeling and Estimation Microelectronics Reliability, v.40, No.12, pp.1997-2018, 2000.

Bendel W.L., Petersen E.D. Proton Upsets in Orbit. IEEE Trans. on Nuclear Science, Vol. 30, No.6, pp.4481-4485, 1983.

Buecher S.P.et. al. Charge collection from focuced picosecond laser pulses. IEEE Trans on Nucl.Sci. Vol.35, No.6, pp.1517-1522, 1988.

Calvel P., Barillot C., Lamothe P., Ecoffer R., Duzellier S., Falguere D.. An Empirical Model for Predicting Proton Induced Upset, IEEE Trans. on Nuclear Science, Vol.44, No 6, pp.2827-2832, 1996

Chumakov A.I. Simplified Estimation of Ion- and Proton-Induced SEU Sensitivity parameters using one/two experimental data. "Proc. of the 5th International Workshop on Radiation Effects on Semiconductor Devices for Space Application. Takasaki, pp.137-140, 2002.

Chumakov A.I., Nikiforov A.Y., Mavritsky O.B., Egorov A.N., Yanenko A.V. Comparison of IC upset thresholds under focused, Local and uniform Laser Irradiation. "Proc. of the 5th International Workshop on Radiation Effects on Semiconductor Devices for Space Application. Takasaki, pp.141-144, 2002

Chumakov A.I., Nikiforov A.Y., Mavritsky O.B., Egorov A.N., Yanenko A.V. Single Event Latchup Threshold Estimation Based on Laser Dose Rate Test Results. IEEE Trans. on Nucl. Sci. Vol.4, No.6, pp. 2034-2039,1997.

Chumakov A.I., Nikiforov A.Y., Telets V.A., Gerasimov V.F., Yanenko A.V., Sogoyan A.V. IC Space Radiation Effects Experimental Simulation and Estimation Methods. Radiation Measurements, vol. 30, pp.547-552, 1999.

Chumakov A.I., Tverskoy M. G. Estimation of ion- and proton-induced SEU rate by two values of saturation cross sections. Sixth European Conference on Radiation and Its Effects on Components and Systems. RADECS 2001. Grenoble

Chumakov A.I., Yanenko A.V. Radiation and Postirradiation Functional Upsets in CMOS SRAM. IEEE Trans. on Nucl. Sci. Vol.43, No.6, pp.3109-3114, 1996.

Dodd P.E. Device simulation of charge collection and single event upset. IEEE Transactions on Nuclear Science. Vol. 43, No. 2, pp. 561-575, 1996.

Fleetwood D.M., Winokur P.S., Meisenheimer T.L. Hardness Assurance for Low-Dose Space Applications. IEEE Trans. Nucl. Sci., Vol.38, No.6, pp. 1560-1565, 1991.

Hilmer R. V. (Ed.). AF-GEOSPACE user's manual version 2.0/Ed. by Air Force Research Laboratory, 2001.

Holmes-Siedle A., Adams L. Handbook of Radiation Effects. - N.Y.: Oxford Unifersity Press, 1993. - 479 p.

Ma T.P., Dressendorfer P.V. (Ed). Ionizing radiation affects in MOS devices and circuits. - N.Y.: John Wiley and Sons, 1989. - 588 p.

Miroshkin V.V., Tverskoy M.G. Some Aspects of Application of Two Parameter SEU Model. IEEE Trans. on Nuclear Science, Vol.42, No.6, pp. 1809-1814, 1995.

Nikiforov A.Y., Telets V.A., Chumakov A.I. Radiation effects in CMOS integrated circuits. Moscow, Radio and Svyaz, 1994.- 180 p.

Petersen E.D. Single-Event Analysis and Prediction. 1997 IEEE Nuclear and Space Radiation Effects. Short Course. Applying Computer Simulation Tools to Radiation Effects Problems. Snowmass Conference Center. Snowmass Village, Colorado. pp.III-1 - III-160, 1997.

Pickel J.C. Single-Event Effects Rate Prediction. IEEE Transactions on Nuclear Science. Vol. 43, No. 2, pp. 483-495, 1996.

Sogoyan A.V., Nikiforov A.Y., Chumakov A.I. Approach to predict CMOS IC radiation degradation parameters taking into account operation conditions and duration. Rus. Microelectronics. Vol.28, No.4, pp.263-275, 1999.

Summers G.P. Correlation of particles induced displacement damage in silicon. IEEE Trans. on Nucl. Sci. Vol.34, No.6, pp. 1135-1139, 1987.

Tylka A.J., Adams J.H., Boberg P.R., Brownstein B., Dietrich W.F., Flueckiger E.O., Petersen E.L., Shea M.A., Smart D.F., Smith E.C.. CREME96: A Revision of the Cosmic Ray Effects on Micro-Electronics Code. IEEE Trans. on Nuclear Science, Vol. 44, No.6, pp.2150-2160, 1997.

http://creme96.nrl.mil/

APPENDIX

List of abbreviations

CCD	Charge-Coupled Device
CMOS	Complementary Metal-Oxide-Semiconductor
DUT	Device Under Test
IC	Integrated Circuit
IRPP	Integral Right Rectangular Parallelepiped
LET	Linear Energy Transfer
MOS	Metal-Oxide-Semiconductor
PRA	Primary Recoil Atom
RAM	Random Access Memory
PC	Personal Computer
RPP	Right Rectangular Parallelepiped
SEE	Single Event Effect
SEL	Single Event Latchup
SEU	Single Event Upset
SRAM	Static Random Access Memory
VLSI	Very Large Scale Integration

List of symbols

A	surface area of the device
E_d.	displacement threshold energy
E	nuclear particle energy
E_o	threshold energy
E_n, E_p	neutron and proton energies
D	absorved total dose
d_{si}	effective thickness of the IC sensitive volume
f	differential path length distribution function
$I(LET)$	integral let spectrum
k_{str}	non-ionizing energy losses
K_τ	lifetime damage constant (radiation constant)
LET_O	threshold value of LET
n_o I	nitial concentration of free charge carriers
R_o	initial carrier removal rate
R	SEU rate
r_p	laser spot radius
T_p	pulse duration
V_{sp}	sensitive IC element volume
ΔQ_{it}	interface trapped charge
ΔQ_{ot}	oxide trapped charge
Φ	fluence of nuclear particles
Φ_n	threshold neutron fluence
$\Phi(LET)$	integral LET spectral ion flux
λ	laser radiation wavelength
ρ_Φ, ρ_o	resistivities after and before irradiation
$\sigma(LET)$	function of IC cross section Vs. LET
σ_{IS}.	ion-induced saturation cross section
$\sigma_p(E_p)$	proton upset cross section
σ_{PS}.	proton-induced saturation cross section
τ_Φ , τ_o	lifetimes after and before irradiation

Chapter 10

The Effects of Space Weather on Radio Systems

with a focus on HF systems

P. S. Cannon*, M. J. Angling, J. A. T. Heaton, N. C. Rogers, A. K. Shukla
Centre for RF Propagation and Atmospheric Research, QinetiQ
Worcs,WR14 3PS, UK

Also at Dept of Electronic and Electrical Engineering, University of Bath, Bath, BA2 7AY, UK

Abstract If the environment were isotropic and stable in time, it would be relatively easy to determine its effects on the propagation of RF waves. Unfortunately, this is not the case. The spatial scales of inhomogenities vary from thousands of kilometres to turbulence with scale sizes of a less than a metre. Likewise the temporal scales vary over many orders of magnitude from many years (solar cycle effects on ionospheric propagation) to hours or even minutes (the scale of weather phenomena). As a consequence of this variability, timely and reliable strategies are required to both specify and accurately forecast the environment and to assess the attendant impact on the operational performance of the RF systems. These strategies can be used to automatically apply corrections to the system operating parameters or advise the user on a course of action that will improve the system functionality. This paper will review these phenomena and describe their impact on a number of systems. The effect that space weather has on high frequency (HF) systems will be addressed in some detail and strategies for mitigating its effects will be described.

Keywords Radio propagation, total electron content, HF communications, ionosphere, space weather

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 185-201.

1. INTRODUCTION

The ionosphere, an area of the atmosphere which extends from ~80 to ~1000 km, can significantly affect the propagation of radio frequency (RF) signals which pass through it or are reflected by it [*Cannon*, 1994a; *Cannon*, 1994b]. The effects are varied but include absorption, refraction, retardation and scintillation. At frequencies above ~1 MHz, the lower *D* region causes absorption and the higher *E* and *F* regions cause a variety of other effects. These effects, which include refraction, signal group delay, signal phase advance, pulse broadening and Faraday rotation of the polarisation vector, all follow an inverse power law and are generally only significant up to a frequency of ~2 GHz. Below ~1 MHz radio systems bounce their signals from the tenuous *D* region; consequently although the height of the layer is important for system operation, absorption is not an issue.

The diverse set of affected systems (Figure 1) include ground-ground high frequency (HF) communications, ground-space communications, GNSS (Global Navigation Satellite Systems), such as GPS (Global Positioning System) and Galileo – particularly single-frequency systems, HF over-the-horizon radars, satellite altimeters and space-based radars [*Goodman and Aarons*, 1990]. HF communications and radar systems rely on the ionosphere for their operation but also have to contend with its effects. Most other systems are degraded by the ionosphere and its variability but for certain specialist applications this variability can be exploited. Loss of phase lock and range errors in GNSS are examples of such deleterious effects.

If the ionospheric environment were isotropic and stable in time, it would be relatively easy to determine its effects on the propagation of RF waves. Unfortunately, this is not the case. The spatial scales of inhomogeneities vary from thousands of kilometres to turbulence with scale sizes of a less than a metre. Likewise, the temporal scales vary over many orders of magnitude from many years (solar cycle effects on ionospheric propagation) to hours or even minutes (the scale of weather phenomena). As a consequence of this variability, timely and reliable strategies are required to both specify and accurately forecast the environment and to assess the attendant impact on the operational performance of the RF systems. These strategies can be used to automatically apply corrections to a systems operating parameters or, via a decision aid, advise the user on a course of action that will improve a systems functionality [*Cannon et al.*, 1997].

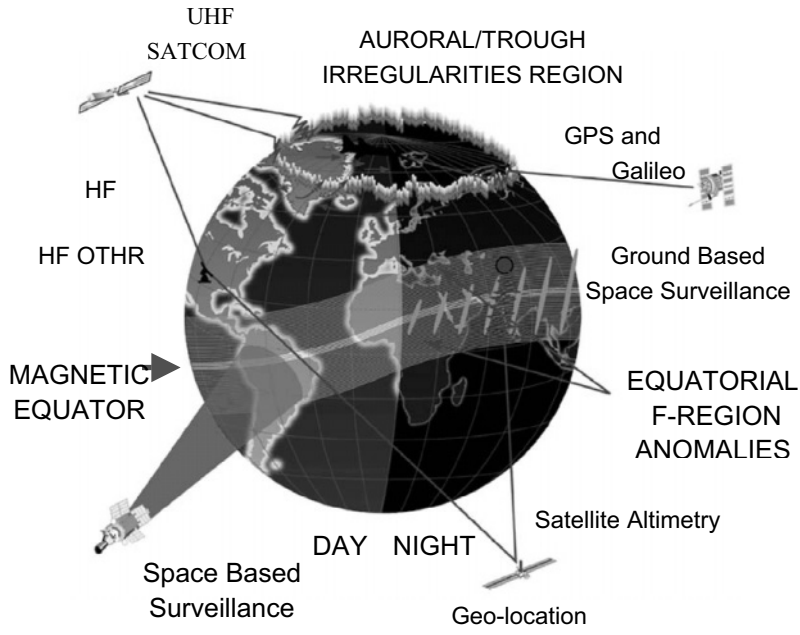

Figure 1. Systems affected by the ionosphere

2. IONOSPHERIC VARIABILITY

2.1 Naturally occurring variability

Natural variability can be categorized into that due to bulk (average) effects and that due to small-scale irregularities with sizes less than a Fresnel radius. The bulk effects, including variations in the refraction and time delay of signals propagating through the ionosphere, can cause varying areas of coverage in HF systems. At UHF they introduce errors in radar, altimetry, geolocation and space-based navigation systems (~30m). Single frequency GNSS transmits model coefficients to mitigate these errors but it is unable to fully compensate for the day-to-day and hour-to-hour variations. Dual frequency GNSS can use the different delays of the two transmitted frequencies to calculate the total electron content (TEC) and make an almost exact correction. Irregularities in the plasma density cause signal scintillation [*Aarons*, 1982]; i.e. random variations in amplitude and phase. These are generally quantified in terms of the standard deviation of phase, σ_φ, and the standard deviation of the signal power, normalised to the average received

power, *S4*. Scintillation can cause data errors in communications, loss of coherence in radar applications and loss of signal lock in GNSS navigation. At the peak of the sunspot cycle fade depths of 30 dB are possible at 400 MHz and 20 dB fades are possible at 1 GHz. Phase changes also limit the ability of synthetic aperture radar (SAR) systems to form a phase coherent aperture and so reduce resolution.

2.2 Artificially induced variability

The ionosphere is a naturally occurring plasma environment, which can be artificially modified by a number of techniques. Modification of the ionosphere by the release of large volumes of chemically reactive gases is one such technique. This possibility was first recognised by the discovery that rocket exhausts deplete the local charged particle density. A second technique makes use of charged particle accelerators both to modify ionospheric properties and create artificial auroras. A third technique uses VLF radiation generated on the ground to stimulate instabilities in the magnetospheric plasma that in turn generate hydromagnetic emissions and cause particle precipitation. This technique arose from the discovery that ground based power transmission lines could affect the particle distribution in the magnetosphere. A fourth technique uses high power ground based transmitters at LF or higher frequencies both to modify the ionosphere and to generate secondary radio emissions. For purpose built facilities the effective isotropic radiated power (EIRP) can be as high as 80-90 dBW. An example of unintended artificial modification of the ionosphere is reported in [*Cannon*, 1982].

3. SCOPE OF PAPER

The previous sections have discussed the impact that the ionosphere has on a broad range of systems. A discussion of the relevant space weather research and development being undertaken to support all of these systems is outside of the scope of this paper. The remainder of this paper will consequently address in more detail the effect that space weather has on high frequency (HF) systems - with an emphasis on work being undertaken in our own laboratory.

4. HF SYSTEMS

4.1 Median models

To ensure circuit reliability, an HF communications operator or provider is assigned a number of frequencies to counter the effects of the day-to-day variability of the ionosphere. The choice of the correct frequency is fundamental to maintaining acceptable communications. The successful selection of the most appropriate frequency depends upon the ability to predict and respond to the prevailing ionospheric conditions.

Space weather affects the operation of HF systems in many ways and to a first approximation this can be dealt with using climatological (monthly median) models of the electron density height profile, such as that described by [*Bradley and Dudeney*, 1973], coupled with a propagation model. In these climatological approaches a simple virtual mirror propagation model is assumed. These concepts are formalized in the secant law, Breit and Tuve's theorem and Martyn's equivalent path theorem, e.g [*McNamara*, 1991]. There are a number of such HF propagation tools available to aid the system manager (or operator). They require as input the geographic co-ordinates of the transmitter and receiver, the level of solar activity, the month and time, and the background noise level at the receiver. Additional inputs may consist of the planetary magnetic index (k_p), the transmitter power, the bandwidth, and the required circuit reliability. Example monthly median propagation tools are: ICEPAC [*Stewart and Hand*, 1994], IONCAP [*Teters et al.*, 1983], VOACAP [*Lane et al.*, 1993], MINIMUF [*Rose et al.*, 1978] and REC533 [*ITU-R*, 1999].

Monthly median models to predict the performance of HF systems were first introduced over 40 years ago and there has been steady progress improving the core ionospheric models [*Bilitza*, 2002]. Examples include NeQuick [*Radicella and Leitinger*, 2001], PIM [*Daniell et al.*, 1995], and various upgrades to the International Reference Ionosphere, IRI [*Bilitza*, 2000].

4.2 Ionospheric variability in the context of HF propagation

Figure 2 shows the maximum usable frequency (MUF) measured by an IRIS oblique ionosonde [*Arthur et al.*, 1997] on a 400 km path from Inskip (53.9ºN, 2.8ºW) to Malvern (52.1°N, 2.3°W), in the UK, over a period of 13 days in October 2002. This period coincided with a period of variable geomagnetic activity. The selected oblique ionograms from the data sets were examined and the MUF of the *F* layer trace was extracted using IRIS

analysis software [*Heaton*, 1999]. Figure 3 shows the minimum and maximum measured MUF alongside the predicted value from REC 533. This period was moderately disturbed geomagnetically with *kp* lying between three and five. Both figures illustrate the day-to-day variability in the MUF with values changing by several megahertz. There is a notable drop in the MUF on the 2nd and 3rd of November, corresponding to increased geomagnetic activity.

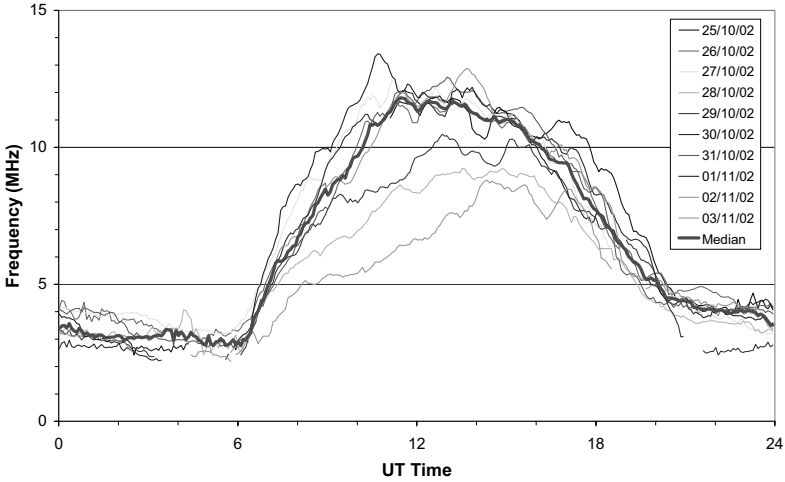

Figure 2: Maximum Usable Frequency as a function of time measured by the IRIS sounder for the Inskip to Malvern path for 25 October to 3 November 2002.

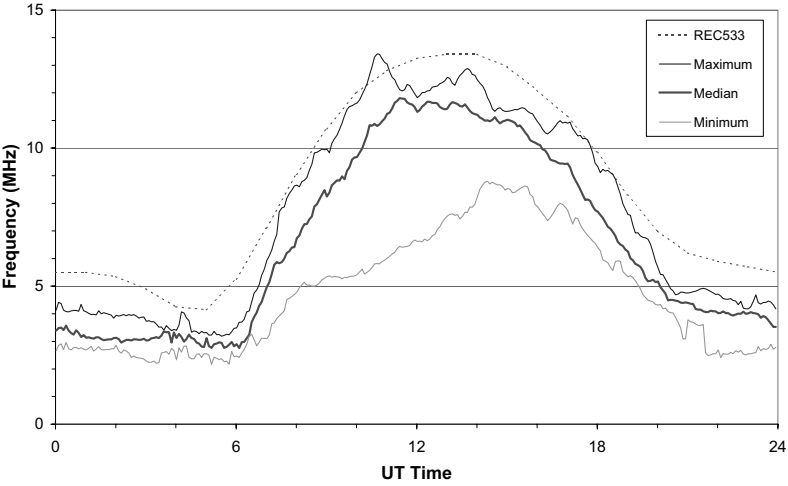

Figure 3. Minimum, median and maximum measured MUF (solid lines) determined from Figure 2, along with the median MUF determined by REC533 (dotted line).

4.3 A simple strategy for ionospheric model updating

We have seen that the MUF of a BLOS HF communications link can be predicted using HF communications prediction programs such as VOACAP and REC533. However, we have also seen that values of the predicted MUF may depart considerably (e.g. up to 50%) from the measured MUF.

A number of simple strategies have also been developed for dealing with the short-term variability in the context of HF propagation. Examples can be found in such propagation tools as: HF-EEMS [*Shukla et al.*, 1997], OPSEND [*Bishop et al.*, 2002], the Teleplan HF frequency management tool (Norwegian Army) and WinHF. Varying levels of sophistication are embodied in these techniques and others are still in development.

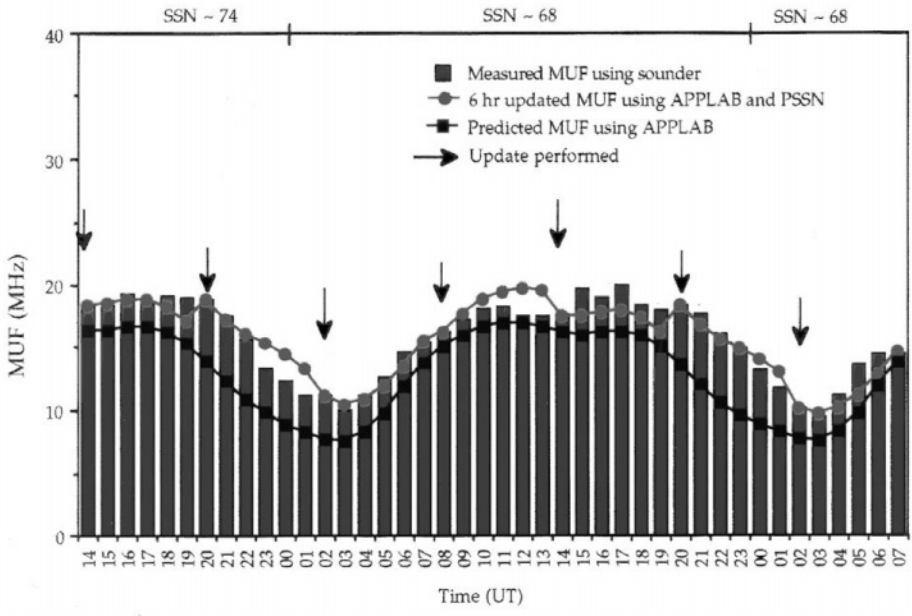

Figure 4: Six hourly updates of the Andoya (Norway) to Cobbett Hill path using PSSNs for 27 to 29 April 1993.

A simple and effective approach to dealing with space weather variability in the context of HF involves comparing measured (e.g. using sounders) and predicted (e.g. using REC533) MUF values. The prediction program input parameters (e.g. sunspot number) are then varied until the predicted MUF agrees with the measured MUF value. The prediction program then uses these best-fit input parameters, termed pseudo-parameters, until the next update is performed. Whilst this technique has been successfully applied to

the mid-latitude (and often benign) ionosphere, the technique is of limited use when applied to high latitude paths.

The limitations of temporal updating have been investigated [*Shukla and Cannon*, 1994]. An oblique sounder was used to make MUF measurements on a path from Andoya (in northern Norway) to Cobbett Hill (south of London, UK) from 27 to 29 April 1993. These measurements were used to update the APPLAB (similar to REC 533) propagation mode every six hours and every 12 hours, Figure 4.The predictions compare more favourably with the measurements when the model is updated more frequently, ie every six hours. Calculations show that 12 hourly updated MUF predictions indicate an improvement of ~38% compared with ~63% MUF prediction improvement observed using 6 hourly updates.

4.4 Adaptive techniques

Another strategy to cope with ionospheric variability due to space weather is to use adaptive systems, such as automatic link establishment (ALE) or automatic radio control systems (ARCS), that can automatically select and change frequency as appropriate. [*Jodalen et al.*, 2001] has used oblique data from the Doppler And Multipath Sounding Network (DAMSON) sounder and modem performance characterizations [*ITU*, 2000], coupled with the ICEPAC HF prediction code [*Stewart and Hand*, 1994], to determine the optimal number of allocated frequencies required by an adaptive HF system. Optimal was considered to be the minimum number of frequencies to achieve maximum reliability. On a 2000 km path it was shown that 5-6 frequencies were sufficient whereas on a short 200 km path 3-4 frequencies were sufficient. Figure 5 shows an example from this study.

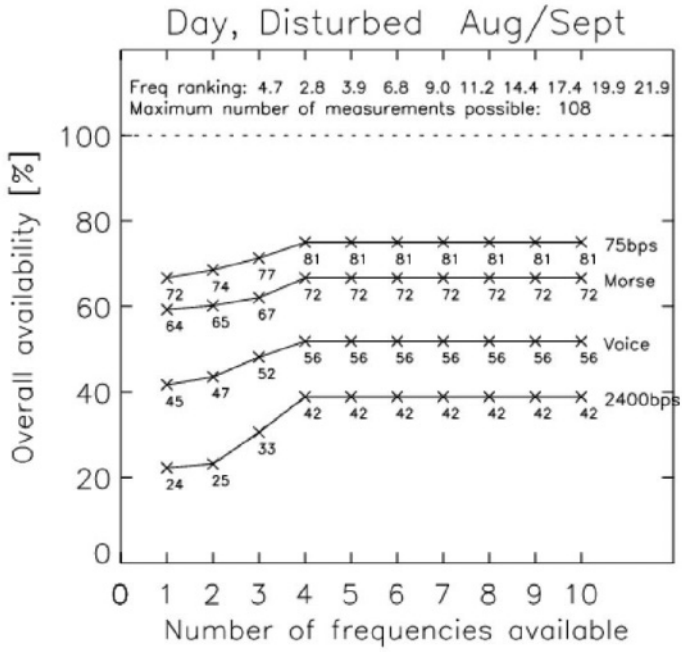

Figure 5. Overall availability of modems when the frequency set consists of 1,2,…..10 frequencies a propagation path from Harstad (Norway) to Kiruna (Sweden).

4.5 Data Assimilation

Climatological models are not sufficiently accurate to support new HF applications where it is important to adapt to the changing ionosphere. For these HF applications, and many of the others described in Section 2, it is necessary to have a much more accurate real-time ionospheric map in order to be able to better predict the propagation of the radio signals. Local area and wide area forecasts have been developed using predictive filtering and heuristic techniques, e.g. [*Francis et al.*, 2000] but more recently data assimilation approaches have been addressed. PRISM [*Daniell et al.*, 1995], was an early attempt to deal with the problem of fusing data from various sources into the model.

Although many ground and spaced based techniques have been developed to characterise the ionosphere, the measurements are generally sparsely distributed. Furthermore the measurements often do not measure ionospheric electron density directly; rather they may measure critical frequencies, integrated total electron contents (TEC) or airglow. Such measurements must be inverted to yield electron density. Often the inversions are underdetermined and it is therefore necessary to apply

constraints to the inversions. This may be done by making assumptions about the ionosphere (i.e. it is spherically symmetric), by using a limited number of functions to represent the ionosphere (e.g. empirical, spherical harmonics), or by assimilating the data into a background model of the atmosphere.

The output of a data assimilation process aims to combine measurement data with a background model in an optimal way [*Rodgers*, 2000]. It is necessary to include a background model because the information that can be extracted from many ionospheric measurements is low compared to the required resolution of the electron density field under investigation (i.e. the problem is mathematically under-determined). Since both the observations and the background model contain errors it is not possible to find the true state of the environment – instead the best statistical estimate of the state must be found. Such techniques are also well suited to sparse data sets. Best Linear Unbiased Estimator (BLUE) and related variational (one, three and four dimensional) data assimilation techniques have been used in meteorology for a number of years and have recently been applied to ionospheric inversion [*Wang et al.*, 1999], [*Pi et al.*, 2003], [*Hajj et al.*, 2003].

Angling and Cannon, [2003] describe the application of data assimilation techniques for combining radio occultation data with background ionospheric models. In simulations where tomographic images provide the truth data and PIM provides the background, a fourfold decrease in the electron density error at 300 km altitude was achieved. A global assimilation simulation has also been conducted using IRI as the truth data. For a constellation of eight RO satellites, a factor of four decrease in the vertical TEC RMS error has also been demonstrated (Figure 6). The same simulation also results in a factor of three decrease in the NmF2 RMS error, and a halving of the hmF2 RMS error.

4.6 Ionospheric channel effects

Because of the detrimental effects of the ionospheric radio channel it was common, until about 10 years ago, for HF users to expect both data rates of little more than 75 bit.s^{-1} and low availabilities. However, with the advent of cheap digital signal processing in the last few years, data rates have increased significantly to 2.4 kbit.s^{-1}, 4.8 kbit.s^{-1}, and beyond in a 3kHz channel. Data rates on benign skywave channels and on ground wave paths can now reach 64 kbit.s^{-1}, albeit sometimes using wider channel bandwidths, and there is an increasing desire to achieve 16kbit.s^{-1} reliably.

In considering the impact that space weather has on HF systems it is easy to lose sight of the fact that the Doppler and multipath characteristics of the signal are critically important issues [*Cannon and Bradley*, 2003]; [*Cannon*

et al., 2002]. Indeed it is these issues that have driven the standardization and imminent introduction of digital HF transmissions for broadcasting.

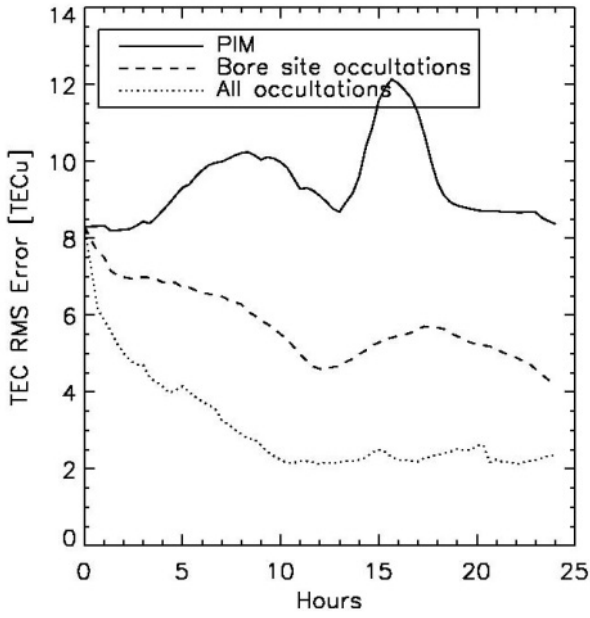

Figure 6. Vertical TEC RMS errors between IRI and PIM (solid line), IRI and a data assimilation process using bore site (±30°) radio occultation measurements (dashed line) and IRI and the same process using all available radio occultation measurements (dotted line).

The design of modern HF systems is a considerable technical challenge, the success of which critically depends on good understanding and modelling of the radio channel that is in turn controlled by space weather.

Multipath propagation may arise because replicas of the transmitted signal arrive at the receiver after reflection from more than one ionospheric layer. Additionally, the transmitted signal may undergo multiple reflections between the ionosphere and the ground, and all of these signals may be received. Each signal (or propagation mode) generally arrives with a different time delay and to complicate the situation further each may exhibit time spreading. Viewed in the frequency domain, multipath distortion dictates the coherency bandwidth of the channel.

Frequency shifts and frequency spread distortion can also be imposed on the transmitted signal by the temporal variability of the ionospheric channel, e.g.[*Basler et al.*, 1988], and this defines the coherency time of the channel. Such effects are particularly prevalent at high and equatorial latitudes where Doppler spreads of many hertz are common. However, slow fading at mid-

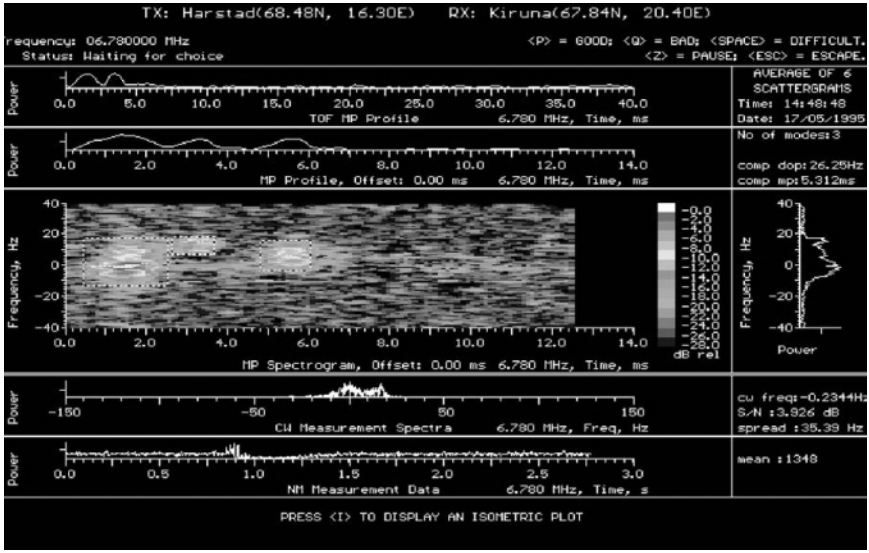

Figure 7. Doppler and multipath measurement for a high latitude path from Harstad (Norway) to Kiruna (Sweden).

latitudes is equally challenging since it can result in long periods of low signal strength.

DAMSON (**D**oppler **A**nd **M**ultipath **SO**unding **N**etwork) [*Davies and Cannon*, 1993] is a relatively low power pulse compression ionospheric sounder. It uses pulse compression sequences, on pre-selected frequencies between remote transmit and receive sites to provide real-time HF channel measurements of the channel scattering function. It therefore provides measurements of absolute time of flight (typically up to 40 ms), multipath (typically up to 12.5 ms with 600 μs resolution in a 3 kHz bandwidth), Doppler shift and spread (typically up to ±40 Hz with 0.6 Hz resolution), signal to noise ratio (SNR) and absolute signal strength (Figure 7).

Data from a high latitude DAMSON network were analysed by [*Willink*, 1997] to determine the diurnal distribution of Doppler spread, multipath spread and SNR on each high latitude path. An important result of this work was the specification of the operating envelope for the low data rate NATO HF modem, STANAG 4415 [*NATO*, 1998].

In analysis by [*Angling et al.*, 1998] the data were broken down into seasons, two time periods and four frequency bands. The times used were all day (0000-2400 UT) and 1900-0100 UT; the latter period approximately corresponded to 2100-0300 CGMLT (Corrected GeoMagnetic Local Time) when the auroral oval was expected to be in its southernmost position and, consequently, when the largest disturbances were expected. A summary of the 95th percentile multipath, Doppler spreads and SNRs were presented for

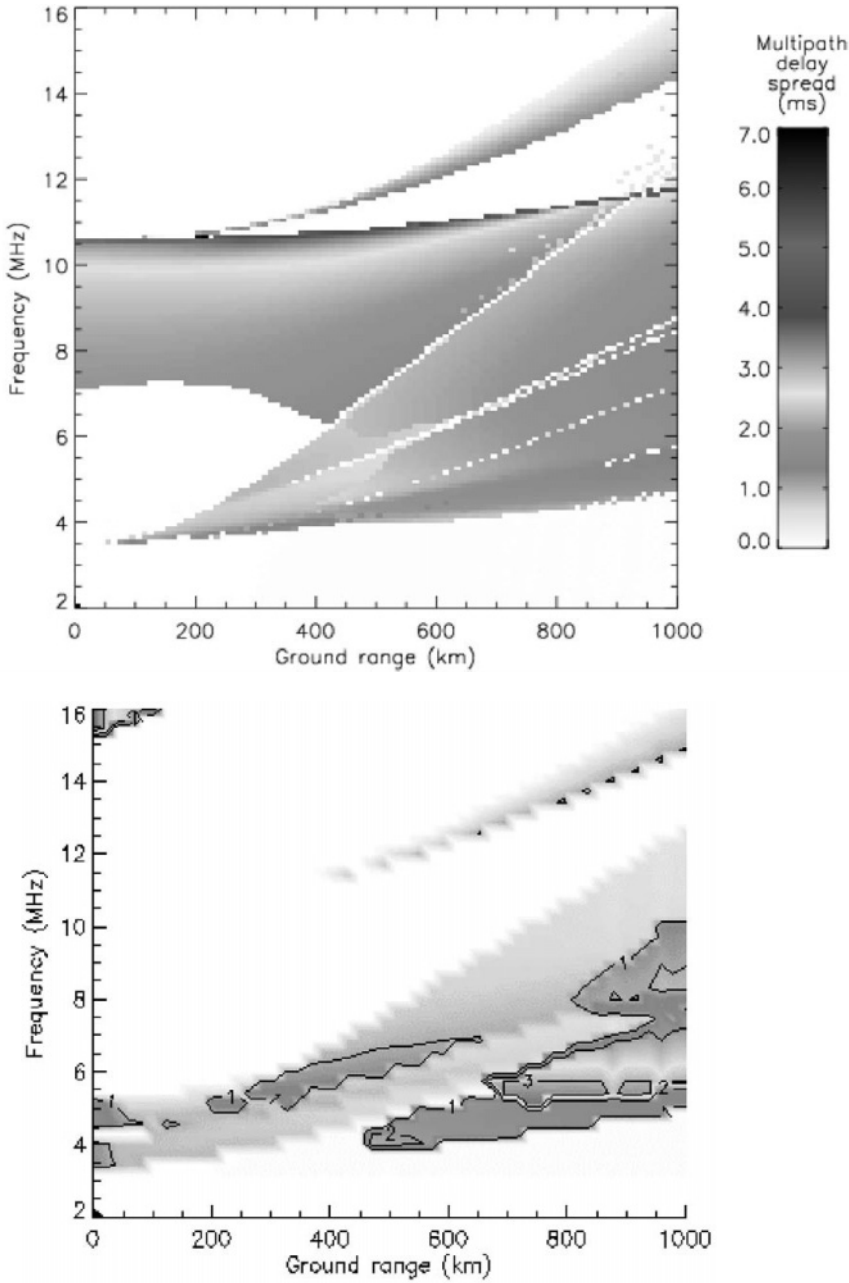

Figure 8. Multipath delay spread determined (upper panel) with SMART ray tracing through a FAIM model ionosphere and compared with (lower panel) "virtual mirror" median model predictions. Propagation is due south of 55°N, 15°E at 11UT on a March equinox day with sunspot number = 122. Modes of propagation with signal strength less than 15 dB below the strongest mode are excluded.

each path. For certain time/frequency/path combinations the 95th percentile Doppler spreads exceeded 73 Hz, while 95th percentile multipath spreads exceeded 11 ms.

4.7 Ray tracing

Ray tracing may be used in HF Over-The-Horizon Radar and HF Direction Finding systems for both frequency planning and the accurate co-ordinate registration of targets or transmitters. Importantly however, with the increasing use of digital HF broadcasting, ray tracing finds another application in as a tool for determining multipath delays, since these are not adequately determined, theoretically, using virtual mirror techniques.

A propagation model based on mirror reflection geometry is well matched to monthly median models but with improved electron density mapping of the effects of space weather improved propagation modeling will also be required. Numerical ray tracing [*Jones and Stephenson*, 1975] will provide the highest accuracy but analytic techniques have been developed which provide comparable accuracy with much shorter computer run times [*Bennett et al.*, 1991; *Norman and Cannon*, 1997; *Norman and Cannon*, 1999].

A comparison between ray-traced multipath spreads with those predicted using a popular 'virtual mirror' type of HF median model is presented in Figure 8. Clear differences in the distribution of multipath spread as a function of frequency and ground range between the analytic ray trace through the FAIM model [*Anderson et al.*, 1989.] (Figure 8a, upper panel) and the “virtual mirror” median model predictions (Figure 8b, lower panel) are evident. For example, the highest multipath spreads of Figure 8a are associated with one-hop and two-hop ionospheric F-layer reflections and these are not evident in the median model predictions (Figure 8b, lower panel). Ray tracing has the potential to considerably improve multipath delay predictions, particularly where group retardation due to underlying layers of the ionosphere is significant, and where high and low angle rays or o and x polarisations are present with similar signal strenth.

However, the accuracy of ray tracing is always limited by the the veracity of the electron density model used. This fact emphasises the importance of space weather studies leading to ionospheric models which can be updated using real-time measurements.

5 SUMMARY

In this paper we have sought to provide some understanding of why space weather is important to so many radio systems and then to illustrate this by reference to the mitigation of space weather especially in the context of HF systems.

6. ACKNOWLEDGEMENTS

The results presented herein cover work conducted by a number of people – past and present – in the Centre for RF Propagation and Atmospheric Research at QinetiQ – formerly DERA. We acknowledge their input with gratitude.

7. REFERENCES

Aarons, J., Global morphology of ionospheric scintillation, *Proc. IEEE*, *70* (4), 360-377, 1982.

Anderson, D.N., J.M.Forbes, and M. Codrescu, A Fully Analytical, Low- and Middle-Latitude Ionospheric Model, *Journal of Geophysical Research*, *94*, 1520, 1989.

Angling, M.J., and P.S. Cannon, Assimilation of Radio Occultation Measurements into Background Ionospheric Models, *Submitted to Radio Science*, 2003.

Angling, M.J., P.S. Cannon, N.C. Davies, T.J. Willink, V. Jodalen, and B. Lundborg, Measurements of Doppler and Multipath Spread on Oblique High-Latitude HF Paths and their use in Characterising Data Modem Performance, *Radio Science*, *33* (1), 97-107, 1998.

Arthur, P.C., M. Lissimore, P.S. Cannon, and N.C. Davies, Application of a high quality ionosonde to ionospheric research, in *HF Radio Systems and techniques*, pp. 135-139, IEE, Nottingham, UK, 1997.

Basler, R.P., P.B. Bentley, R.T. Price, R.T. Tsunoda, and T.L. Wong, Ionospheric Distortion of HF Signals, *Radio Science*, *23* (4), 569-579, 1988.

Bennett, J.A., J. Chen, and P.L. Dyson, Analytic Ray Tracing for the Study of HF Magneto-ionic Radio Propagation in the Ionosphere,, *App. Computational Electromagnetics J.*, *6*, 192-210, 1991.

Bilitza, D., International Reference Ionosphere 2000, *Radio Science*, *36* (2), 261-275, 2000.

Bilitza, D., Ionospheric Models for Radio Propagation Studies, in *Review Of Radio Science 1999-2002*, edited by W.R. Stone, pp. 625-679, IEEE and Wiley, 2002.

Bishop, G., T. Bullett, K. Groves, S. Quigley, P. Doherty, E. Sexton, K. Scro, R. Wilkes, and P. Citrone, Operational Space Environment Network Display (OpSEND), in *10th International Ionospheric Effects Symposium,*, Alexandria, Virginia, USA, 2002.

Bradley, P.A., and J.R. Dudeney, A simplified model of the vertical distribution of electron concentration in the ionosphere, *J. Atmos. Terr. Phys.*, *35*, 2131-2146, 1973.

Cannon, P.S., Ionospheric ELF radio signal generation due to LF and/or MF radio transmissions: Part I Experimental results, *J. Atmos Terr Phys*, *44*, 819-829, 1982.

Cannon, P.S., Propagation in the ionosphere (A), in *Propagation Modelling and Decision Aids For Communications, Radar and Navigation Systems*, pp. 1A1-1A10, NATO-AGARD, 1994a.

Cannon, P.S., Propagation in the ionosphere (B), in *Propagation Modelling and Decision Aids For Communications, Radar and Navigation Systems*, pp. 1B1-1B17, NATO-AGARD, 1994b.

Cannon, P.S., M.J. Angling, and B. Lundborg, Characterisation and Modelling of the HF Communications Channel, in *Review Of Radio Science 1999-2002*, edited by W.R. Stone, pp. 597-623, IEEE-Wiley, 2002.

Cannon, P.S., and P. Bradley, Ionospheric Propagation, in *Propagation of Radio Waves (2nd Edition)*, edited by L.W. Barclay, pp. Chapter 16, IEE, London, 2003.

Cannon, P.S., J.H. Richter, and P.A. Kossey, Real time specification of the battlespace environment and its effects on RF military systems, in *Future Aerospace Technology in Service of the Alliance*, AGARD multi-panel symposium, Paris, 1997.

Daniell, R.E., L.D. Brown, D.N. Anderson, M.W. Fox, P.H. Doherty, D.T. Decker, J.J. Sojka, and R.W. Schunk, PRISM: A global ionospheric parameterisation based on first principles models, *Radio Science*, *30*, 1499, 1995.

Davies, N.C., and P.S. Cannon, DAMSON- A System to Measure Multipath Dispersion, Doppler Spread and Doppler Shift, in *AGARD Symposium on Multi-Mechanism Communication Systems*, pp. 36.1-36.6, NATO AGARD/RTO, 7 Rue Ancelle, 92200 Neuilly sur Seine, France, Rotterdam, Netherlands, 1993.

Francis, N.M., P.S. Cannon, A.G. Brown, and D.S. Broomhead, Nonlinear prediction of the ionospheric parameter foF2 on hourly, daily and monthly timescales., *J. Geophys. Res.*, *105* (A6), 12839-12849, 2000.

Goodman, J.M., and J. Aarons, Ionospheric effects on modern electronic systems, *Proc. IEEE*, *78* (3), 512-528, 1990.

Hajj, G.A., B.D.Wilson, C.Wang, W.Pi, and I.G.Rosen, Ionospheric data assimilation of ground GPS TEC by use of the Kalman filter, *Submitted to Radio Scienceha*, 2003.

Heaton, J.A.T., IRIS Analysis Package Users Guide,DERA/CIS/CIS1/SUM990673, DERA, Malvern, Worcs, UK, 1999.

ITU, Testing of HF Modems with Bandwidths of up to about 12 kHz using Ionospheric Channel Simulators, pp. 11,Rec. ITU-R F.1487, ITU, Geneva, Switzerland, 2000.

ITU-R, HF Propagation Prediction Method,ITU-R P.533-7, International Telecommunications Union, Geneva, Switzerland, 1999.

Jodalen, V., T. Bergsvik, P.S. Cannon, and P.C. Arthur, The performance of HF modems on high latitude paths and the number of frequencies necessary to achieve maximum connectivity, *Radio Science*, *36* (6), 1687, 2001.

Jones, R.M., and J.J. Stephenson, A Three-Dimensional Ray Tracing Computer Program for Radio Waves in the Ionosphere,OT Report 75-76, US Dept. of Commerce, Office of Telecommunication, 1975.

Lane, G., F.J. Rhoads, and L. DeBlasio, Voice of America coverage analysis program, (VOACAP). A guide to VOACAP,,*B/ESA Technical Report 01-93*, US Information Agency, Bureau of Broadcasting, Washington, D. C. 20547-0001, 1993.

McNamara, L.F., *The ionosphere: communications, surveillance and direction finding*, Krieger Publishing Company, Malabar, Florida, 1991.

NATO, STANAG 4415: Characteristics of a Robust Non-Hopping Serial Tone Modulator/Demodulator for Severely Degraded HF Radio Links, NATO, Brussels, 1998.

Norman, R.J., and P.S. Cannon, A Two-dimensional Analytic Ray Tracing Technique Accommodating Horizontal Gradients, *Radio Science*, *32* (2), 387-396, 1997.

Norman, R.J., and P.S. Cannon, An evaluation of a new 2-D analytic ionospheric ray tracing technique - SMART, *Radio Science*, *34* (2), 489-499, 1999.

Pi, X., C. Wang, G.A. Hajj, G. Rosen, B.D. Wilson, and G.J. Bailey, Estimation of E × B drift using a global assimilative ionospheric model: an observation system simulation experiment,, *J. Geophys. Res.*, *108(A2), doi:10.1029/2001JA009235*, 1075, 2003.

Radicella, S.M., and R. Leitinger, The evolution of the DGR approach to model electron density profiles, *Advances in Space Research*, *27* (1), 35-40, 2001.

Rodgers, C.D., *Inverse methods for atmospheric sounding: theory and practice*, World Scientific Publishing, Singapore, 2000.

Rose, R.B., J.N. Martin, and P.H. Levine, MINIMUF-3: A simplified HF MUF prediction algorithm,Tech. Rep. NOSC TR 186, Naval Ocean Systems Center,, San Diego, California, USA, 1978.

Shukla, A.K., and P.S. Cannon, Prediction Model Updating Using the ROSE-200 Oblique Ionospheric Sounder at Mid and Higher Latitudes, in *HF, Radio Systems and Techniques Conf*, pp. 89-103, IEE, York, UK, 1994.

Shukla, A.K., P.S. Cannon, S. Roberts, and D. Lynch, A tactical HF decision aid for inexperienced operators and automated HF systems, in *7th International Conference on HF Radio Systems and Techniques*, pp. 383, IEE, Nottingham, UK, 1997.

Stewart, F.G., and G. Hand, Technical description of the ICEPAC Propagation prediction program, Private Communication, Institution of Telecommunications Sciences, Boulder, Colorado, USA, 1994.

Teters, L.R., J.L. Lloyd, G.W. Haydon, and D.L. Lucus, Estimating the performance of telecommunications systems using the ionospheric transmission channel: Ionospheric Communications Analysis and Prediction Program (IONCAP) User's manual,*NTIA report 83-127, NTIS order no N70-24144*, NTIA, Springfield, VA, USA, 1983.

Wang, C., G.A. Hajj, X. Pi, and L.J. Romans, The Development of an Ionospheric Data Assimilation Model, in *Ionospheric Effects Symposium*, edited by J. Goodman, Alexandria, Virginia, USA, 1999.

Willink, T.J., Analysis of High-Latitude HF Propagation Characteristics and their Impact for Data Communications,CRC Report 97-001, Industry Canada, 1997.

Chapter 11

Effects of Solar Radio Bursts on Wireless Systems

Dale E. Gary[1], Louis J. Lanzerotti[1,2], Gelu M. Nita[1], David J. Thomson[3]
1. *Center for Solar-Terrestrial Research, New Jersey Institute of Technology Newark, NJ 07102, USA*
2. *Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA*
3. *Queens University, Kingston, Ontario K7L 3N6, Canada*

Abstract We review the state of current understanding of the potential for interference and interruption of service of wireless communications systems due to solar radio bursts. There have been several reported instances of an enhanced rate of dropped cell-phone calls during solar bursts, and the design of current base station systems make them vulnerable to problems near sunrise and sunset for antennas facing in the direction of the Sun during outbursts. It is likely that many cases of interference have gone unreported and perhaps unrecognized. We determine the level of radio noise that can cause potential problems, and then discuss how often bursts of the required magnitude might happen. We illustrate the range of radio flux behavior that may occur, in both frequency and time, with data from the Solar Radio Spectropolarimeter and the Owens Valley Solar Array. We find that bursts that can cause potential problems occur on average once every 3.5 days at solar maximum, but also occur at a reduced rate of 18.5 days between events at solar minimum. We investigate the rate of occurrence as a function of frequency, which is relevant for future wireless systems that will operate at higher frequencies than the present systems.

Keywords Solar radio bursts, radio communications, cell phones, space weather.

1. INTRODUCTION

In the last 150 years, the number of technologies embedded in space-affected environments has vastly increased, and continue to do so. The sophistication of the technologies and how they relate to the environments in which they are embedded means that ever more detailed understanding of both the technologies and the physical environments is needed (e.g., Lanzerotti

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 203-213.

2001). This paper summarizes research into the potential effects of solar radio emissions on some wireless technologies. This work is motivated by the evidence presented in Figure 1 (from Lanzerotti et al. 1999) for an increase in dropped call rates on one day in the hour at sunrise in the wireless system of one state of the United States that appeared to be associated with a solar radio event during that interval.

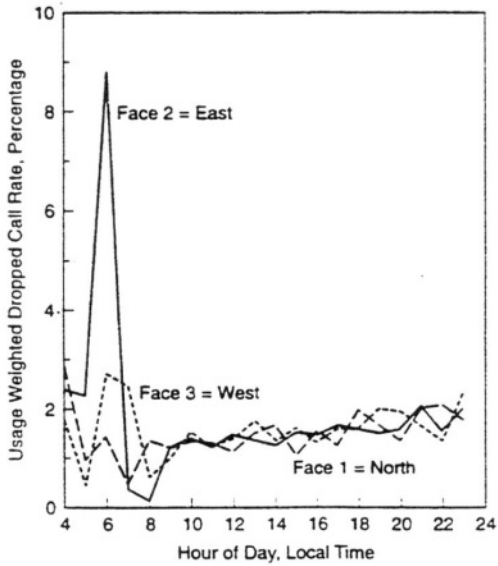

Figure 1. Usage-weighted dropped call rate for a wireless system base station, showing an enhanced level of dropped calls on the east-facing receivers near local sunrise. From Lanzerotti et al. (1999).

Microwave emissions from the sun were first reported by Southworth (1945). Radio bursts at the time of solar activity were discovered to be the source of interference in World War II radar systems (rather than by deliberate enemy jamming; Hey, 1946). Indeed, the first occurrence of this "natural" jamming was the result of intense solar activity (on 28 February 1942) that produced the first-ever measured ground level enhancement of "cosmic" rays (Forbush, 1946). Since their discovery, solar radio bursts have long been of applied-research interest for numerous reasons, including their possible predictive use for solar particle events (e.g., Castelli et al., 1973)—which in turn can cause radiation effects on spacecraft and polar-flying aircraft, and enhanced ionization in the ionosphere. Solar radio bursts continue to be of applied research interest in view of their potential interference in radar systems, satellite communication systems, and wireless.

The evidence for solar-burst influence on existing wireless systems remains indirect due to the proprietary nature of the data pertaining to system outages. Our approach, outlined in section 2, has been to examine the potential for solar bursts to affect cell-phone base stations based on our understanding of the noise and error-correction characteristics of the cell-phone waveform. We find that solar bursts exceeding about 1000 sfu (solar flux units, 1 sfu = 10^{-22} W m^{-2} Hz^{-1}) can potentially cause significant interference when the Sun is within the base-station antenna beam, which can happen for east- or west-facing antennas during sunrise and sunset at certain times of the year. In section 3 we describe the temporal and frequency characteristics of large solar bursts, using examples from the Bell Labs/NJIT Solar Radio Spectropolarimeter (SRSP) and the Owens Valley Solar Array (OVSA), operated by New Jersey Institute of Technology. We give attention to these characteristics across the entire microwave frequency band, not just the operating bands of current systems, in anticipation of future wireless systems that will undoubtedly go to higher frequencies. In section 4 we provide an overview of results of the studies we have carried out to establish the occurrence rate of solar bursts exceeding 1000 sfu. We conclude in section 5 with a discussion of the impact of solar bursts on future wireless technology.

2. SOLAR RADIO BURST FLUX THRESHOLD FOR IMPACT ON WIRELESS SYSTEMS

The discussions in Bala et al. (2002), in Nita et al. (2002) and in Lanzerotti et al. (2003) present some of the considerations of noise levels for wireless systems. For an ambient operating temperature T = 273 K, the nominal thermal noise power level P_T for a receiver of bandwidth B = 1 Hz is 3.8 x 10^{-21} W (~ −174 dBm), or 38 sfu Hz m^2. A single polarization antenna of gain G that is immersed in an isotropic radio flux of F W m^{-2} will have a receiver power of (Kummer and Gillespie, 1978)

$$P_R = G\,B\,\lambda^2 F/(8\pi) \quad \text{W Hz}^{-1} \tag{1}$$

where λ is the carrier wavelength. If F_{eq} is defined as an "equivalent" solar flux in sfu where the thermal and the solar noise levels are equal, then

$$kT\,B = G\,B\,\lambda^2 F_{eq}/(8\pi). \tag{2}$$

From (2), F_{eq} will be ~ 960 sfu for a typical cellular base station with an antenna $G \sim 10$ that is operating near 1 GHz ($\lambda^2 \sim 0.1$ m^2). Thus the total input noise $P_T + P_R$ will be more than 3 dB above thermal for $F > F_{eq}$. F_{eq} could range from ~ 1000 sfu to as low as ~ 300 sfu for an operating frequency $f \sim 1$ GHz since gains of cell site antennas can typically range from 10 to 30.

Thus, we take a flux density of ~ 10^3 sfu, the point where the noise floor rises by a factor of two, as the limit beyond which a solar radio burst can potentially affect cell-phone systems. We will investigate the likelihood of occurrence of bursts of this magnitude in Section 4, but first we give a few examples of dynamic spectra (flux density resolved in frequency and time) for large solar radio bursts to provide an overview of their characteristics.

3. SOME EXAMPLES OF SOLAR RADIO BURST DYNAMIC SPECTRA

Before discussing the occurrence rate of bursts, which merely refers to a single flux measurement at the time of maximum flux, it is worthwhile to give an overview of the spectral dynamics of typical large bursts. Some bursts show a simple, single peak in both time and frequency, for which the potential for impact on wireless systems is limited to a short duration at a small range of frequencies. However, the larger bursts typically show multiple temporal and spectral peaks, with significant complexity.

Figure 3, from the Bell Labs/NJIT SRSP radiotelescope, shows an example of a rather simple burst observed at 120 frequencies in the range 1.46-15.5 GHz, reaching a maximum flux density of about 1100 sfu near 5 GHz. Note, however, the narrow spike at the beginning of the burst, which appears in only one sample (time resolution 2 s) and reaches nearly 1300 sfu. The peak flux density may have been higher still if better time resolution were used. It is worth keeping in mind that the burst statistics we discuss in the next section were taken at a few fixed frequencies and with unexceptional time resolution, so generally the peak flux densities may be expected to be higher than the reported ones.

In Figure 4 we show a more typical large burst, observed with OVSA. Here the peak flux density exceeds 6000 sfu, and the burst displays multiple peaks in both time and frequency. In addition to the main spectral component at cm wavelengths (at frequencies $f > 3$ GHz), which is due to gyrosynchrotron radiation from electrons spiralling in the coronal magnetic field, there is also a strong and variable low-frequency component ($f < 3$ GHz) due to plasma processes in a higher coronal source. The burst remains

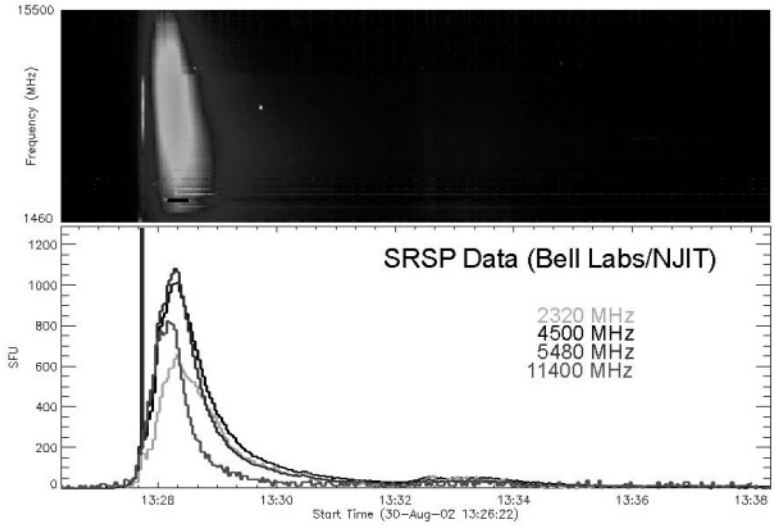

Figure 3. A simple solar burst on 2003 Aug 30, as observed with SRSP. The dynamic spectrum is shown in the upper panel, where colors represent the radio flux density, shown on a logarithmic color scale. Time is plotted horizontally, and frequency is plotted on the vertical scale, increasing upward. Time profiles at 4 representative frequencies are shown in the bottom panel.

above our 1000 sfu threshold for minutes at a time. Other bursts have been measured to remain above the threshold for as much as an hour.

4. STATISTICS OF OCCURRENCE OF SOLAR RADIO BURSTS

To establish the occurrence rate of solar bursts as a function of peak flux density and frequency, we have performed several studies whose results are summarized here. The work is described in more detail in Bala et al. (2002), Nita et al. (2002), and Nita et al. (2003). For the first two studies, we used the world-wide database of solar bursts compiled by the National Geophysical Data Center (NGDC) of the National Oceanic and Atmospheric Administration (NOAA). The database comprises 40 years of burst reports gathered from dozens of reporting stations around the world, from 1960-1999. The reports are limited to a single flux density per operating frequency for each burst, taken at the time of maximum flux density at that frequency.

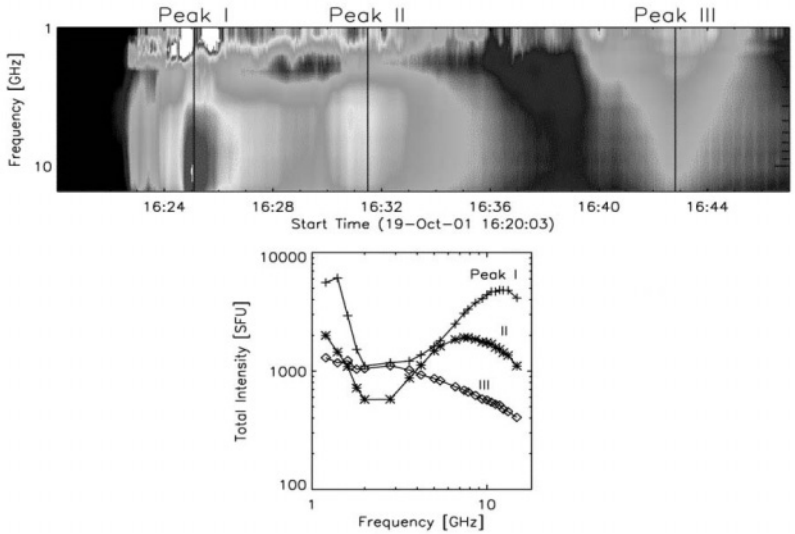

Figure 4. A large solar burst on 2001 Oct 19, as observed with OVSA. The dynamic spectrum is shown in the upper panel, as in Fig. 3, except frequency is plotted increasing downward. Vertical lines are drawn at the times of three main peaks in the emission, and the instantaneous spectra at each time are shown in the lower panel.

The dataset contains over ½ million entries concerning some 150,000 bursts. The third study consists of a database of 412 events observed over two years from 2001-2002 with OVSA. This database gives more detailed information in the form of high-resolution spectra as a function of time over the entire burst duration, as shown in Fig. 4, and was used to confirm and extend the findings from the earlier studies.

Table 1 shows some statistics of the entire NOAA dataset for all 40 years (first row), broken down by solar cycle (next three rows), and broken down by phase of the cycle (bottom two rows). The number of events per hour (last column) shows a factor of 3 increase from solar minimum to solar maximum, reaching about 0.83 events/hr. This is the observed rate of occurrence for all reported events, regardless of their flux density (down to 10-50 sfu, depending on the reporting station).

Wheatland (2000) among others has pointed out that the event occurrence rate (or its inverse, the waiting time between events) should not be determined from an average over a long period of time because the phenomenon is non-stationary—that is, the rate itself varies over time. Nita et al. (2002) did an analysis of waiting times by forming a distribution of time of occurrence of each burst relative to its predecessor, and found that the distribution of mean waiting times takes the form of an exponential. In

Year range	N_{events}	$\langle \Delta t_{events} \rangle$ (min)	N_{events}/T (events/hr)
1960-1999 (All 40 yrs)	155396	135.39	0.44
1966-1975 (Cycle 20)	39074	134.60	0.44
1976-1985 (Cycle 21)	59175	88.88	0.67
1986-1995 (Cycle 22)	45391	115.87	0.52
12 solar max years: 69-72, 79-82, 89-92	87516	72.11	0.83
12 solar min years: 66-67, 74-77, 84-87, 94-95	28261	223.30	0.27

Table 1. Statistics of events for the entire NOAA database.

this case one can deduce a characteristic waiting time using a functional form suggested by Wheatland (2000). The characteristic waiting time between bursts found by this method was about 80 minutes. This is to be compared with the overall *observed* waiting time for the entire dataset ($\langle \Delta t_{events} \rangle$ for all 40 years in Table 1) of 135 minutes, which is some 1.7 times longer.

The simplest explanation for this discrepancy is that a rather large fraction of bursts was missed. Nita et al. (2002) looked for evidence for missed bursts by plotting the distribution of occurrence time vs. time of day. Since solar bursts would not be expected to favor any particular hour of the day, any non-random distribution of time of day might be attributable to missed bursts at some geographical locations. Figure 5 shows the results of such hourly distributions. There are, indeed, significant peaks in the hourly distributions which indicate that some Earth longitudes are better covered for solar radio measurements than others. Nita et al. (2002) made the assumption that the peaks of these distributions represent the true rate (i.e., observing stations at those longitudes did not miss any events), and therefore that the ratio of filled area to total area at this peak number level represents the total number of missed events. Plots like those in Fig. 5 can thus be used to derive a "geographical correction factor," C_{geo}. It is remarkable that although it varies over time, C_{geo} was found to be close to the factor of 1.7 expected from the waiting time analysis. We note in passing that the distributions of Fig. 5 seem to show an increasing percentage of missed events at U.S. longitudes in solar cycle 22 as compared to earlier cycles.

We are now ready to show the main result of Nita et al. (2002) that pertains to solar burst impact on wireless systems. This main result is the so-called cumulative distribution of events, plotted separately for solar maximum years and solar minimum years in Figure 6. A cumulative distribution has the property that a given bin contains the cumulative number of bursts at that flux density *and higher*. The number on the vertical axis is the number of events per day. The plot on the left is for solar maximum years, and the plot of the right is for solar minimum years. The actual data are shown by the binned line, while the best power law fit is shown by the solid diagonal lines. The fit is useful for parametrizing the distribution for analytical use, and the parameters of the fit are shown in each panel. The dotted lines show the same fits, but multiplied by the relevant geographical correction factor, C_{geo}, also shown in each panel.

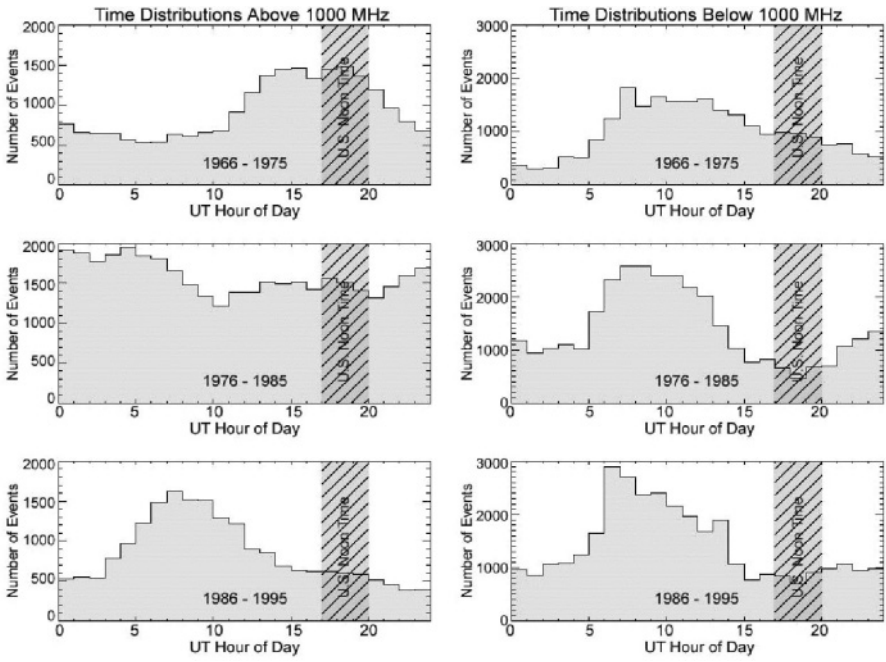

Figure 5. Plots of hourly occurrence rate of bursts in the NOAA database for events with peak frequency $f > 2$ GHz (left panels) and for events with $f < 2$ GHz (right panels). The distributions are shown separately for the three complete solar cycles 20, 21 and 22. The UT hours corresponding to noon-time time zones for the U.S. are shown as a hatched region in each plot.

As we noted in section 2, the threshold for potential impact on wireless systems is ~1000 sfu, so to determine the occurrence rate of solar bursts above this flux density at solar maximum, one merely reads the

number from the graph as shown by the arrows in the left panel of Figure 6 to get a value of about 0.28 events/day, or one event every 3.5 days, on average. To obtain this number we used the dotted line, which is the fit to the observed points corrected for missed events by multiplying by the geographical factor C_{geo}. A similar procedure (right panel) shows that the number of bursts at solar minimum falls to one event every 18.5 days.

Figure 6 shows the data and fits for all bursts above 2 GHz, meaning that bursts are counted regardless of the frequency at which they occur. Designers of current and future wireless systems may wish to evaluate potential interference only for bursts within their operating band. For this reason, Nita et al. (2002) tabulate the fit parameters $N(S > 1$ sfu) and λ, as shown in Figure 6, for burst distributions in other frequency ranges. To apply the tabulated values for a given flux density threshold S_0, one inserts the parameters into the equation

$$N(S>S_0) = N(S>1)\, S_0^{\lambda+1}.$$

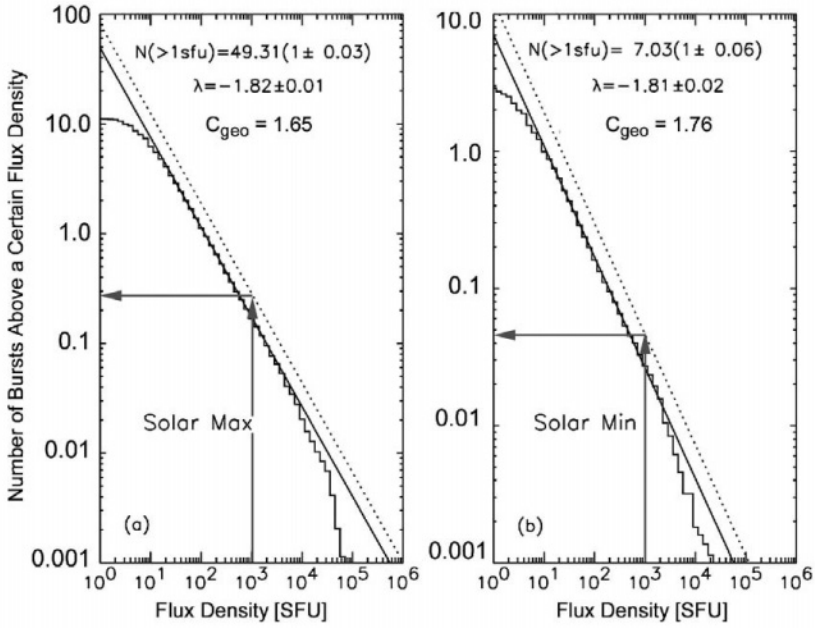

Figure 6. Cumulative distributions of number of bursts per day greater than a given flux density near solar maximum (left panel) and solar minimum (right panel). The data are shown by the histograms, and the solid lines show the best power law fit to the distributions. The fit parameters are given in the annotation. The fall of the data away from the fit at low flux densities (< 20 sfu) is due to the instrumental sensitivity limit of the reporting stations. The lack of large bursts may be a real solar effect. The dotted lines show the power law fit after multiplying by C_{geo}. The arrows show how to read the number of bursts for a given flux density threshold (10^3 sfu in this case—see text).

Recently, Nita et al. (2003) have completed another statistical study of solar bursts, this time with full temporal and spectral resolution based on two years (2001-2002) of OVSA data (see Fig. 4). Although the aim of the work was basic understanding of solar burst phenomena, some characteristics of the results are relevant to the topic of this chapter. In particular, they found that there is a relatively sharp dividing line between gyrosynchrotron bursts (above 2.6 GHz) and decimetric bursts due to coherent processes (below 2.6 GHz). The coherent bursts have flux densities that can have almost any value, so high-flux bursts are equally possible at any frequency below 2.6 GHz. Above this frequency, however, the bursts are due to gyrosynchrotron emission and show a rather clear, frequency-dependent limit of $\sim 300(f_{\mathrm{GHz}})^{2}$ sfu over the range 2.6-18 GHz. The limit is empirical, not fundamental, and appears to be due to the product of the practical limit of burst size (< 1.3 arc-minutes diameter) and the typical brightness temperature ($\sim 10^{9}$ K). Rare individual bursts can probably exceed this flux limit, but the finding offers some indication that frequencies just above 2.6 GHz are least likely to have high flux densities.

5. CONCLUSIONS

We have examined the statistical properties of solar bursts from the point of view of their potential impact on wireless systems, in particular cell-phone base stations. An analysis of the noise floor of typical base stations shows that bursts exceeding ~1000 sfu will double the noise and hence may begin to cause problems for the system if the horizon-looking antennas are pointed at the rising or setting Sun. Our analysis shows that such bursts occur on average once every 3.5 days during solar maximum and once every 18.5 days at solar minimum.

Since a given base station of a wireless system is at risk for only a short period (about 1 hour) around sunrise and sunset, a typical station may be affected at roughly $1/12^{\mathrm{th}}$ of this rate, or once per 42 days at solar maximum and once per 222 days at solar minimum. Thus, the impact may be deemed small. However, any optimism should be tempered by the facts that (1) a large geographical area will see the rising or setting Sun simultaneously, and so any impacts may be felt system-wide and (2) systems spanning multiple time zones are at risk for correspondingly longer times. Note also that the largest bursts may attain peak flux densities 10-100 times the limit of 1000 sfu that we identified as having a potential impact, so on rare occasions the impact may be more severe.

As technological systems continue to proliferate, it is wise to keep all potential environmental influences in mind. Solar radio bursts represent one aspect of Space Weather that can easily be overlooked, but may nevertheless cause problems for certain technologies. We can look forward to wireless systems moving to higher frequencies in the future. Our work indicates that wireless system operating frequencies just above 2.6 GHz are the most favorable for avoiding impacts from solar bursts, but the impacts below 2.6 GHz and above about 10 GHz are significantly higher.

6. ACKNOWLEDGEMENTS

This work was supported through an NSF Space Weather program grant ATM-0077273 to New Jersey Institute of Technology. The Owens Valley Solar Array is supported by NSF grant AST-0307670 and NASA grant NAG5-11875 to New Jersey Institute of Technology.

7. REFERENCES

Bala, B., Lanzerotti, L. J., Gary, D. E. & Thomson, D. J., Noise in wireless systems produced by solar radio bursts, Radio Science, 37(2), 10.1029/2001RS002488, 2002.

Castelli, J. P., J. Aarons, D.A. Guidice and R.M. Straka, The Solar Radio Patrol Network of the USAF and Its Application. Proc. IEEE, 61, 1307, 1973.

Forbush, S. E., Three recent cosmic-ray increases possibly due to charged particles from the Sun, Phys. Rev., 70, 771, 1946.

Hey, J. S., Solar Radiations in the 4--6 metre Radio Wavelength band, Nature, 158, 234, 1946.

Lanzerotti, L. J., Thomson, D. J., & Maclennan, C. G., Modern Radio Science, Oxford, 25, 1999.

Lanzerotti, L. J., Space Weather Effects on Communications, Space Storms and Space Weather Hazards, Proc. NATO Advanced Study Institute on Space Storms and Space Weather Hazards, (ed: I.A. Daglis), Kluwer Academic Publishers, 313, 2001.

Lanzerotti, L. J., Gary, D. E., Nita, G. M., Thomson, D. J. & Maclennan, C. G., Noise in wireless systems from solar radio bursts, Advances in Space Research, in press, 2003.

Nita, G. M., Gary, D. E., Lanzerotti, L. J. & Thomson, D. J., The Peak Flux Distribution of Solar Radio Bursts, Astrophysical Journal, 570, 2002

Nita, G. M., Gary, D. E. & Lee, J., Statistical study of two years of solar flare radio spectra obtained with OVSA, Astrophysical Journal, in press, 2003

Southworth, G.C., Microwave Radiation from the Sun, J. Franklin Inst., 239, 285, 1945.

Wheatland, M. S., The Origin of the Solar Flare Waiting-Time Distribution, Astrophysical Journal, 536, L109, 2000.

Chapter 12

Space Weather Effects on Aircraft Operations

An Introductory Overview

Captain J Bryn L Jones
Mullard Space Science Laboratory, (UCL), Holmbury St Mary, Dorking,Surrey, UK
SolarMetrics Limited, Cheltenham, Glos, UK
Virgin Atlantic Airways Limited, Crawley UK

Abstract European airlines are now monitoring occupational exposure of aircrew to cosmic radiation, but the impacts of SW on other systems used for day-to-day operations are not considered. Future civil and military aircraft will use increasingly complex avionics and will operate in a satellite-based air-space management network. However, to maximise effectiveness of this globally interoperable system and maintain safe operations will require a greater understanding of the SW risks with increasing technology. This chapter describes the SW effects on aircraft operations and includes brief details of an ESA project to quantify those risks.

Keywords Aircraft, avionics, cosmic radiation, solar energetic particles, space weather, FANS, Network-Centric, air-space management, operational risks.

1. INTRODUCTION

The introduction of European Union (EU) legislation requiring the monitoring of aircrew exposure to Cosmic Radiation (CR), including any variations at aircraft altitudes due to solar activity, has briefly raised the profile of SW in the airline industry. However, SW phenomena can affect all areas of aircraft operations including avionics, communications and satellite navigation systems.

Since man first launched into the air 100 hundred years ago, day to day aircraft operations, whether civil or military, have developed into incredibly complex and structured events, requiring careful planning and control to

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 215-234.

ensure above all else that Flight Safety is maintained at all times. The envisaged developments in aircraft and support systems over the next 10 to 20 years suggest that aircraft operations are about to enter a new era of technology infrastructure utilising satellites and micro-electronics. This increasing reliance upon such technologies means that all aspects of civil and military aircraft operations will become increasingly susceptible to the impacts of Space Weather (SW).

Despite the World Trade Centre terrorist attacks of 9/11 and the current poor economic situation for airline companies the demands for commercial air travel continue to grow at ~5% per annum. Airline companies are continually investing in new technologies and services in a bid to improve efficiency and drive down the costs per seat mile. This includes the introduction of onboard flight data and performance monitoring systems, Electronic Flight Bags for pilots, enhancing the passenger experience with on-demand in-flight entertainment, and wireless and satellite communication systems for aircraft operations, email and internet.

At the same time the management of the air traffic infrastructure is being developed and enhanced to provide more airspace for more aircraft. Improved Command, Control, Communication and Information (C^3I), via satellites, bring the benefits of increasing navigational accuracy and greater connectivity while outside of ATC radar coverage, thus reducing the separation between aircraft during all stages of the flight, and ensuring the availability of accurate real-time information for operational and commercial decisions. In a similar manner, future military capabilities are being planned around the concept of multi-national "Network-Centric" operations, utilising the latest technologies in avionics, secure data-link communications and pilot-sensory integration. For military aircraft operations this will provide a quantum leap in the situational awareness of the pilots, mission planners and Command Centres for their whole Theatre of Operations.

With the introduction of new ultra-long-haul "over-the-pole" routes, "more-electric" aircraft in the future, and the increasing use of satellites in the operation, the need for a better understanding of the SW impacts on future aircraft operations will become all the more compelling as we look towards the next solar maximum around 2011/12. These impacts will be discussed further in this chapter.

2. SPACE WEATHER EFFECTS

The accepted definition of SW is "conditions on the sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of space-borne and ground-based technological

systems and can endanger human life and health" (US National Space Weather Strategic Plan, Aug 1995). Included in this definition are the effects of Galactic Cosmic Rays (GCRs), that originate from exploding stars outside our solar system, but which also affect technological systems, and endanger human life and health because their flux is modulated by solar processes. From this very broad definition, it is useful to define just the major environmental events that are applicable to aircraft operations. And while a description of the underlying physical processes of these events (and their interactions within the complete Sun-to-Earth SW model), are beyond the scope of this chapter, their variability and magnitude will be mentioned later when describing the level of risk to operations.

The SW events that concern aircraft operations most are those that increase the radiation environment, and those that disrupt operational systems. Additionally, recent research now suggests there may be a greater influence of the solar cycle and SW upon terrestrial weather (Heavner et al., 2000). Therefore, the most important SW phenomena are GCRs, Solar Energetic Particles (SEPs), Coronal Mass Ejections (CMEs) and Geomagnetic Storms (and more directly, the ionospheric disturbances).

2.1 Galactic Cosmic Rays

Galactic Cosmic Rays are primarily 89% protons (ionised hydrogen atoms), the remainder comprising 10% alpha particles (ionised helium atoms), and about 1% heavy ions. Their energies can extend up to 10^{20}eV, but most effects result from those GCRs of some 100s MeV to ~ 20GeV. These high energy cosmic rays collide with the upper atmosphere starting around 130,000ft (40km), where they produce a cascade of "secondary" particles (pions, muons, neutrons, electrons, positrons and gamma rays) that shower down through the atmosphere to the Earth's surface. The number of particles reaching the Earth's surface is related to the energy of the cosmic ray that struck the upper atmosphere. It is these highly ionising GCRs and secondary particles that gives rise to the cosmic radiation hazard to humans at aircraft altitudes and can cause Single Event Effects (SEEs) in aircraft avionics. The corresponding level of radiation dose reaches a maximum intensity at around 66,000ft (20km) and then slowly drops off by sea level; however, the 11-year solar cycle also gives ± 20% variation in dose from solar minimum to maximum due to the anti-correlation in cosmic ray density. Figure 1 shows the variation in cosmic ray intensity with solar activity.

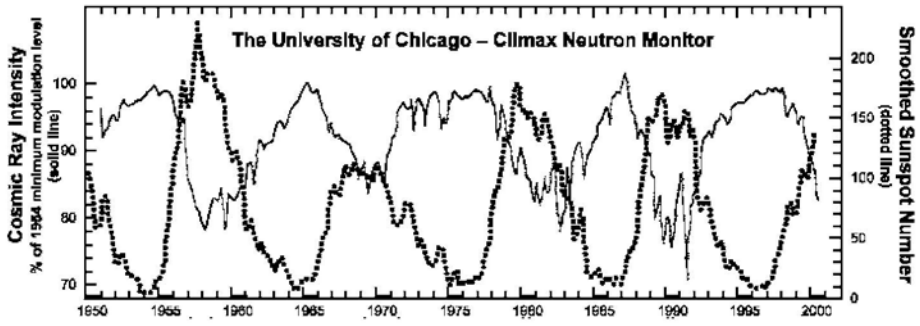

Figure 1. Long term Climax neutron monitor observations and smoothed sunspot numbers SOHO mass-memory single-event upsets. (Courtesy University of Chicago)

The cosmic rays show an inverse relationship to the sunspot cycle because the Sun's magnetic field is stronger during sunspot maximum and shields the Earth from cosmic rays. The Earth's magnetic field also provides some shielding such that cosmic ray fluences vary with geomagnetic latitude. Observed dose rates at an altitude of 26,000ft (8 km) near the equator are about 1 to 1.5μSv/hr, but at temperate latitudes this increases up to about 3 microSv (μSv) per hour, reaching a constant at about 50°. At 39,000ft (12km), the values are greater by about a factor of two. Figure 2 shows how the global radiation dose distribution varies with geomagnetic latitude at a constant altitude of 35,000ft. (Sievert (Sv) is the unit of Dose Equivalent, which is a measure of the quantity of energy deposited in a unit mass of matter, such as biological tissue, taking into account the radiobiological effectiveness of the various types of ionising radiation). The plots in Figure 3 clearly show the effects of altitude and latitude on received total counts (which can be translated into a radiation dose) on one particular Virgin Atlantic Airways flight from London to Johannesburg, measured with a Tissue Equivalent Proportional Counter (TEPC). Several of these TEPCs are currently being flown onboard Virgin aircraft as part of a UK Government sponsored collaborative study to measure cosmic radiation at aircraft altitudes. This study, involving Mullard Space Science Laboratory (MSSL – University College London), Civil Aviation Authority (CAA), National Physical Laboratory (NPL) and Virgin Atlantic, started in January 2000 and will have flown monitors on over 1000 world-wide flights when it ends in 2004.

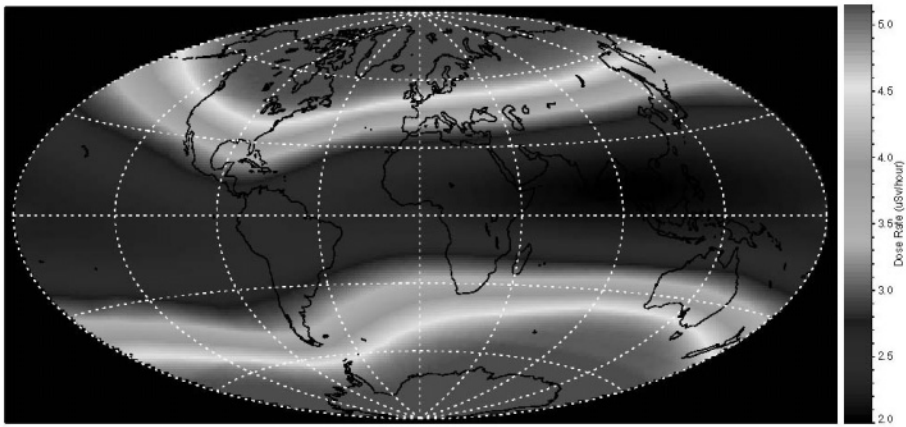

Figure 2. Projected global dose rate at 35,000 ft (Model results from the Federal Aviation Administration's CARI-6).

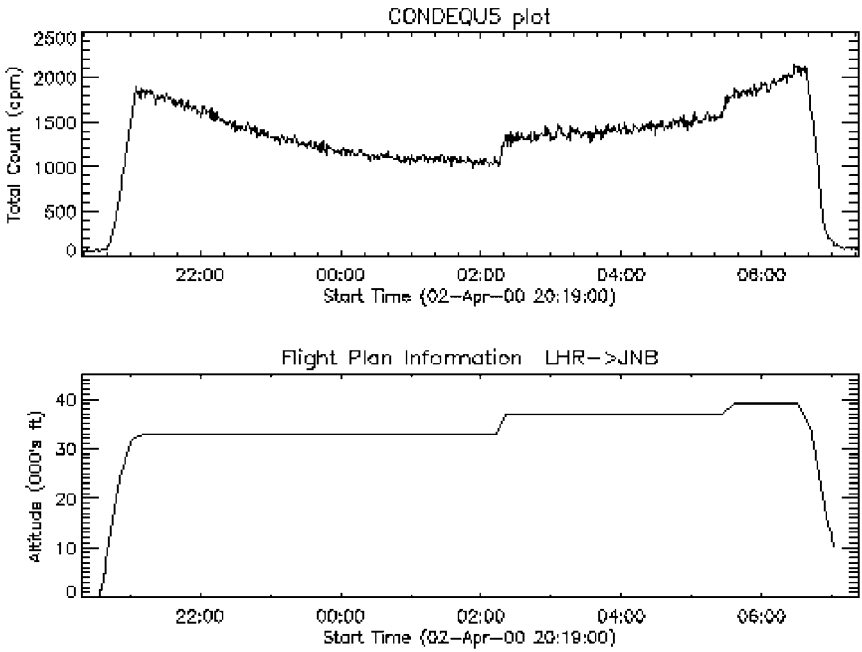

Figure 3. Airbus A340 flight from London to Johannesburg showing effects of altitude and latitude variation on cosmic ray dose.

2.2 Solar Energetic Particles

Solar flares with lifetimes ranging from hours for large gradual events, down to tens of seconds for the most impulsive events, release ultraviolet, x-ray and radio emissions, reaching the Earth in about 8 minutes, and producing ionospheric disturbances in the sunlit hemisphere of minutes to hour's duration. Large flares, known as Solar Energetic Particle events (SEPs), can release very energetic particles (primarily protons), which then arrive in our atmosphere within 30 minutes. The Earth's magnetic field does offer some protection, but these particles can spiral down the field lines, entering the upper atmosphere in the polar regions where they produce additional ionisation in the ionosphere and increase the radiation at aircraft altitudes. The largest SEP events tend to occur on either side of solar maximum and can cause a significant rise in the neutron monitor count rates at ground level, which are classed as Ground Level Events (GLEs). Figure 4 shows a plot of the number of GLEs recorded for the last 5 solar maxima.

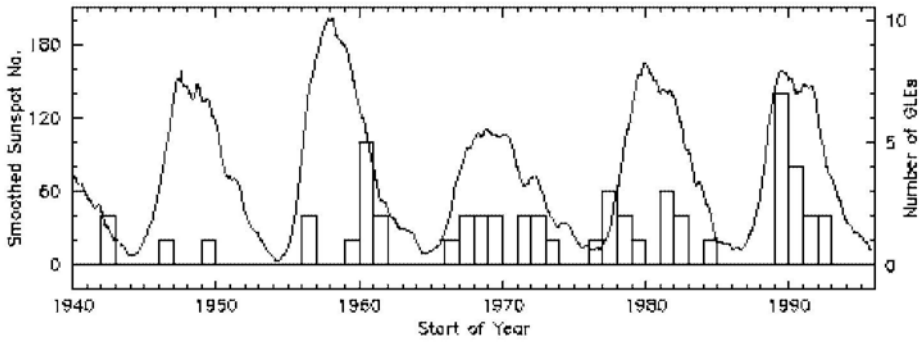

Figure 4. Monthly smoothed sunspot numbers and annual frequency of Ground Level Events (histogram) for the period 1940 - 1995. (Cramp 2000)

Figure 5 shows an image of the 14 July 2000, Bastille Day SEP event along with the GOES-10 proton flux measurements for energies from 370 MeV to > 850 MeV for the same event. The red and blue lines depict the time intervals during which two London-Tokyo Virgin flights (14th and 21st July 2000 respectively) were airborne. During the earlier flight there is a significant enhancement of the proton flux across all the energy channels. In contrast by the 21st July the flux has returned to background levels.

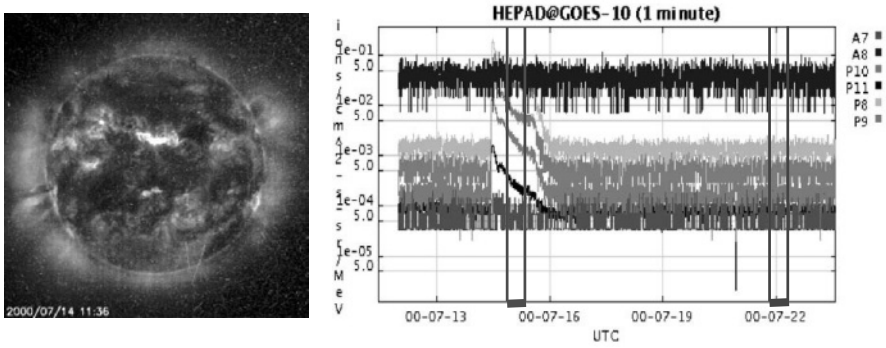

Figure 5. GOES/HEPAD measurements of solar protons and He^+ for the 14th July 2000 Bastille Day event. Virgin flights on the 14th and 21st July are indicated by the red and the blue intervals respectively. [A7: (He^+) 630 - 850 MeV, A8: (He^+) > 850 MeV, P10: 640 - 850 MeV, P11: > 850 MeV, P8: 370 - 480 MeV, P9: 480 - 640 MeV]

2.3 Coronal Mass Ejections

The explosive release of Coronal Mass Ejections (CMEs) from the Sun's outer atmosphere over the course of several hours, can also rapidly shower the Earth with accelerated energetic particles and cause severe disturbances in the physical characteristics of the solar wind (e.g. density, composition, and magnetic field strength). Of primary importance are the Earth directed events, known as halo CMEs, which generally produce the most severe SW impacts. Accurately predicting their time of arrival and severity is however, still very difficult. The CME shock accelerated particles may lead to increases in atmospheric radiation levels at aircraft altitudes and the arrival of the CME at Earth can significantly enhance the particle population in the radiation belts, which in turn can increase the radiation doses to satellites.

2.4 Geomagnetic Storms

Because the solar wind varies over time scales as short as seconds, the boundary between interplanetary space and the Earth's magnetosphere is extremely dynamic. One to four days after a solar disturbance a plasma cloud reaching the Earth will buffet the magnetosphere resulting in a Geomagnetic Storm. The magnitude and orientation of the solar plasma's magnetic field when it impacts the magnetopause will affect the severity of

these storms; however, predictions rely primarily on data measured by the ACE satellite, which only gives approximately 45 minutes warning.

3. CURRENT OPERATIONAL IMPACTS

These SW events, briefly described above, can impact on aircraft operations in many ways. Of primary importance within the industry at the moment is the radiation hazard to aircrew and passengers. However, other operational hazards include the SW effect on avionics, communications, GPS navigation systems and possible variations in terrestrial weather. Table 1 summarises these cause and effects for specific areas.

Impact Areas	Primary Event	Effects
Hazard to Humans	GCRs	Radiation dose to aircrew/passengers
	SEPs & CMEs	Increased doses to aircrew/passengers
Avionics	GCRs, SEPs,	Upsets in aircraft electronics
Communications	SEPs	PCAs – HF Communications disruption
	CMEs, Geomagnetic Storms	HF, VHF
Satellite Navigation	GCRs, SEPs	Radiation damage to onboard systems
	Geomagnetic Storms	Upsets in space electronics
	Ionospheric Disturbances	Navigation services disruption
Terrestrial Weather	GCRs	Cloud & precipitation
	Solar cycle u/v	Jetstreams
	SEPs	Ozone
	Ionospheric disturbances & sub-storms	Thunderstorm Red sprites, Blue jets

Table 1. The various SW events and their effects given for specific impact areas.

3.1 Hazard to Humans

Monitoring occupational exposure to natural sources of ionising radiation is included in the recommendations provided by the International Commission for Radiological Protection (ICRP). This includes exposure to the background cosmic radiation received while flying. Within the European Union (EU) this has been translated into National legislation by each of the Member States, such that “operators” (of aircraft) should assess and monitor the exposure of their aircrews to both GCRs and SEPs.

Ionising radiation is measured in terms of absorbed dose, the energy deposited per unit mass. Equal absorbed doses of different types of radiation

cause biological effects of different magnitudes, and the sensitivity of different tissues of the body differ. To account for this, tissue absorbed doses are multiplied by radiation weighting factors to give equivalent doses, and then by tissue weighting factors to give the effective dose in Sieverts (Sv) to the whole body.

Under present international guidelines, the recommended dose limit for aircrew is a 5-year maximum 100 mSv dose, averaged to 20 mSv per year, but with no more than 50 mSv in any single year. In the EU a maximum limit of 6 mSv/yr has been adopted by the airlines for record keeping purposes, which is workable with current flight profiles and annual block hours. Should an individual exceed this 6 mSv level then airlines must keep their records for 30 years (even if the person is deceased) or until age 75, whichever is the longer period of time. For a pregnant crewmember, starting when she reports her pregnancy to management, her work schedule should be such that the equivalent dose to the child is as low as reasonably achievable and unlikely to exceed 1 mSv during the remainder of the pregnancy. However, if future generations of large commercial aircraft are designed for increased range or to utilise the available airspace at higher altitudes, then we can expect to see significant increases in the doses as shown in Figure 6.

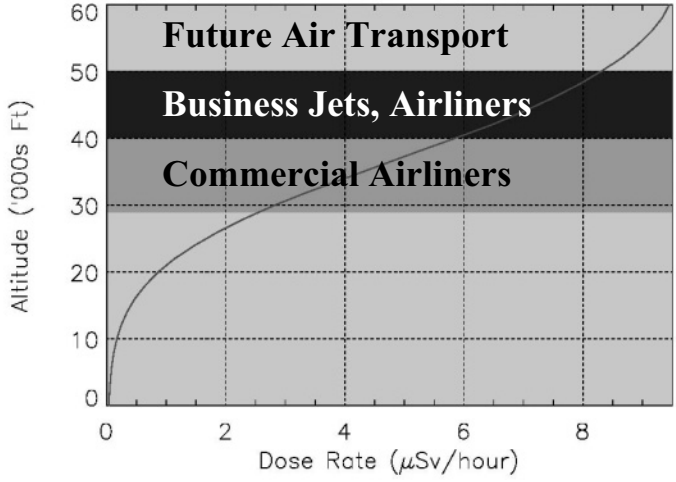

Figure 6. Change in dose rate due to cosmic radiation (GCR component) as a function of altitude and aircraft operational type.

Through improvements of GCR dose calculations, or preferably by actual measurement, it may be possible to make efficient scheduling of current aircrew without the need to employ significant numbers of additional crews. However, with each significant increase in aircraft performance (altitude, range), the need to adjust an individual's roster will become more likely; with a corresponding increase in employee numbers. Quicker flights will reduce doses, but significant increases in cruising speeds will need to be achieved: Boeing's scrapped Sonic Cruiser flying at Mach 0.98 (98% of the speed of sound) was forecast to reduce flight times by 15-20%, but as it would have operated at higher altitudes, the route doses were likely to increase by 30-40%. Only supersonic aircraft will keep route doses down for any significant increase in operating altitude. Concorde operated up to 58,000ft; however, its routes tended to be at mid-latitudes and of short duration, thus aircrew annual doses were lower than those who operate on the ultra-long haul aircraft.

Typically, a London to Los Angeles flight in current commercial aircraft accumulates ~65μSv (6μSv/hr) from the GCR component. However, the impact of SEPs, even at current cruising altitudes, could significantly increase the dose. Using CREAM (Cosmic Radiation Effects and Activation Monitor) data collected onboard a Concorde flight from London to New York on 29th September 1989, recent calculations (Dyer et al., 2001) performed on large SEP events (23rd February 1956, 29th September 1989) estimated that the additional radiation dose received at 40,000ft (12km) on a subsonic transatlantic flight would have been approximately 10 mSv and 2 mSv respectively. This work also shows that the flux of these SEP events has a very steep dependence on altitude and cut-off rigidity (a measure of a particle's resistance to bending in the Earth's magnetic field), which is very important when considering any increase in aircraft operating altitude. Measurements taken during GLE60 on 15th April 2001 onboard a subsonic aircraft did show an increase in the dose rate by a factor of 2 for approximately one hour. Besides this recent data and that collected onboard Concorde, very few SEP events have so far been captured with onboard measurements as only Concorde has had a regulatory requirement to carry a monitor. In the event of future SEPs producing significantly elevated dose rates at aircraft altitudes it has been proposed that calculation of SEP doses to crew may be done retrospectively using computer calculations. However, such a technique requires data from a large number of geomagnetically-dispersed, ground-level neutron monitors, and such calculations would also require validation against flight observations using active monitors. Despite this clear evidence that SEPs can increase the radiation dose, and will become increasingly important in higher flying aircraft, there is no current requirement to permanently carry active monitors onboard any aircraft.

Without the availability of active monitors, current aircraft operations are reliant upon very inaccurate solar flare alerting models for warning of an increased radiation hazard. Should solar activity occur above set thresholds, then a Solar Radiation Alert would be issued by the FAA, as shown in Figure 7. The current advice for all aircraft operating anywhere in the North and South hemisphere shaded regions would be to descend below 29,000ft; however, the practicalities of completing this procedure safely within busy airspace regions using current Air Traffic Management systems make it almost impossible. At present, reacting to such radiation hazard warnings, especially if they are inaccurate and geographically un-focussed, is also not commercially viable for current operations as the increased fuel burn at the lower altitudes could result in diversions costing up to $100,000 a time.

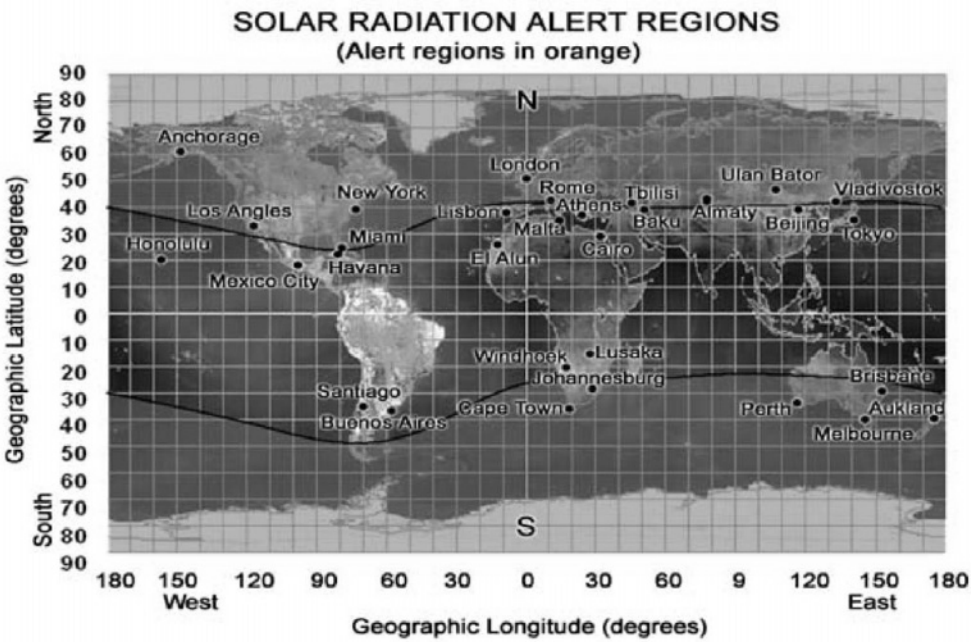

Figure 7. A Solar Radiation Alert indicates a solar particle event is in progress that may lead to a substantial increase in the ionising radiation at aircraft altitudes in the North and South shaded areas shown on the map.

As an indication of what the civilian operation may need to consider in the future with higher cruising altitudes, Table 2 is an example of a notional go/no-go radiation risk matrix for high altitude military operations. This table integrates the applicable high flying background cosmic radiation with the NOAA SW scales for SEPs. Obviously the acceptable level of risk for military operations will always be significantly different to that for

civilian operations, but the principle behind such risk assessment indices is still applicable.

Hi-Flyer	SPE – 0	SPE – S1	SPE – S2	SPE – S3	SPE – S4	SPE – S5
5mR/hr	GREEN	GREEN	AMBER	RED	RED	RED
10mR/hr	AMBER	AMBER	AMBER	RED	RED	RED
50mR/hr	AMBER	AMBER	AMBER	RED	RED	RED
100mR/hr	RED	RED	RED	RED	RED	RED

Table 2. Example of a notional military High Altitude Radiation Index (HARI). 100mR = 1mSv. (Courtesy 9th RW ORM)

3.2 Avionics

The electronic components of aircraft avionic systems are susceptible to damage from the highly ionising interactions of cosmic rays, solar particles and the secondary particles generated in the atmosphere. And as these components become increasingly smaller then the risk of damage also increases. The heavier and most energetic particles can deposit enough charge in a small volume of silicon to change the state of a memory cell, a one becoming a zero and vice versa. This can corrupt systems leading to erroneous commands. These soft errors are referred to as Single Event Upsets (SEU). Sometimes a single particle corrupts more than one bit to give Multiple Bit Upsets (MBU). Certain devices could be triggered into a state of high current drain, leading to burn-out and hardware failure; such effects are termed single event latch-up or single event burn-out. All these interactions of individual particles are referred to as Single Event Effects (SEE).

Satellites incorporating sensitive Random Access Memory (RAM) chips have had upset rates from one per day at quiet times to several hundred per day during SEP events. In-flight measurements of SEU sensitivity in 4Mb SRAM produced a rate of 1 upset per 200 flight hours, and agreed well with the expected upset rate variations due to changing latitude. Research has already shown that 100MB of modern RAM found in laptops and PC's may suffer upsets every 2hrs at 40,000ft, or 1 upset/minute in 1GB of memory due to the 1989 SEP event. This problem is expected to increase as more, low-power, small feature size electronics are deployed in "more electric" aircraft.

3.3 Communications

Many communication systems utilise the ionosphere to reflect radio signals over long distances. Ionospheric storms can affect radio communication at all latitudes. Some radio frequencies are absorbed, while others are reflected, leading to rapidly fluctuating signals and unexpected propagation paths. Solar flare ultraviolet and x-ray bursts, solar energetic particles, or intense aurora can all bring on these conditions. A sudden increase of x-ray emissions from a flare will increase the ionisation in the lower ionosphere on the sunlit side of the Earth and cause a Sudden Ionospheric Disturbance (SID) of radio signals. At HF, and occasionally at VHF, an SID may appear as a short-wave fade, which can last for minutes to hours. If the flare is sufficiently large the effects may become especially strong, causing a total communications blackout. SEPs produce a particular type of disturbance called Polar Cap Absorption (PCA) that can last for many days. When very energetic particles enter the atmosphere over the polar regions, the enhanced ionisation produced at these low altitudes is particularly effective in absorbing HF radio signals and can render HF communications impossible throughout the polar regions. Several US airlines indicated that they have either had to cancel, delay or divert trans-polar flights due to such SW events.

Other airspace regions of civil operations also rely heavily on HF communications. The North Atlantic and Pacific Ocean regions use HF for aircraft position reporting to maintain separation while outside of ATC radar coverage. Even relatively minor SW disturbances can seriously disrupt the HF signal causing significant impact on these oceanic region procedures. The increasing use of a new Automatic Dependent Surveillance (ADS), an automated satellite reporting system, is reducing the reliance upon HF in such regions, but in itself brings with it the problems of SW impacts on satellite-based systems.

Within normal radar coverage, civil aircraft operations use VHF frequencies. Although less prone to interference, VHF signals can be lost in the noise produced by solar flares: a point not generally considered when investigating temporary losses of communication between aircraft and ATC. Action focuses primarily on aircraft equipment serviceability, with the majority resulting in a “no fault found”. Such transient losses of communication could result in aircraft separation minima being eroded as ATC avoidance transmissions are missed, or within the military sphere, a friendly aircraft is engaged as hostile due to lack of response. Robust communications will become more important for future civil and military integrated airspace management systems, which are described in the section on technological developments.

3.4 Satellite Navigation

There are now plans to use GPS for navigating aircraft so that the separation between aircraft can be reduced, and to position the aircraft on approach. There are also studies in progress on the longer-term goal of landing aircraft by GPS. However, the accuracy of the GPS signal, which must pass through the ionosphere, is obviously affected by any ionospheric variations due to solar and geomagnetic activity. Dual-frequency GPS receivers actually measure the effect of the ionosphere on the GPS signals and can better adjust to, but not eradicate, these difficult circumstances. This is accomplished by using a network of fixed ground based GPS receivers, separated by a few hundred km, to derive a map of the ionosphere. The map is then transmitted to the aircraft so that the GPS receiver on board can make an accurate ionospheric correction.

On a smaller scale, irregularities in the density of the ionosphere that produce scintillations occur in varying amounts, depending on latitude. For example, the equatorial region, (the latitude zone that spans 15-20° either side of the magnetic equator) is the site of some of the greatest ionospheric irregularities, even when magnetic storms do not occur. Seemingly unpredictable episodes of density enhancements in the upper ionosphere can occur there in the evening hours and can cause radio waves to be misdirected. These scintillations make GPS operations difficult.

GPS signals are generally immune to ionospheric changes in response to large infusions of x-rays following a solar flare. However, GPS and all other satellites (including communications) must contend with all the detrimental effects of the SW environment: surface and deep dielectric charging, total radiation dose and SEEs. Many of these effects are continuous throughout the solar cycle, but some, in particular SEEs, may become more pronounced during the next solar maximum as the satellites will have been designed to new standards. They are likely to be smaller and lighter; requiring them to have less shielding, less redundancy and more compact electronic components. The satellites designed to support the global airspace system, like the aircraft themselves, will have become more susceptible to the SW environment.

3.5 Terrestrial Weather

Besides the ionospheric disturbances directly caused by flares and magnetic storms, the ionosphere exhibits irregular variations related to the dynamics of the underlying atmosphere. These depend upon the combination of traditional "weather" near the ground, which produces waves in the

atmosphere like the waves in the deep ocean, and the winds between the ground and the upper-atmosphere levels that act like a filter to the passage of those waves. While this aspect of SW may appear to have a non-solar origin, its effects are most pronounced when the upper-atmosphere winds or lower-ionosphere composition is enhanced by the energy inputs from the active Sun.

Optical phenomena called “red sprites” and “blue jets” have been observed at altitudes extending from the tops of strong thunderstorms (at around 15-kilometers altitude) to the lower ionosphere (about 95-km altitude). Possibly related to these optical signatures, intense electromagnetic pulses (10,000 times stronger than lightning-related pulses) have been detected over thunderstorm regions by satellites. These observations suggest that there may be a stronger connection between global thunderstorm activity and the ionosphere and upper atmosphere than previously suspected. Interest in their effects will depend on the future use of this region of Earth-space.

The effect of terrestrial weather on current aircraft operations still has the greater impact or risk when compared with SW effects. Thunderstorms, microbursts, hurricanes, icing, jetstreams, fog, etc., all impact the commercial operation as well as the safety. The aviation community is very well versed at planning around terrestrial weather: many scientific tools and technological systems are available to mitigate the risks. However, the importance of the solar cycle and variability upon the Earth’s climatology is only just beginning to be considered as an integral part of the study of the Sun-Earth system. Through these studies it is hoped to establish the mechanisms that may link solar variability with terrestrial weather, thereby including the need for regular SW information in any operational brief.

4. TECHNOLOGICAL DEVELOPMENTS

The future picture of aircraft operations, whether civil or military, is of one comprising a global multi-network infrastructure that provides for future growth, but ensures safety and security. It will have a satellite-based, air space management system with global connectivity, offering better C^3I, “free-flight” routing for commercial aircraft and complete “space-to-mud” situational awareness for multi-national military operations. Combined with information technology, airspace and aircraft management will become strategic and predictable. As described in the preceding sections, the technology required for that capability means that all critical components of the operation will be at risk from SW. Table 3 provides examples of technological developments for future aircraft systems and operations.

Impact Area	Civil Aviation	Military Aviation
Radiation	Higher, Longer Flights	Medium / High Level Operations
Avionics	Electronic Flight Bags, Aircraft Systems, Passenger IFE	Human sensor integration, Electronic Mission Bags, Radars
Communications	Secure Data Link, SatCom, Internet	Secure Data Link, SatCom
SatNav	Reduced separation, Free-Flight	Improved position accuracy
Future C^3I Ops	Future Air Navigation System (FANS)	Network-Centric

Table 3. Examples of technological developments for future aircraft operations.

Electronic Flight Bags (EFB) or the "paper-less office" is already a reality for some pilots. Instead of briefing from charts and maps, they access everything needed for their flight from a laptop: airfield, weather and flight plan data are automatically downloaded; landing charts, aircraft and operational manuals can also be automatically updated. The next generation of EFBs will be permanent installations, networked to the aircraft flight management servers; the pilot will access the systems via a keyboard and mouse or touch screens. Eye and voice recognition are already in use in the latest military aircraft.

The passenger in-flight experience now includes satellite phones, multi-channel video and games and recently email access. Soon to arrive are wireless high speed internet connections, on demand digital video and satellite television broadcasts. All of this will be controlled by a central cabin environment server, probably utilising more microchip processing power than the aircraft flight systems.

The development and implementation of FANS for future civil operations incorporates the needs of Communications, Navigation, Surveillance/Air Traffic Management (CNS/ATM) into one global system. Achieving this safely will require maintaining system integrity of every component of that system.

Briefly consider the impact of a large SEP event on a global free-flight civil airspace structure. Some pilots may elect to descend to avoid the incoming radiation health hazard, likely causing other aircraft to alter course and altitude as their onboard navigation and flight management systems communicate directly between aircraft. Which aircraft components are hardened against SEEs? Collision avoidance systems, cockpit systems, engine controls?

In a similar way, the concept of Network-centric warfare relies upon the information sharing capabilities of a real-time global Command, Control, Communication, Computer and Intelligence, Surveillance and Reconnaissance (C^4ISR) system. In essence this means that overall mission success is no longer focused on the aircraft or weapon platform, but on to the network. Combining all military assets onto an information sharing net can dramatically increase the combat effectiveness, especially if the resources are drawn from many different nationalities. Such a system will improve mission and target co-ordination, navigation accuracy, and reduce system-induced friendly fire and collateral damage. However, reliance upon such an information rich network can leave an aircraft isolated or blind if the net goes down. Therefore, the system technology will require an increasing level of reliability and security, which will make it increasingly affected by the SW environment.

5. ASSESSING THE RISKS TO FUTURE OPERATIONS

Whether we are concerned about civil aircraft separation minima or military aircraft mission success, a chain of minor system failures induced by SW impacts could have potentially dire consequences. The role of any aviation Safety Management System (SMS) is to ensure such chains are broken before they break the aircraft. This means putting procedures and checks in place for all the identified hazards. However, to do this we must first assess the level of risk that SW events are likely to place upon technology and operations in the future.

The ESA Space Environments and Effects Analysis section of the European Space Agency has recently (April 2003) initiated a SW application-based pilot project. The aim of this pilot project is to extend the SW community in Europe, develop outreach activities, collaborations and key SW applications based on existing or easily adaptable sources of data. One project, which is part of ESA's Service Development Activities (SDA), is called "Space weather Operational Airline Risk Service" (SOARS). This is a collaboration between the Mullard Space Science Laboratory (University College London) and Virgin Atlantic Airways Ltd, a UK long-haul airline. Technical expertise is also being provided by ESYS, QinetiQ, the UK Met Office, the National Physical Laboratory and SolarMetrics Limited. The aim of the project is to assess the quantifiable risks and the impacts the airline industry can expect to see on current and future commercial operations. It will then develop a prototype service for the airlines through operational risk modelling.

To ensure that industry requirements are met for maintaining safe and secure operations, while at the same time reducing any commercial impacts of SW, will require the identification and assessment of all SW hazards and their impact on the different operational areas. Each SW hazard needs to be investigated and catalogued, and expressions of probability and severity combined to derive risk matrices, which can then be applied to the "at risk" elements of the operation. The "at risk" areas can be placed into 3 major categories: Engineering (safety critical/non-critical flight systems, passenger service systems), Flight Operations (Health, Flight Safety, C^3I,), FANS (CNS/ATM) and Network-Centric Ops. Within each of these categories it would be necessary to drill down to the individual core components to find the areas of critical weakness in the system or operation.

Thereafter, using these assessments, it should be possible to decide what SW data and information would be required upon which to build a prototype Airline Space Weather Service. In addition, these risk management exercises will also provide the necessary information to carry out a cost/benefit analysis for any SW service, which is required under a third phase of ESA's pilot project. It is envisaged that one potential source of required SW data could be a Space Weather European Network (SWENET), which in collaboration with other international resources could support a Global Space Weather Organisation (GSWO), analogous to, or even incorporated into, the World Meteorological Organisation (WMO).

However, there is one important proviso. The utilisation of SW information, if not properly implemented, controlled and standardised, could in fact be the cause of disruption and reduced safety to the operation due to pilots reacting in isolation to any SW warning. Aviation is one of the most heavily regulated industries, both nationally and internationally, in terms of its safety, security and operational procedures. The use of any SW information in a similar manner to terrestrial weather, should therefore, be co-ordinated, implemented and agreed to by the many world-wide governing bodies, from IATA and ICAO to Air Traffic Control and Aviation Regulatory Authority's.

And before any such bespoke services are introduced for aircraft operations, there is the need to develop a substantial Educational Outreach Programme (EOP) aimed specifically at the airline industry. The publicity and industry contacts generated by the SOARS project will go some way to begin this process, but only a European or international EOP will truly make any impact on the lack of awareness and understanding of the SW impacts on our technology-reliant society.

6. CONCLUSIONS

SW already affects aircraft operations, most notably the exposure of humans to cosmic radiation and any variations caused by SEP events. The SW hazards to current avionics, communication and navigation systems are also considered to be scientifically significant, although as yet there appears to be insufficient industry interest or understanding to support further research. However, with the introduction of any new ultra-long-haul, higher-flying, more electric aircraft, aviation is likely to see more of its safety margins being eroded by SW impacts. And with the plans for FANS (CNS/ATM) and Network-Centric operations reliant upon a global satellite infrastructure the SW risks will continue to increase proportionally with the technological developments.

Aviation is built upon one overriding principle: safety. Therefore, to ensure that safety (and security of operations) is maintained in the future, the industry must begin to utilise SW information, and make plans for it to be integrated safely into the daily operation. This should also include the requirement to have permanently installed active monitors' onboard aircraft to give almost worldwide, full-time coverage of SW events.

One of ESA's SW pilot projects, SOARS, intends to establish the industry requirements for bespoke SW services through operational risk modelling of all the SW hazards. This work should also provide a limited EOP to improve the awareness and understanding of the SW environment and its many impacts on aircraft operations.

7. ACKNOWLEDGEMENTS

I would like to thank Clive Dyer (QinetiQ), Roger Iles (SolarMetrics) for input to this paper. The image of Figure 7 is supplied courtesy of FAA/CAMI and the notional data supplied in Table 2 is provided courtesy of 9th RW ORM. I would also like to thank Virgin Atlantic Airways for their continued support of my research activities.

8. REFERENCES

Bartlett, D. T. Radiation protection concepts and quantities for the occupational exposure to cosmic radiation. *Rad. Prot. Dosim.* **86**, 263-268 (1999).

Bentley, R. D., R. H. A. Iles, J. B. L. Jones, et al., Monitoring cosmic radiation on aircraft, in press, *Solar Cycle and Space Weather,* S*OLSPA,* Naples, Italy, May 2001.

Council Directive 96/29/EURATOM of 13 May 1996: Laying down the basic safety standards for protection of the health of workers and the general public against the danger arising from ionising radiation. *Official Journal of the European Communities* **19**, L159, 1-114, Article 42 (June 1996).

DfT, UK Government, Protection of air crew from cosmic radiation: Guidance material http://www.dft.gov.uk/stellent/groups/dft_aviation/documents/page/dft_aviation_503468.hcsp

Dyer, C. S., Radiation effects on spacecraft and aircraft, in proceedings, *Solar Cycle and Space Weather,* S*OLSPA,* Naples, Italy, May 2001.

Dyer, C. S., F. Lei, Monte Carlo calculations of the influence on aircraft radiation environments of structures and solar particle events, presented at Nuclear and Space Radiation Effects Conference, July 2001.

Dyer, C. S., P. Truscott, C. Sanderson, B. Colwell, A. Chugg, R. Jones, I. MacDiarmid, K. Johansson, Cosmic radiation effects on avionics, an increasing hazard in the new millennium, presented at 22nd International Congress on Aeronautical Sciences, Harrogate, UK, September 2000.

Heavner, M.J., D.D. Sentman, D.R. Moudry, E.M. Wescott, C.L Siefring, J.S. Morrill, and E.J. Bucsela, "Sprites, Blue Jets, and Elves: Optical Evidence of Energy Transport Across the Stratopause," AGU Monograph 123 "Atmospheric Science Across the Stratopause," p69-82, 2000.

ICRP. 1990 Recommendations of the International Commission on Radiological Protection. *Publication 60* (Oxford: Pergamon Press) (1991).

Iles, R. H. A., R. D. Bentley, R. Hunter, J. B. L. Jones, G. C. Taylor, D. J. Thomas, L. K. Harra, A. J. Coates, The effect of solar particle events at aircraft altitudes, presented at *Looking Towards a Future European Space Weather Programme,* ESA-ESTEC Space Weather Workshop December 2001.

Jones, J. B. L., R. D. Bentley, R. Hunter, R. H. A. Iles, G. C. Taylor, D. J. Thomas, The practical issues of utilising a European space weather programme for airline operations, presented at *Looking Towards a Future European Space Weather Programme,* ESA-ESTEC Space Weather Workshop December 2001.

Jones, J. B. L., R. D. Bentley, R. Hunter, R. H. A. Iles, G. C. Taylor, D. J. Thomas, Space weather and commercial airlines, in press, *COSPAR 2002*, Houston, USA, October 2002

Lewis, B. J., M. J. McCall, A. R. Green, L. G. I. Bennett, M. Pierre, U. J. Schrewe, K. O'Brien, E. Felsberger, Aircrew exposure from cosmic radiation on commercial airline routes. *Radiat. Prot. Dosim.* **93**, 293-314 (2001).

Chapter 13

Ground Effects of Space Weather

Space weather effects on electric power transmission grids and pipelines

Risto Pirjola, Ari Viljanen, Antti Pulkkinen, Sami Kilpua*, Olaf Amm

Finnish Meteorological Institute, Geophysical Research Division
P. O. Box 503, FIN-00101 Helsinki, Finland

*Now at:
GeoForschungsZentrum Potsdam
Telegrafenberg, D-14473 Potsdam, Germany

Abstract Space storms produce geomagnetically induced currents (GIC) in technological systems at the Earth's surface, such as electric power transmission grids, pipelines, communication cables and railways. Thus GIC are the ground end of the space weather chain originating from the Sun. The first GIC observations were already made in early telegraph equipment about 150 years ago, and since then several different systems have experienced problems during large magnetic storms. Physically, GIC are driven by the geoelectric field induced by a geomagnetic variation. The electric and magnetic fields are primarily created by magnetospheric-ionospheric currents and secondarily influenced by currents induced in the Earth that are affected by the ground conductivity. The most violent magnetic variations occur in auroral regions, which indicates that GIC are a particular high-latitude problem but lower-latitude systems can also experience GIC problems. In power networks, GIC may cause saturation of transformers with harmful consequences extending from harmonics in the electricity to large reactive power consumption and even to a collapse of the system or to permanent damage of transformers. In pipelines, GIC and the associated pipe-to-soil voltages can enhance corrosion and disturb corrosion control measurements and protection. Modelling techniques of GIC are discussed in this paper. Having information about the Earth's conductivity and about space currents or the ground magnetic field, a GIC calculation contains two steps: the determination of the geoelectric field and the computation of GIC in the

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 235-256.

system considered. Generally, the latter step is easier but techniques applicable to discretely-earthed power systems essentially differ from those usable for continuously-earthed buried pipelines. Time-critical purposes, like forecasting of GIC, require a fast calculation of the geoelectric field. A straightforward derivation of the electric field from Maxwell's equations and boundary conditions seems to be too slow. The complex image method (CIM) is an alternative but the electric field can also be calculated by applying the simple plane wave formula if ground-based magnetic data are available. In this paper, special attention is paid to the relation between CIM and the plane wave method. A study about GIC in Scotland and Finland during the large geomagnetic storm in April 2000 and another statistical study about GIC in Finland during SSC events are also briefly discussed.

Keywords Geomagnetically induced currents, GIC, geoelectric field, geomagnetic disturbances, geoelectromagnetics, plane wave, complex image method

1. INTRODUCTION

At the Earth's surface, space weather manifests itself as geomagnetically induced currents (GIC) flowing in long conductors, such as electric power transmission networks, oil and gas pipelines, telecommunication cables and railways systems. In power grids, GIC cause saturation of transformers, which tends to distort and increase the exciting current. It in turn implies harmonics in the electricity, unwanted relay trippings, large reactive power consumption, voltage fluctuations etc., leading finally to a possible black-out of the whole system, and to permanent damage of transformers (Kappenman and Albertson, 1990; Kappenman, 1996; Erinmez et al., 2002b; Molinski, 2002).

In buried pipelines, GIC and the associated pipe-to-soil voltages contribute to corrosion and disturb corrosion control surveys and protection systems (Boteler, 2000; Gummow, 2002). Telecommunication devices have also experienced GIC problems (Karsberg et al., 1959; Boteler et al., 1998; Nevanlinna et al., 2001). As optical fibre cables do not carry GIC, space weather risks on telecommunication equipment are probably smaller today than previously. However, it should also be noted that metal wires are used in parallel with optical cables for the power to repeat stations. There are not many studies of GIC effects on railways, and to the knowledge of the authors of this paper, the only publicly and clearly documented case has occurred in Sweden where GIC resulted in misoperation of railway traffic lights during a geomagnetic storm in July 1982 (Wallerius, 1982). (A private communication with a Russian scientist indicates that space weather has caused problems in Russian railway systems, too.)

GIC have a long history since the first observations were already made in early telegraph systems about 150 years ago (Boteler et al., 1998). In general, GIC is a high-latitude problem, which is supported by the fact that the most famous destructive GIC event occurred in the Hydro-Québec power system in Canada (Czech et al., 1992; Bolduc, 2002). GIC values in a system are, however, not directly related to the proximity of the auroral zone but the ground resistivity and the particular network configuration and resistances also affect. GIC values usually greatly vary from site to site and from system to system. Furthermore, GIC magnitudes that are a potential risk for a power transmission system are highly dependent on transformer structures and on engineering details of the network. For example, the largest GIC measured in the Finnish 400 kV power system is about 200 A but Finland's transformers have not experienced GIC problems (Elovaara et al., 1992; Lahtinen and Elovaara, 2002). Probably, the largest GIC anywhere and ever measured is 320 A in Sweden during the geomagnetic storm in April 2000 (Erinmez et al., 2002b). The value of 600 A in Sweden mentioned by Stauning (2002) is evidently not correct (private communication with a Swedish engineer).

There are engineering means which may be used for preventing harmful GIC in a system. For example, the dc-like GIC cannot flow through series capacitors installed in power transmission lines. However, determining the locations of capacitors in a power grid is not straightforward (Erinmez et al., 2002a; Pirjola, 2002). Thus, the flow of GIC cannot easily be blocked in a system, and efforts should be concentrated on estimating expected GIC magnitudes at different sites and on forecasting large GIC events.

The horizontal geoelectric field induced at the Earth's surface drives GIC. Therefore, model developments in GIC research should aim at calculating the geoelectric field. After knowing this field, the determination of GIC in a system is a simpler task although a discretely-earthed power grid and a continuously-earthed buried pipeline require different techniques (Lehtinen and Pirjola, 1985; Pulkkinen et al., 2001).

As described by Faraday's law, the geoelectric field is induced by a temporal variation of the magnetic field during a geomagnetic disturbance or storm. Both the magnetic field and the electric field are primarily produced by ionospheric-magnetospheric currents, but they also have a secondary contribution from currents in the Earth affected by the Earth's conductivity structure. In principle, knowing the space currents and the Earth's conductivity permits the determination of the electric and magnetic fields at the Earth's surface by using Maxwell's equations and appropriate boundary conditions. Such a straightforward method is presented by Häkkinen and Pirjola (1986). In practice, however, the ionospheric-magnetospheric currents and the conductivity of the Earth are not known precisely, and even if they were known, the exact formulas would not allow fast enough

computations needed for forecasting purposes. The complex image method (CIM) has shown to be a suitable technique for geoelectromagnetic calculations because it is accurate and fast (Boteler and Pirjola, 1998; Pirjola and Viljanen, 1998).

The simplest relation between surface electric and magnetic fields is obtained by making the plane wave assumption, which rigorously means that the primary electromagnetic field originating from space current is a vertically-downwards propagating plane wave. Assuming further that the Earth has a layered structure and operating in the frequency domain, the electric field is obtained by multiplying the magnetic field by the surface impedance. It has been shown that the assumption of a vertical plane wave need not be strictly fulfilled for the plane wave technique to work in practice (Cagniard, 1953; Wait, 1954; Dmitriev and Berdichevsky, 1979; Wait, 1980). Thus, the plane wave method provides a good tool for the calculation of the geoelectric field if magnetic data are available.

In Section 2, we summarize the methods to be used for determining the geoelectric field and for computing GIC. Special attention is paid to the relation between the plane wave technique and CIM. There are a great variety of different space current systems which can produce a significant magnetic disturbance, a geoelectric field and GIC in technological systems. The spherical elementary current system (SECS) method is a novel useful tool for determining (equivalent) ionospheric currents from ground magnetic observations during different space weather events (Amm, 1997; Amm and Viljanen, 1999; Pulkkinen et al., 2003a). A step forwards in understanding GIC processes and forecasting them is to classify space weather events by considering their GIC impacts. In Section 3, we summarize a study of GIC during the large magnetic storm in April 2000, and briefly discuss observations of GIC in the Finnish natural gas pipeline during sudden storm commencements (SSC), which are global geomagnetic disturbances.

2. MODELLING THE GEOELECTRIC FIELD AND GIC

2.1 Calculation of the geoelectric field

2.1.1 Plane wave model

GIC are usually considered in systems located in a limited area. Therefore models used in this connection have a regional character permitting the use of a flat-Earth model. The standard coordinate system has its *xy* plane at the

Earth's surface with the x , y and z axes pointing northwards, eastwards and downwards, respectively. Let us assume that the primary electromagnetic field originating from ionospheric and magnetospheric sources is a plane wave propagating vertically downwards and that the Earth is uniform with a permittivity ε, a permeability μ and a conductivity σ. Considering a single frequency ω (i.e. the time dependence is $\exp(i\omega t)$), it is simple to derive the following relation between the y component of the electric field E_y and the x component of the magnetic field B_x at the Earth's surface (or similarly between E_x and B_y):

$$E_y = -\frac{\mu\omega}{\mu_0 k} B_x \tag{1}$$

where the propagation constant k is given by

$$k = \sqrt{\omega^2 \mu\varepsilon - i\omega\mu\sigma} \tag{2}$$

In geoelectromagnetics always $\sigma >> \omega\varepsilon$, and μ can be set equal to the vacuum value μ_0. Thus

$$k = \sqrt{-i\omega\mu_0\sigma} \tag{3}$$

and

$$E_y = -\sqrt{\frac{\omega}{\mu_0\sigma}} e^{i\frac{\pi}{4}} B_x \tag{4}$$

Inverse-Fourier transforming equation (4) into the time domain yields

$$E_y(t) = -\frac{1}{\sqrt{\pi\mu_0\sigma}} \int_0^{\infty} \frac{g(t-u)}{\sqrt{u}} du \tag{5}$$

$$= -\frac{1}{\sqrt{\pi\mu_0\sigma}} \int_{-\infty}^{t} \frac{g(u)}{\sqrt{t-u}} du$$

where the time derivative of $B_x(t)$ is denoted by $g(t)$. Equations (4) and (5) show that the electric field decreases with an increasing Earth conductivity. This indicates that GIC should be taken into account in particular in resistive

areas. This conclusion is, however, not self-evident since the ground conductivity also has an influence on earthing resistances of a power system thus affecting the GIC flow (Pirjola and Viljanen, 1991). It is seen from equation (5) that the electric field at a given moment t is not only related to the time derivative of the magnetic field at the same moment but earlier values also affect with a decreasing weight (the square root factor in the denominator). The inverse-Fourier transform may, of course, be performed for the exact formula (1) as well leading to an expression which approximately reduces to equation (5) (Pirjola, 1982, p. 23).

If the Earth is not uniform but has a layered structure the term $\mu\omega/k$ in (1) has to be replaced by the (plane wave) surface impedance $Z = Z(\omega)$ (see e.g. Wait, 1981, pp. 43-55), so that

$$E_y = -\frac{Z}{\mu_0} B_x \tag{6}$$

If the Earth's structure also depends on the x and y coordinates, as for example in coastal areas, the situation becomes much more complicated, and the independence of x and y of the fields disappears.

Equations (4) and (6) form the basis of the magnetotelluric sounding method of the conductivity structure of the Earth (Cagniard, 1953). A lot of discussion has concerned the validity of the plane wave assumption of the primary field (e.g. Mareschal, 1986; Pirjola, 1992) since a vertical plane wave is certainly not true near a concentrated ionospheric current, like an auroral electrojet. It, however, appears that the magnetotelluric equations (4) and (6) are applicable to a wide range of events.

2.1.2 Transfer function between horizontal electric and magnetic fields

Let us now assume that the primary electromagnetic field incident on a uniform Earth depends on the x coordinate (but for simplicity not on y), which is the case if, for example, an auroral electrojet is modelled by an east-west line current. Considering a single frequency ω and making a spatial Fourier transform from the x coordinate to the wavenumber b, equation (6) is satisfied with

$$Z = Z(b,\omega) = \frac{i\omega\mu}{\sqrt{b^2 - k^2}} \tag{7}$$

where k is given by formula (2) (Pirjola, 1982, p. 51). Setting b equal to zero in equation (7) gives the surface impedance included in equation (1) as expected.

Inverse-Fourier transforming into the x domain yields

$$E_y(x,\omega) = \int_{-\infty}^{\infty} E_y(b,\omega) e^{ibx}\, db = -\frac{1}{\mu_0} \int_{-\infty}^{\infty} Z(b,\omega) B_x(b,\omega) e^{ibx}\, db \tag{8}$$

Using the convolution theorem, this can be written as

$$E_y(x,\omega) = -\frac{1}{\mu_0} \int_{-\infty}^{\infty} Z(x-x') B_x(x',\omega)\, dx' \tag{9}$$

$$= -\frac{1}{\mu_0} \int_{-\infty}^{\infty} Z(x') B_x(x-x',\omega)\, dx'$$

where the transfer function is given by

$$Z(x) = Z(x,\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(b,\omega) e^{ibx}\, db \tag{10}$$

Equations (8), (9) and (10) do not presume that the Earth is uniform but any layered structure is possible. The reference Wait (1981, pp. 43-55) mentioned in Section 2.1.1 is associated with a wavenumber-dependent surface impedance, i.e. not only with the plane wave case. The treatment included in equations (8), (9) and (10) is analogous to that used for calculating the electric field at the seafloor in terms of the surface magnetic field in the two-dimensional case (Pirjola et al, 2000).

In the case of a uniform Earth, a substitution of equation (7) into (10) gives

$$Z(x) = \frac{\omega\mu}{2} H_0^{(2)}(kx) \tag{11}$$

where $H_0^{(2)}$ denotes the Hankel function of the second kind and of the zeroth order (see e.g. Pirjola, 1982, p. 128). Fig. 1 shows $|Z(x)|$ in normalized units for the following three period-conductivity combinations: 300 s / 10^{-3} Ω^{-1}m^{-1}; 300 s / 10^{-2} Ω^{-1}m^{-1}; 30 s / 10^{-2} Ω^{-1}m^{-1}. (In absolute units, the largest values of $|Z(x)|$ obtained when x is small are much higher for the smaller

period 30 s than for 300 s while changing the conductivity from 10^{-3} $\Omega^{-1}m^{-1}$ to 10^{-2} $\Omega^{-1}m^{-1}$ does not matter so much.)

Fig. 1 suggests that an increase of the conductivity and a decrease of the period, implying an increase of $|k|$, make the transfer function more concentrated for small values of x. This indicates that, for small values of the period and for highly-conducting Earth structures, $Z(x)$ approaches the Dirac delta function $\delta(x)$ (multiplied by a constant Z_A). Furthermore, it seems clear that this conclusion is not limited to the case of a uniform Earth but is valid more generally, too.

Substituting $Z(x) = Z_A\delta(x)$ into equation (9) yields

$$E_y(x,\omega) = -\frac{Z_A}{\mu_0} B_x(x,\omega) \tag{12}$$

Consequently, comparing with equation (6), E_y and B_x locally, i.e. for a given x value, satisfy the plane wave relation, and the constant Z_A may clearly be identified with the local plane wave surface impedance. The delta function form of $Z(x)$ is also obtained from equation (10) by assuming that $Z(b,\omega)$ can be regarded as independent of b, which is another way to express the plane wave assumption.

It has been shown by Dmitriev and Berdichevsky (1979) that the plane wave formula (12) is true if the electric and magnetic fields are linear functions of the x and y coordinates in the area considered. This provides an extension to the validity of the plane wave model for calculating the geoelectric field. The issue is also discussed in detail by Pulkkinen (2003).

In practice, the area under consideration, e.g. that occupied by a power grid, is convenient to be divided into blocks each of which has a layered-Earth conductivity structure. Using the plane wave relation (12) for each block separately makes Z_A vary from site to site, i.e. it depends on the x coordinate (and more generally also on y). Now it is important to note that this dependence, of course, does not make Z_A equal the transfer function $Z(x)$ included in equations (9) and (10).

2.1.3 Complex image method

The basic idea of the complex image method (CIM) is to replace the real Earth by a perfect conductor located at a complex depth, which thus enables the calculation of the secondary contribution to surface electric and magnetic fields simply by considering a mirror image of the primary ionospheric-magnetospheric source current. CIM was introduced to geoelectromagnetic applications already by Thomson and Weaver (1975). However, CIM was

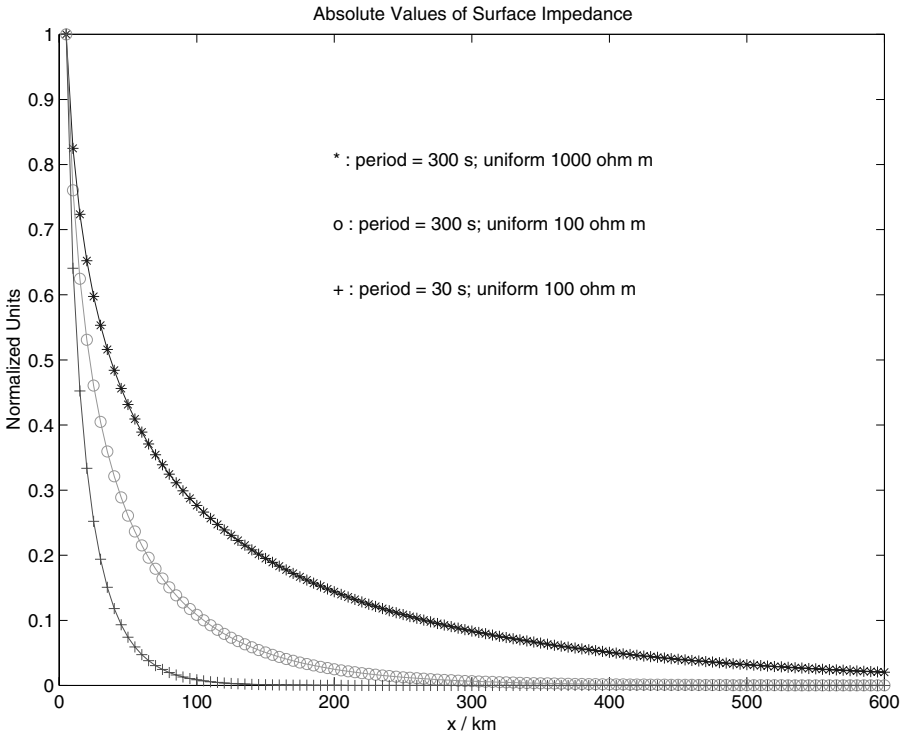

Figure 1. Normalized absolute values of the transfer function $Z(x)$ between the horizontal magnetic and electric fields as functions of the x coordinate. The Earth is uniform with a conductivity 10^{-3} $\Omega^{-1}m^{-1}$ or 10^{-2} $\Omega^{-1}m^{-1}$, and the periods considered are 300 s and 30 s.

not used in connection with GIC studies until the discussions by Boteler and Pirjola (1998) and Pirjola and Viljanen (1998). Considering a single frequency ω, the depth of the perfect conductor is given by the complex skin depth $p = p(\omega)$ defined by

$$p = \frac{Z}{i\omega\mu_0} \tag{13}$$

where $Z = Z(\omega)$ is the plane wave surface impedance included in equation (6).

Thomson and Weaver (1975) derive CIM for an arbitrary horizontal divergence-free current distribution above the Earth's surface. Boteler and Pirjola (1998) provide a detailed validation of CIM in case of a horizontal infinitely long line current above a layered Earth, and in particular, demonstrate that CIM requires that $|pb|$ is so small that $(pb)^3 \approx 0$ where b is a

characteristic wavenumber of the primary field. Pirjola and Viljanen (1998) generalize CIM for an U-shaped current, i.e. a horizontal current of a finite length with vertical currents at its ends. At high-latitudes vertical currents are a good approximation of geomagnetic-field-aligned currents. Setting U-shaped currents on an ionospheric grid enables the construction of any ionospheric currents distribution, so CIM is applicable to studies of complicated space weather events. The spherical elementary current system (SECS) method referred to above can be used for investigating ionospheric currents based on ground magnetic observations.

A crucial point in the discussion by Pirjola and Viljanen (1998) is the proof of the equivalence of a vertical current with a horizontal current distribution, which makes it possible to utilize the CIM result by Thomson and Weaver (1975) concerning horizontal currents. In other words, Pirjola and Viljanen (1998) argue that the current system depicted in Fig. 2 and having a time dependence $\exp(i\omega t)$ produces no magnetic field and no horizontal electric field at the surface of a layered Earth. This result is true within the neglect of the displacement currents, which is an acceptable approximation in geoelectromagnetics. From the theoretical viewpoint, it is important to note that neither the primary current system nor the secondary induced current system creates a magnetic field at the Earth's surface while both create an electric field but the horizontal components of these two electric contributions (practically) cancel each other at the surface. The fact that the current shown in Fig. 2 has no magnetic effect below it is a well-known result by Fukushima (1976). An aim of future research should be to consider the equivalence more generally and to try to construct a primary current (and charge) system which alone would not produce any horizontal electric field at the Earth's surface.

2.1.4 Relation between CIM and the plane wave model

Fig. 3 shows a comparison between computations of the east component of the electric field at the Earth's surface based on the CIM and on the plane wave method. One hour during the large magnetic storm on July 15, 2000, is considered in the figure. The layered Earth model used corresponds to central Finland with the following layer thicknesses and resistivities: [12, 22, 16, 50, 50, ∞] km and [30000, 3000, 50, 1000, 5000, 1] Ωm. Since both CIM and the plane wave formula operate in the frequency domain, Fast Fourier Transforms were applied to obtain the electric field curves as functions of time. The calculations are based on first determining ionospheric equivalent currents from ground magnetic data by utilizing the spherical elementary current system (SECS) method (Amm, 1997; Amm and Viljanen, 1999; Pulkkinen et al., 2003a). It would, of course, be interesting and important to

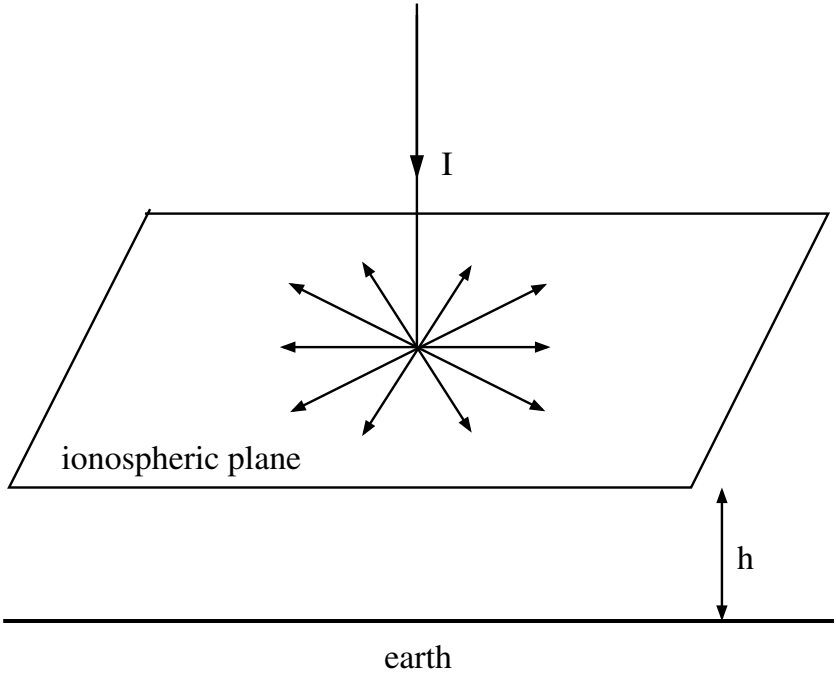

Figure 2. Current system consisting of a vertical part and a radial horizontal distribution in the ionosphere at the height *h* above the Earth's surface. The system does not cause any magnetic or horizontal electric field at the surface of a layered Earth (Pirjola and Viljanen, 1998).

compare the modelled curves shown in Fig. 3 with actual measured data. However, such recordings are not available for the particular site and event considered, so a validation against real data for the two models remains to be an objective of future studies.

The excellent agreement between the two curves in Fig. 3 suggests that CIM and the plane wave method have a close relation. This will now be investigated theoretically.

Let us consider a horizontal divergence-free current distribution at the height *h* above the Earth's surface:

$$\mathbf{j}(x,y,z) = (j_x(x,y)\mathbf{e_x} + j_y(x,y)\mathbf{e_y})\delta(z+h) \qquad (14)$$

where the time factor $\exp(i\omega t)$ is not written explicitly. We assume that the Earth is characterized by the complex skin depth p as expressed by equation (13). By utilizing CIM, the contribution from induced currents to the surface fields is obtained by removing the Earth and assuming a current opposite to the current of equation (14) lying at the complex depth $z = h + 2p$. Hence, the vector potential **A** has the form

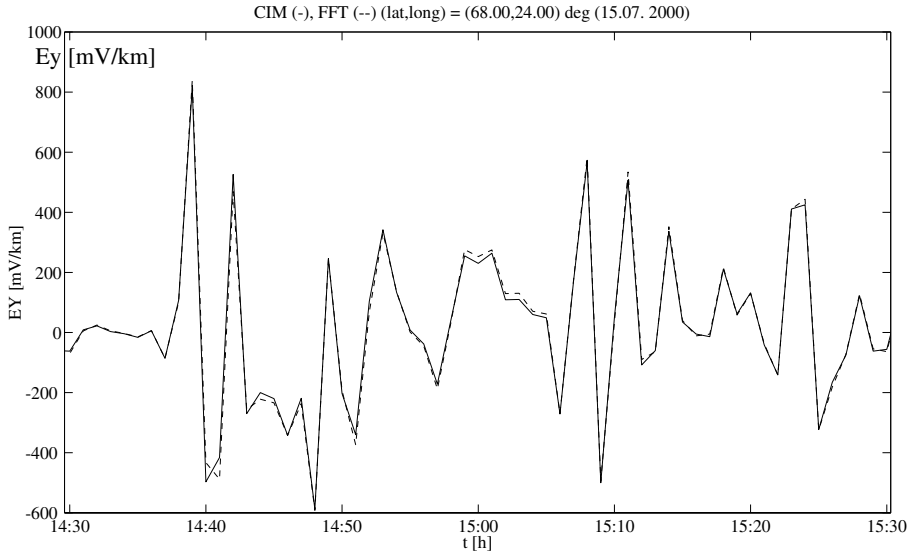

Figure 3. East component of the electric field at a site in northern Finland during one hour on July 15, 2000. Solid line (labelled by "CIM" on the top): calculated based on the complex image method. Dashed line (labelled by "FFT" on the top): based on the plane wave model by using the CIM magnetic field as the input. The Earth model corresponds to central Finland representing a resistive structure (see the text).

$$\mathbf{A}(x,y,z) = A_x(x,y,z)\mathbf{e_x} + A_y(x,y,z)\mathbf{e_y} \tag{15}$$

where

$$A_{x/y}(x,y,z) = f_{x/y}(x,y,z+h) - f_{x/y}(x,y,z-(h+2p)) \tag{16}$$

and f_x and f_y are functions related to and determined by the current components j_x and j_y. Since, according to the above assumption, $\nabla\cdot\mathbf{j} = 0$ there are no charges, and so the scalar potential is zero. Thus

$$\mathbf{E}(x,y,z) = -\frac{\partial \mathbf{A}}{\partial t} = -i\omega\mathbf{A}(x,y,z) \tag{17}$$

and

$$\mathbf{B}(x,y,z) = \nabla \times \mathbf{A}(x,y,z) \tag{18}$$

We now consider the y component of the electric field at the Earth's surface (the treatment of E_x would be similar):

$$E_y(Surface) = -i\omega(f_y(x,y,h) - f_y(x,y,-h-2p)) \tag{19}$$

The tangential component of an electric field always vanishes at the surface of a perfect conductor, so $E_y(x,y,z=p) = 0$. (Note that CIM is actually a mathematical trick implying, e.g., the possibility of putting an equal sign here between the real coordinate z and the complex number p.) Thus, using the relation $E_y(x,y,z=p) = 0$ in formulas (16) and (17) permits writing equation (19) as

$$E_y(Surface) = -i\omega(D_2 - D_1) \tag{20}$$

where

$$D_1 = f_y(x,y,h+p) - f_y(x,y,h) \tag{21}$$

and

$$D_2 = f_y(x,y,-h-p) - f_y(x,y,-h-2p) \tag{22}$$

Let us now assume that $|p|$ is small enough to allow Taylor expansions of D_1 and D_2:

$$D_1 = p\frac{\partial f_y(x,y,z+h)}{\partial z}\Big|_{z=0} + \frac{1}{2}p^2\frac{\partial f_y{}^2(x,y,z+h)}{\partial z^2}\Big|_{z=0} + O(p^3) \tag{23}$$

and

$$D_2 = p\frac{\partial f_y(x,y,z-(h+2p))}{\partial z}\Big|_{z=0} + \frac{1}{2}p^2\frac{\partial f_y{}^2(x,y,z-(h+2p))}{\partial z^2}\Big|_{z=0} + O(p^3) \tag{24}$$

In a non-conducting medium and neglecting the displacement currents, the z dependence of the fields is given by $\exp(\pm\kappa_0 z)$ where $\kappa_0 = \sqrt{b^2+q^2}$ with

b and q being the wavenumbers associated with the x and y coordinates (see e.g. Pirjola, 1982, pp. 79-80). Equations (23) and (24) then reduce to

$$D_1 = \pm p\kappa_0 f_y(x,y,h) + \frac{1}{2}(p\kappa_0)^2 f_y(x,y,h) + O((p\kappa_0)^3) \tag{25}$$

and

$$D_2 = \pm p\kappa_0 f_y(x,y,-(h+2p)) \tag{26}$$

$$+\frac{1}{2}(p\kappa_0)^2 f_y(x,y,-(h+2p)) + O((p\kappa_0)^3)$$

Assume now that $|p|$ is so small that for all relevant values of b and q the term proportional to $(p\kappa_0)^2$ and the higher-order terms are negligible in (25) and (26). It is seen from equation (13) that a small value of $|Z|$ makes $|p|$ small as well. An increase of the Earth's conductivity tends to decrease $|Z|$, so the present approximation holds true especially for highly-conducting areas. Going back to the derivatives included in equations (23) and (24), we now obtain from equations (20), (25) and (26)

$$E_y(Surface) = -i\omega p\left(\frac{\partial f_y(x,y,z-(h+2p))}{\partial z}\right. \tag{27}$$

$$\left. -\frac{\partial f_y(x,y,z+h)}{\partial z}\right)\Big|_{z=0}$$

Using equations (13), (16) and (18), formula (27) can be written as

$$E_y(Surface) = -i\omega p\left(-\frac{\partial A_y}{\partial z}\right)\Big|_{z=0} = -i\omega p(\nabla \times \mathbf{A})_x\Big|_{z=0} \tag{28}$$

$$= -i\omega p B_x(Surface) = -\frac{Z}{\mu_0} B_x(Surface)$$

We have thus derived the plane wave relation (6) starting from CIM. Consequently, the good agreement between the CIM and plane wave curves in Fig. 3 also has a theoretical argument.

A crucial point in the above derivation of the plane wave relation (28) from CIM (16) is the assumption that $|p|$ is sufficiently small to make $(p\kappa_0)^2$ negligible. The derivation of CIM seems to have a similar, but not exactly the same, requirement, i.e. $(p\kappa_0)^3$ must be ignored, as is explicitly demonstrated by Boteler and Pirjola (1998) in the case of a line current source (note that $q = 0$ so κ_0 that equals b then.) This further couples CIM and the plane wave method together. A final remark about the correspondence of these two techniques is the observation that the derivation of CIM necessarily presumes p to be independent of the wavenumber (whereas it originally depends on the wavenumber since Z in equation (13) may be a function of the wavenumber) and that the wavenumber-independent plane wave value is the natural choice for p (Pirjola and Boteler, 2002).

2.2 Calculation of GIC

2.2.1 Power system

A power system is a network consisting of earthed nodes (transformer stations) that are connected to each other by conductors (transmission lines). Such a grid can be described by an earthing impedance matrix and a network admittance matrix (Lehtinen and Pirjola, 1985). In GIC computations the characteristic frequencies are so small that the matrices are real and only depend on resistances. The geoelectric field is incident on the network and the resulting currents (= GIC) flowing into and from the ground and along the conductors are obtained in a straightforward manner from matrix equations presented by Lehtinen and Pirjola (1985).

However, there are a couple of issues that require to be emphasized. Firstly, the geoelectric field is generally not a potential field, which implies that no single-valued "Earth-surface potential" exists and the geovoltage between two points at the Earth's surface depends on the path along which the geoelectric field is integrated (see Pirjola, 2000). Secondly, power grids are three-phase systems, so that the lines between nodes actually consist of three parallel conductors, and the node is the area where the conductors contact the transformer windings. The earthing resistance of a node is the sum of the (total) resistance of the windings, of the resistance of a possible neutral point reactor and of the actual grounding resistance of the station. Special care is needed when there are several transformers in parallel, when autotransformers are included, etc. (Mäkinen, 1993; Pirjola, 2003).

2.2.2 Pipeline

When calculating GIC and the accompanying pipe-to-soil voltages, a buried pipeline is convenient to be handled as a transmission line. Its parallel impedance per unit length Z is given by the resistance of the metallic pipeline and the transverse admittance per unit length Y is determined by the properties of the insulating coating covering the pipeline. Important parameters are the characteristic impedance ($= \sqrt{Z/Y}$) and the propagation constant ($= \sqrt{ZY}$). The inverse of the propagation constant (typically in the order of tens of km) gives the adjustment distance, i.e. the size of the area near an inhomogeneity of the system where significant pipe-to-soil voltages are expectable.

The geoelectric field affecting the pipeline network everywhere is a distributed source. This means that a GIC calculation requires the application of the distributed-source transmission line (DSTL) theory (Pulkkinen et al., 2001).

Inhomogeneities of a pipeline, like bends, changes of the pipeline material or of the pipeline size, and branches of the pipeline network, are important regarding corrosion issues. This means that models in which the pipeline is approximated by an infinitely long cylinder are not appropriate. In the DSTL theory, inhomogeneities of a pipeline network are convenient to be treated by applying Thévenin's theorem, which enables going through a whole pipeline network section by section (Pulkkinen et al., 2001). An algorithm applicable to model computations of GIC and pipe-to-soil voltages in a complicated pipeline systems is also presented by Pulkkinen et al., (2001).

3. GIC EFFECTIVENESS OF SPACE WEATHER EVENTS

It has been often stated that the east-west auroral electrojet is the most important ionospheric current producing ground effects of space weather in auroral regions. It is true that the magnetic north component (B_x) is statistically clearly larger than the east component (B_y) indicating the significance of the electrojet (Viljanen, 1997; Viljanen et al., 2001). However, the time derivatives $\mathrm{d}B_x/\mathrm{d}t$ and $\mathrm{d}B_y/\mathrm{d}t$, which play an important role for the geoelectric field and GIC, are roughly of an equal magnitude. This indicates that the large-scale electrojet is accompanied by smaller-scale rapidly-varying north-south and east-west currents. This further shows that a model of a mere electrojet simulated by an infinitely long line or sheet

current is not satisfactory regarding GIC estimation. Equal magnitudes of dB_x/dt and dB_y/dt also imply that the north and east components of the geoelectric field are statistically equal, too, so that the common statement that east-west power lines and pipelines would be more prone to GIC problems is incorrect (Pirjola, 2000).

Pulkkinen et al. (2003b) present a detailed study about GIC observations at three sites in the Scottish power grid, at two sites in the Finnish power system and at a site in the Finnish natural gas pipeline during the large geomagnetic storm in April 2000. The area covered is thus of a large regional scale. Most of the highest GIC values can be identified with substorm intensifications but no clear characteristics to be associated with all peak GIC values are found. Both localized ionospheric current systems as well as larger-scale propagating structures are seen. Only the sudden storm commencement (SSC) at the beginning of the event produced a simultaneous GIC at all sites. Pulsations were also drivers of GIC. The GIC magnitudes varied from site to site with the largest values observed being about 20 A during this particular storm. The durations of the peak values of GIC were typically in the order of minutes, which is also a piece of information important when considering possible harmful GIC impacts.

SSCs are sometimes considered particularly significant from the GIC point of view. This is supported, for example, by the observation that one of the largest GIC values measured in Finland (175 A / 10 s mean value) occurred at a 400 kV transformer in northern Finland during an SSC event on March 24, 1991 (Viljanen and Pirjola, 1994). This event started a storm at a later stage of which the largest GIC ever observed in Finland (201 A / one-minute mean value) was recorded at a 400 kV transformer in southwestern Finland.

Recently we have performed a study in which GIC flowing in the Finnish natural gas pipeline and recorded since November 1998 were considered during SSC events (Kilpua, 2003). After a careful selection, the analysis contained 79 events, all of which thus represent the sunspot maximum time. The largest GIC observed in the pipeline (32 A / 10 s mean value on November 6, 2001) was due to an SSC and was thus involved in the study. SSCs are usually not seen clearly in the midnight sector, which explains the fact that no GIC event during UT hours from 21 to 24 was included in the study. (The local time is UT plus two hours.) Furthermore, the UT distribution of SSC/GIC events clearly differs from the diurnal distribution of all GIC events. Thus, SSCs do not dominate when considering GIC statistically, which is in agreement with the rarity of SSCs. By investigating ionospheric (equivalent) currents based on the SECS method and the time derivatives of the ground magnetic field, attempts were made to find a systematic behaviour of different SSC events occurring at the same time of

the day. Some correlation was found between events at UT hours 5 or 15 but more definite results obviously require much further research. Anyway, possibilities of developing GIC forecasting techniques based on SSC events do not look promising at least yet.

4. CONCLUDING REMARKS

Geomagnetically induced currents (GIC) are the ground manifestation of space weather. Thus, besides the practical importance of GIC research, GIC also provide additional data in connection with space research. It should be noted that the history of observations and investigations of GIC is much longer than the time of roughly ten years during which intensive research under the term “space weather” has been going on. The first GIC observations were already made in early telegraph equipment about 150 years ago. An increase in GIC research interests occurred after the harmful GIC effects on American power grids during a large geomagnetic storm in March 1940, and more active research on the topic has been done since the 1970s. The famous GIC catastrophe in the Hydro-Québec power system in March 1989 remarkably increased GIC research all over the world, in particular in North America.

Probably the best and most practical way to decrease GIC risk and to avoid problems is developing forecasting methods based on observations of the solar wind by satellites at the L1 point located at about 1.5 million km from the Earth towards the Sun. Such a forecast would typically provide a time of the order of 30 to 60 minutes for taking countermeasures against a coming GIC event.

A crucial parameter is the geoelectric field at the Earth’s surface, which is the driving force of GIC. Thus, recent research efforts have concentrated on fast and accurate enough calculation techniques of the geoelectric field. The complex image method (CIM) is a tool in this respect since it allows an efficient computation of the surface electric and magnetic fields from information about the Earth’s conductivity and about ionospheric currents as the input. The latter may be determined, for example, from ground-based magnetic recordings by using the spherical elementary current system (SECS) method. The simple plane wave technique also seems to be a very appropriate way to calculate the electric field from ground magnetic data.

An application of real-time magnetic observations naturally only permits a nowcasting of GIC but it might also be possible to forecast the ground magnetic field from solar wind data based, e.g., on neural networks, and it combined with SECS and CIM or the plane wave method would lead to the forecasting of GIC as well. Neural networks are certainly worth investigating

in this connection. However, a final aim in the farther future should be to provide forecasting techniques that are based on physical models of the coupling between the solar wind and the magnetosphere-ionosphere system.

A shortcoming that both CIM and the plane wave technique suffer from is that they only work for layered-Earth models, so for example areas near ocean-continent boundaries cannot be investigated. Therefore, theoretical modelling developments should contain methods to deal with horizontal variations of the Earth's structure. On the other hand, however, using always a local (layered-Earth) surface impedance seems to yield sufficiently accurate results.

In conclusion, pieces for GIC calculation and forecasting exist, and they should only be put together to cover the whole chain as efficiently and usefully as possible. In any case, possibilities of forecasting GIC magnitudes at individual sites reliably are still extremely difficult today. A technique of providing GIC nowcasts, forecasts and warnings is used for protecting the National Grid Company's power system against GIC in England (Erinmez et al., 2002a; 2002b).

A step towards a better understanding of the GIC risk is to classify space weather events according to their GIC effectiveness and characters. A study associated with GIC in Scotland and Finland during the large geomagnetic storm in April 2000 as well as a statistical investigation about the correlation of GIC data with sudden storm commencement (SSC) events discussed in this paper give some hints but additional research is inevitably needed before definite conclusions.

5. ACKNOWLEDGEMENTS

We wish to thank the Fingrid and Gasum companies for support and collaboration in studies of GIC in the Finnish high-voltage power system and in the Finnish natural gas pipeline during many years. The support from the Academy of Finland to one of us (AP) is also acknowledged.

6. REFERENCES

Amm, O., Ionospheric Elementary Current Systems in Spherical Coordinates and Their Applications, J. Geomag. Geoelectr., 49, 947-955, 1997.

Amm, O., and A. Viljanen, Ionospheric disturbance magnetic field continuation from the ground to the ionosphere using spherical elementary current systems, Earth Planets Space, 51, 431-440, 1999.

Bolduc, L., GIC observations and studies in the Hydro-Québec power system, J. Atm. Solar-Terr. Phys., 64, 16, 1793-1802, 2002.

Boteler, D. H., Geomagnetic effects on pipe-to-soil potentials of a continental pipeline, Advances in Space Research, 26, 1, 15-20, 2000.

Boteler, D. H., and R. J. Pirjola, The complex-image method for calculating the magnetic and electric fields produced at the surface of the Earth by the auroral electrojet, Geophysical Journal International, 132, 31-40, 1998.

Boteler, D.H., R.J. Pirjola, and H. Nevanlinna, The effects of geomagnetic disturbances on electrical systems at the earth's surface, Advances in Space Research, 22, 17-27, 1998.

Cagniard, L., Basic theory of the magnetotelluric method of geophysical prospecting, Geophysics, 18, 605-635, 1953.

Czech, P., S. Chano, H. Huynh, and A. Dutil, The Hydro-Québec system blackout of 13 March 1989: system response to geomagnetic disturbance, EPRI Report, TR-100450, Proceedings of Geomagnetically Induced Currents Conference, Millbrae, California, USA, November 8-10, 1989, 19.1-19.21, 1992.

Dmitriev, V. I., and M. N. Berdichevsky, The Fundamental Model of Magnetotelluric Sounding, Proc. IEEE, 67, 1034-1044, 1979.

Elovaara, J., P. Lindblad, A. Viljanen, T. Mäkinen, R. Pirjola, S. Larsson, and B. Kielén, Geomagnetically Induced Currents in the Nordic Power System and Their Effects on Equipment, Control, Protection and Operation, CIGRE, 1992 Session, 30 August - 5 September, 1992, Paris, 36-301, 11 pp., 1992.

Erinmez, I. A., J. G. Kappenman, and W. A. Radasky, Management of the geomagnetically induced current risks on the national grid company's electric power transmission system, J. Atm. Solar-Terr. Phys., 64, 5-6, 743-756, 2002a.

Erinmez, I. A., S. Majithia, C. Rogers, T. Yasuhiro, S. Ogawa, H. Swahn, and J. G. Kappenman, Application of modelling techniques to assess geomagnetically induced current risks on the NGC transmission system, CIGRE, Session - 2002, 39-304, 10 pp., 2002b.

Fukushima, N., Generalized theorem for no ground magnetic effect of vertical currents connected with Pedersen currents in the uniform-conductivity ionosphere, Report of Ionosphere and Space Research in Japan, 30, 35-40, 1976.

Gummow, R. A., GIC effects on pipeline corrosion and corrosion control systems, J. Atm. Solar-Terr. Phys., 64, 16, 1755-1764, 2002.

Häkkinen, L., and R. Pirjola, Calculation of Electric and Magnetic Fields Due to an Electrojet Current System above a Layered Earth, Geophysica, 22, 1&2, 31-44, 1986.

Kappenman, J.G., Geomagnetic Storms and Their Impact on Power Systems, IEEE Power Engineering Review, May 1996, 5-8, 1996.

Kappenman, J. G., and V. D. Albertson, Bracing for the Geomagnetic Storms, IEEE Spectrum, March 1990, 27-33, 1990.

Karsberg, A., G. Swedenborg, and K. Wyke, The influences of earth magnetic currents on telecommunication lines, Tele (English edition), Televerket (Swedish Telecom), 1, Stockholm, Sweden, 1-21, 1959.

Kilpua, S., Nopeiden maailmanlaajuisten geomagneettisten kentänmuutosten aiheuttamat vaikutukset maanpinnalla (in Finnish), M.Sc. thesis, University of Helsinki, Finland, 57 pp., 2003.

Lahtinen, M., and J. Elovaara, GIC Occurrences and GIC Test for 400 kV System Transformer, IEEE Transactions on Power Delivery, 17, 2, 555-561, 2002.

Lehtinen, M., and R. Pirjola, Currents produced in earthed conductor networks by geomagnetically-induced electric fields, Ann. Geophys., 3, 4, 479-484, 1985.

Mäkinen, T., Geomagnetically induced currents in the Finnish power transmission system, Finnish Meteorological Institute, Geophysical Publications, No. 32, Helsinki, Finland, 101 pp., 1993.

Mareschal, M., Modelling of natural sources of magnetospheric origin in the interpretation of regional induction studies: a review, Surveys in Geophysics, 8, 261-300, 1986.

Molinski, T. S., Why utilities respect geomagnetically induced currents, J. Atm. Solar-Terr. Phys., 64, 16, 1765-1778, 2002.

Nevanlinna, H., P. Tenhunen, R. Pirjola, J. Annanpalo, and A. Pulkkinen, Breakdown caused by a geomagnetically induced current in the Finnish telesystem in 1958, J. Atm. Solar-Terr. Phys., 63, 10, 1099-1103, 2001.

Pirjola, R., Electromagnetic induction in the earth by a plane wave or by fields of line currents harmonic in time and space, Geophysica, 18, 1-2, 1-161, 1982.

Pirjola, R., On magnetotelluric source effects caused by an auroral electrojet system, Radio Science, 27, 463-468, 1992.

Pirjola, R., Geomagnetically induced currents during magnetic storms, IEEE Transactions on Plasma Science, 28, 6, 1854-1866, 2000.

Pirjola, R., Fundamentals about the flow of geomagnetically induced currents in a power system applicable to estimating space weather risks and designing remedies, J. Atm. Solar-Terr. Phys., 64, 18, 1967–1972, 2002.

Pirjola, R., Effects of space weather on high-latitude ground systems, Submitted to Advances in Space Research, 13 pp., 2003.

Pirjola, R. and D. Boteler, Calculation methods of the electric and magnetic fields at the Earth's surface produced by a line current, Radio Science, 37, 3, 10.1029/2001RS002576, 14-1-14-9, 2002.

Pirjola, R. J., and A. T. Viljanen, Geomagnetic Induction in the Finnish 400 kV Power System, Environmental and Space Electromagnetics, Proceedings of the URSI International Symposium on Environmental and Space Electromagnetics, Tokyo, Japan, September 4-6, 1989, edited by H. Kikuchi, Springer–Verlag, Chapter 6.4, 276-287, 1991.

Pirjola, R., and A. Viljanen, Complex image method for calculating electric and magnetic fields produced by an auroral electrojet of a finite length, Ann. Geophys., 16, 1434-1444, 1998.

Pirjola, R. J., A. T. Viljanen, and D. H. Boteler, Electric field at the seafloor due to a two-dimensional ionospheric current. Geophys. J. Int., 140, 286-294, 2000.

Pulkkinen, A. Geomagnetic induction during highly disturbed space weather conditions: Studies of ground effects, Ph.D. thesis, University of Helsinki, Finland, Finnish Meteorological Institute, Contributions, To be published in August 2003, 78 pp., 2003.

Pulkkinen, A., O. Amm, A. Viljanen, and BEAR Working Group, Ionospheric equivalent current distributions determined with the method of spherical elementary current systems, J. Geophys. Res., 108, doi: 10.1029/2001JA005085, 2003a.

Pulkkinen, A., A. Thomson, E. Clarke, and A. McKay, April 2000 geomagnetic storm: ionospheric drivers of large geomagnetically induced currents, Ann. Geophys., in press, 2003b.

Pulkkinen, A., R. Pirjola, D. Boteler, A. Viljanen, and I. Yegorov, Modelling of space weather effects on pipelines, J. Appl. Geophys., 48, 233-256, 2001.

Stauning, P., High-voltage power grid disturbances during geomagnetic storms, Proceedings of the Second Solar Cycle and Space Weather Conference (SOLSPA 2001), Vico Equense, Italy, 24-29 September 2001, European Space Agency, SP-477, 521-524, 2002.

Thomson, D. J., and J. T. Weaver, The Complex Image Approximation for Induction in a Multilayered Earth, J. Geophys. Res., 80, 1, 123-129, 1975.

Viljanen, A., The relation between geomagnetic variations and their time derivatives and implications for estimation of induction risks, Geophys. Res. Lett., 24, 6, 631-634, 1997.

Viljanen, A., and R. Pirjola, Geomagnetically induced currents in the Finnish high-voltage power system, A geophysical review, Surveys in Geophysics, 15, 383-408, 1994.

Viljanen, A., Nevanlinna, H., Pajunpää, K. and Pulkkinen, A., Time derivative of the horizontal geomagnetic field as an activity indicator, Ann. Geophys., 19, 1107-1118, 2001.

Wait, J. R., On the Relation between Telluric Currents and the Earth's Magnetic Field, Geophysics, 19, 281-289, 1954.

Wait, J. R., Electromagnetic surface impedance for a layered earth for general excitation, Radio Science, 15, 1, 129-134, 1980.

Wait, J. R., Wave Propagation Theory, Pergamon Press, 348 pp., 1981.

Wallerius, A., Solen gav Sverige en strömstöt (in Swedish), Ny Teknik - teknisk tidskrift, 29, 3, 1982.

Chapter 14

Space Weather and the Vulnerability of Electric Power Grids

An Overview of the Increasing Vulnerability Trends of Modern Electric Power Grid Infrastructures and the potential consequences of Extreme Space Weather Environments

John G. Kappenman

Metatech Corporation, Applied Power Solutions Division
5 W. First St, Suite 301, Duluth, Mn, USA

Abstract A number of trends are causing overall increases in geomagnetically-induced currents (GIC's) and associated threats from geomagnetic storms for electric power grids. GIC threats have been a concern for power grids at high-latitude locations due to disturbances driven by electrojet intensifications. However, other geomagnetic storm processes such as SSC and ring current intensifications are also proving to cause GIC concerns for the power industry at low-latitude locations as well. In addition to threats arising from various regions of the space environment, the response of local ground and power system design have important roles that can significantly increase risk from geomagnetic storms. In particular a number of long-term trends in power system design and operation have been continually acting to increase geomagnetic storm risks. These design implications have acted to greatly escalate GIC risks for power grids at all latitude locations. As a result, GIC impacts may now be of concern even to power grids that have never considered the risk of GIC previously because they were not at high latitude locations. The paper will provide a comprehensive overview of these risk issues as they apply to many world power systems and particularly review the potential impacts to power system operations due to extreme geomagnetic disturbance events.

Keywords Geomagnetically induced currents, GIC, geoelectric field, geomagnetic disturbances, electric power grids, reactive power demand, MVARs, harmonics, SSC, electrojet intensification, ring current intensification.

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 257-286.

1. INTRODUCTION

Continuous advances have occurred in the understanding of space weather or more specifically geomagnetic storm environments and the ability of modelling these environments and the environmental interactions with electric power networks. As these efforts have progressed, it has also become evident that devastating impacts due to these storms events are plausible. These disturbances have caused catastrophic impacts to technology systems in the past (e.g., the power blackout in Quebec in March 1989). More importantly, as detailed examinations have been undertaken concerning the interaction of geomagnetic storm environments with power grids and similar infrastructures, the realization has developed that these infrastructures are becoming more vulnerable to disruption from electromagnetic interactions for a wide variety of reasons. This direction of these trends suggests that even more severe impacts can occur in the future for reoccurrences of historically large storms on present-day systems.

While more details will be provided in later sections of this paper, a brief overview of how these geomagnetic disturbance environments actually interact with large regional power grids indicates the complex nature of the threat. When geomagnetic storms occur they result in slowly varying (1-1000 seconds) geomagnetic field disturbances that can have very large geographic footprints. These magnetic field disturbances will induce electric fields at the Earth's surface over these same large regions. Across the U.S. and most other locations around the world, complex topologies of long-distance transmission lines have been built. These grids include transformers at generating plants and substations that have grounded neutrals. The transformer neutrals provide a path from the network to ground for these slowly varying electric fields (less than 1 Hz) to induce a current flow through the network phase wires and transformers.

These currents (known as geomagnetically-induced currents – GICs) are generally on the order of 10s to 100s of amperes during a geomagnetic storm. Though these quasi-DC currents are small compared to the normal AC current flows in the network, they have an impact that becomes enormously amplified on the operation of transformers in the network. Under normal conditions, even the largest transformer requires only a few amperes of AC excitation current to energize its magnetic circuit. GIC when present, also acts as an excitation current for these magnetic circuits, therefore GIC levels of only 1 to 10 amperes can initiate magnetic core saturation during one-half of the AC cycle in an exposed transformer, causing extremely large AC currents to be drawn from the power grid. As GIC levels increase, the levels of saturation of the transformer core also increase.

When a transformer saturates due to the presence of GIC, it effectively becomes an amplifier of highly distorted AC current. Because the disturbance can span a very large area, this large injection of AC distortion behaviour can be produced simultaneously in a large number of exposed transformers. The simultaneous injections of these AC distortions from many transformers can cause widespread operational and reliability problems throughout the grid. This amplified AC current from saturation effects can pose risks to power networks directly due to increased reactive power demands that can lead to voltage regulation problems. But a nearly equal concern arises from collateral impacts stemming from highly distorted waveforms (rich in harmonics) from saturated transformers that are injected into the network. These distortions can cascade problems by disrupting the performance of other network apparatus and causing them to trip off-line just when they are most needed to preserve network integrity (i.e. relay & protection system mis-operation). If the spatial coverage of the disturbance is large, many transformers will be simultaneously saturated, a situation that can rapidly escalate into a network-wide voltage collapse. In addition, individual transformers may be damaged from overheating due to this unusual mode of operation, which can result in long-term outages to key transformers in the network. In short, the threats to the infrastructure from geomagnetic storms include the possibility of widespread power blackouts, damage to expensive and difficult to replace transformers, and damage to equipment connected to the grid.

In order to assess the risks that modern electric power grids face in regards to the space environment it is necessary to consider a wide variety of risk modifiers and multipliers. These risk modifiers start with a consideration of the various space environment disturbance processes that can cause differing degrees of impulsive geomagnetic disturbance environments at differing latitude locations from auroral locations to equatorial locations. Other risk modifiers include the nature of the electromagnetic interaction between geomagnetic field disturbances and the solid-earth geophysics of the terrestrial environment that produces the geo-electric field. This particular risk modifier may be responsible for the highest degree of uncertainty because of the relative lack of information on the conductivity properties to depth of many regions of the Earth. And finally what is emerging as one of the largest risk escalators for electric power grids, is the greatly magnified exposure risks due to geomagnetic storms, which has developed from the evolution of power grid design and operational factors.

2. GEOLOGICAL RISK FACTORS & GEO-ELECTRIC FIELD RESPONSE

Considerable prior work has been done to model the geomagnetic induction effects in ground-based systems [Albertson, Lanzerotti-1983, Pirjola]. As an extension to this fundamental work, numerical modelling of ground conductivity conditions have been demonstrated to provide accurate replication of observed geo-electric field conditions over a very broad frequency spectrum [Kappenman-1997]. Past experience has indicated that 1-D Earth conductivity models are sufficient to compute the local electric fields. Lateral heterogeneity of ground conductivity conditions can be significant over meso-scale distances [Kappenman-2001]. In these cases, multiple 1-D models can be used in cases where the conductivity variations are sufficiently large.

Ground conductivity models need to accurately reproduce geo-electric field variations that are caused by the considerable frequency ranges of geomagnetic disturbance events from the large magnitude/low frequency electrojet-driven disturbances to the low amplitude but relatively high frequency impulsive disturbances commonly associated with SSC events. This variation of electromagnetic disturbances therefore require models accurate over a frequency range from 0.3 Hz to as low as 0.00001 Hz. At these low frequencies of the disturbance environments, diffusion aspects of ground conductivities must be considered to appropriate depths. Therefore skin depth theory can be used in the frequency domain to determine the range of depths that are of importance. It is clear that for constant Earth conductivities, the depths required are more than several hundred kilometres, although the exact depth is a function of the layers of conductivities present at a specific location of interest.

It is generally understood that the Earth's mantle conductivity increases with depth. In most locations, ground conductivity laterally varies substantially at the surface over meso-scale distances; these conductivity variations with depth can range 3 to 5 orders of magnitude. While surface conductivity can exhibit considerable lateral heterogeneity, conductivity at depth is more uniform, with conductivities ranging from values of .1 to 10 S/m at depths from 600 to 1000km [Campbell-1987, Masse-1987]. If sufficient low-frequency measurements are available to characterize ground conductivity profiles, models of ground conductivity can be successfully applied over meso-scale distances and can be accurately represented by use of layered conductivity profiles or models. For illustration of the importance of ground models on the response of geo-electric fields, a set of four example ground models have been developed that illustrate the probable

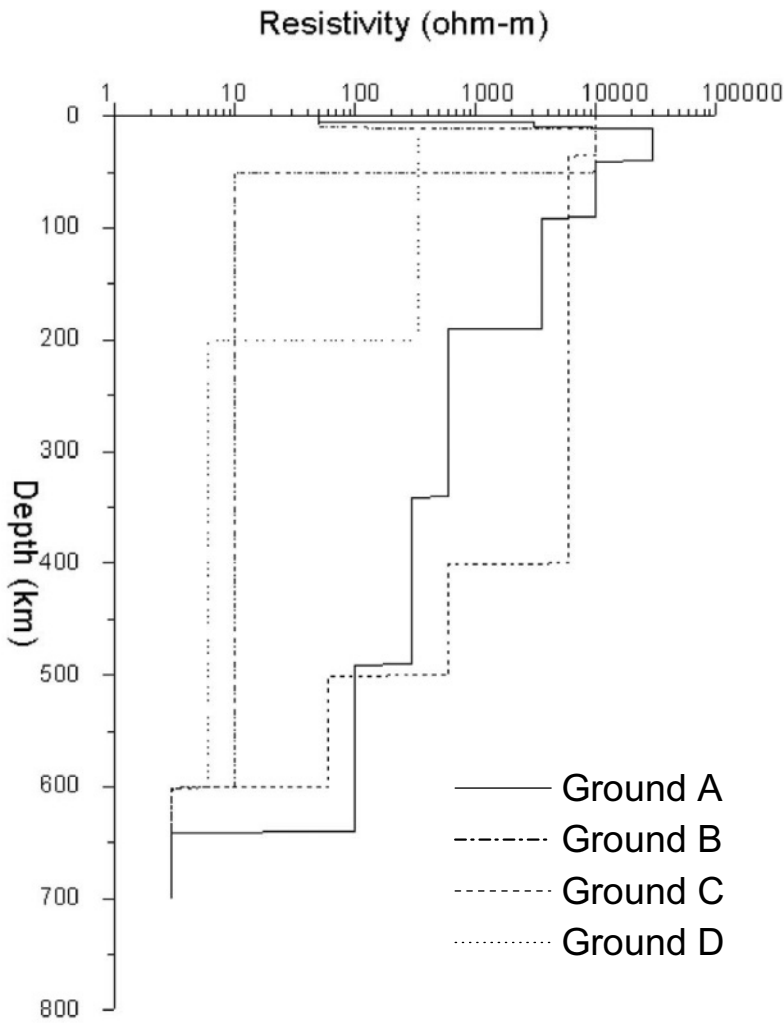

Figure 1. Resistivity profiles versus depth for four example layered earth ground models.

lower to upper quartile response characteristics of most known ground conditions, considering there is a high degree of uncertainty in the plausible diversity of upper layer conductivities. Figure 1 provides a plot of the layered ground conductivity conditions for these four ground models to depths of 700 km. As shown, there can be as much as four orders of magni-

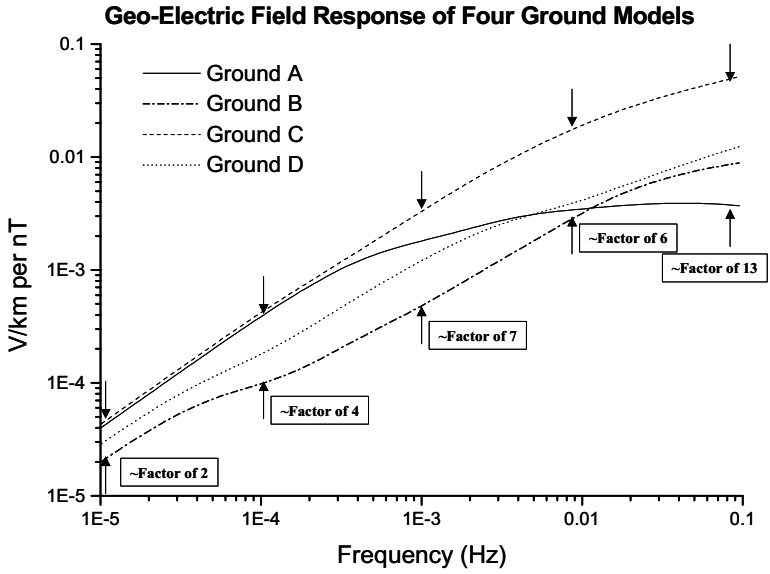

Figure 2. Frequency response of four example ground models of Figure 1, Max/Min geo-electric field response characteristics shown at various discrete frequencies.

tude variation in ground resistivity at various depths in the upper layers. Models A and B have very thin surface layers of relatively low resistivity. Models A and C are characterized by levels of relatively high resistivity until reaching depths exceeding 400km, while models B and D have high variability of resistivity in only the upper 50 to 200km of depth [Campbell-1980, Rasmussen-1987, 1988].

Figure 2 provides the frequency response characteristics for these same four layered earth ground models of Figure 1. Each line plot represents the geo-electric field response for a corresponding incident magnetic field disturbance at each frequency. While each ground model has unique response characteristics at each frequency, in general all ground models produce higher geo-electric field responses as the frequency of the incident disturbance increases. Also shown on this plot are the relative differences in geo-electric field response for the lowest and highest responding ground model at each decade of frequency. This illustrates that the response between the lowest and highest responding ground model can vary at discrete frequencies by more than a factor of 10. Also because the frequency content of an impulsive disturbance event can have higher frequency content (for instance due to a SSC), the disturbance is acting upon the more responsive portion of the frequency range of the ground models [Kappenman-2003].

Therefore, the same disturbance energy input at these higher frequencies produces a proportionately larger response in geo-electric field. For example in most of the ground models, the geo-electric field response is a factor of 50 higher at 0.1 Hz compared to the response at 0.0001 Hz.

From the frequency response plots of the ground models as provided in Figure 2, some of the expected geo-electric field response due to geomagnetic field characteristics can be inferred. For example, Ground C provides the highest geo-electric field response across the entire spectral range, therefore, it would be expected that the time-domain response of the geo-electric field would be the highest for nearly all B field disturbances. At low frequencies, Ground B has the lowest geo-electric field response while at frequencies above 0.02 Hz, Ground A produces the lowest geo-electric field response. Because each of these ground models have both frequency-dependent and non-linear variations in response, the resulting form of the geo-electric field waveforms would be expected to differ in form for the same B field input disturbance. In all cases, each of the ground models produces higher relative increasing geo-electric field response as the frequency of the incident B field disturbance increases. Therefore it should be expected that a higher peak geo-electric field should result for a higher spectral content disturbance condition.

A large electrojet-driven disturbance is capable of producing an impulsive disturbance as shown in Figure 3, which reaches a peak delta B magnitude of ~2000 nT with a rate of change (dB/dt) of 2400 nT/min. This disturbance scenario can be used to simulate the estimated geo-electric field response of the four example ground models. Figure 4 provides the geo-electric field responses for each of the four ground models for this 2400 nT/min B field disturbance. As expected, the Ground C model produces the largest geo-electric field reaching a peak of ~15 V/km, while Ground A is next largest and the Ground B model produces the smallest geo-electric field response. The Ground C geo-electric field peak is more than 6 times larger than the peak geo-electric field for the Ground B model. It is also evident that significant differences result in the overall shape and form of the geo-electric field response. For example, the peak geo-electric field for the Ground A model occurs 17 seconds later than the time of the peak geo-electric field for the Ground B model. In addition to the differences in the time of peak, the waveforms also exhibit differences in decay rates. As is implied from this example, both the magnitudes of the geo-electric field responses and the relative differences in responses between models will change dependent on the source disturbance characteristics.

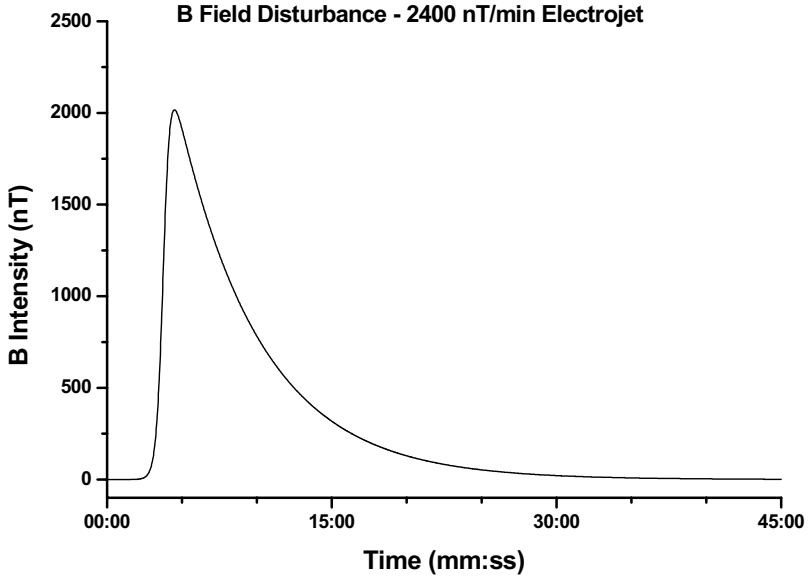

Figure 3. Waveform of example electrojet-driven geomagnetic field disturbance with 2400 nT/min rate of change intensity.

3. POWER GRID DESIGN & NETWORK TOPOLOGY RISK FACTORS

While ground conductivity conditions are important in determining the geo-electric field response, and in determining levels of GICs and their resulting impacts. Power grid design is also an important factor in the vulnerability of these critical infrastructures, a factor in particular that over time has greatly escalated the effective levels of GIC and operational impacts due to these increased GIC flows.

Power systems are designed and operated with a focus on maintaining a balance between generation and demand at all times in a distributed manner. Sufficient reserves are provided throughout the system so that it can tolerate the loss of any one component at any time (the N-1 criterion). Power system designers and operators expect these systems to be challenged by the elements, and where those challenges were fully understood in the past, the system design has worked extraordinarily well.

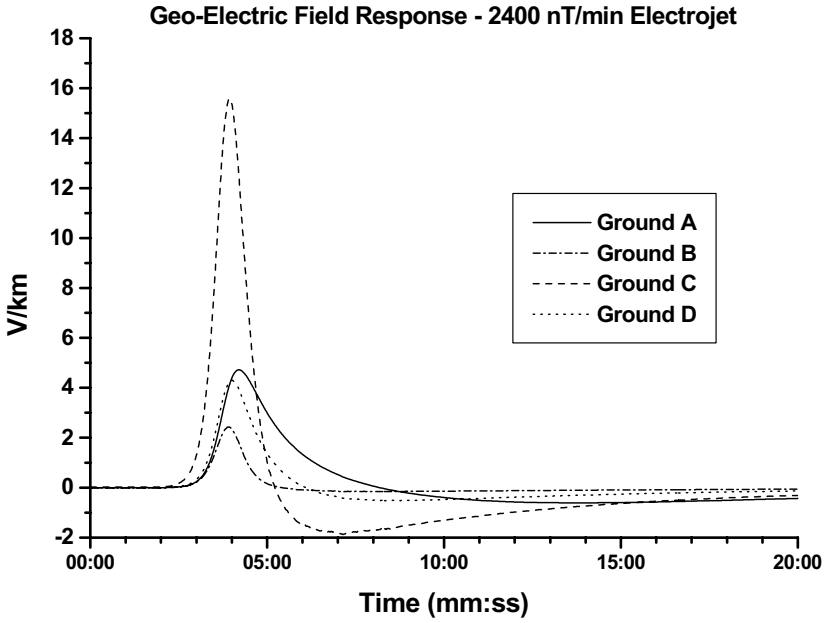

Figure 4. Geo-electric field response of the four example ground models to the 2400 nT/min disturbance conditions of Figure 3.

The primary design approach undertaken by the industry for decades has been to weave together a tight network, which pools resources and provides redundancy to reduce failures. In essence, unaffected neighbouring grids help out the temporarily weakened portion of the grid.

Ironically, the designs that have worked to make the electric power industry strong for ordinary weather, introduce key vulnerabilities to the electromagnetic coupling phenomena of geographically widespread geomagnetic disturbances. Since large interconnected power grids and intense geomagnetic disturbances can both have continental footprints, the design concept of unaffected neighbouring system and sharing the burden of storm-caused stresses are not always realizable. Unlike ordinary weather patterns that arise due to thermodynamic conditions, the electromagnetic interactions of impulsive geomagnetic field disturbances can develop very rapidly and when present are inherently near-instantaneously observed across the exposed system.

The extent of the change or growth in vulnerability in the US and other major world power grid infrastructures over time are due to a number of factors stemming from either growth in the infrastructure base or technology changes within the existing base that introduce new impact problems.

Figure 5 shows the growth of the US high voltage transmission grid over the last 50 years. The high voltage transmission grid is the portion of the power network that spans long distances. This geographically widespread infrastructure readily couples through multiple ground points to the geo-electric field produced by disturbances in the geomagnetic field. As shown in Figure 5, from Solar Cycle 19 (late 1950's), through Solar Cycle 22 (early 1980's), the high voltage transmission grid has grown nearly tenfold. Similar development rates of transmission infrastructure have occurred simultaneously in other developed regions of the world.

As this network has grown in size, it has also grown in complexity and sets in place a compounding of risks that are posed to the US power grid infrastructure for GIC events. Some of the more important changes in technology base that can increase impacts from GIC events include higher design voltages, changes in transformer design and other related apparatus. The operating levels of high-voltage networks have increased from the 100-200kV thresholds of the 1950's to 400 to 765kV levels of present-day networks. With this increase in operating voltages, the average per unit length circuit resistance has decreased while the average length of the grid circuit increases. In addition, power grids are designed to be tightly interconnected networks, which present a complex and in many cases a system that is continental in size. These interrelated design factors have acted to substantially increase the levels of GIC that are possible in modern power networks.

All high voltage bulk power grids throughout the world utilize a three-phase configuration for delivery of power over the long-distance high voltage transmission networks from power generation facilities to end-users of the electricity. In this delivery process, transformers are used to step up or step-down voltage levels, as it is most efficient to transmit long distances at high voltage (69kV to 765kV), but producing and using electricity has to be done at very low voltages (120 to 4000 volts). These transformers introduce the path by which GIC enters and exits the power grid. Also, the GIC when it flows through these transformers is the root cause of all power system problems as these transformers saturate due to this quasi-DC current. This saturated mode of operation can cause distortions or disruptions to the operation of the entire power grid.

GIC levels are determined by the size and the resistive impedance of the power grid circuit itself when coupled with the level of geo-electric field that results from the geomagnetic disturbance event. Given a geo-electric field imposed over the extent of a power grid, a current will be produced entering the neutral ground point at one location and exiting through other ground points elsewhere in the network.

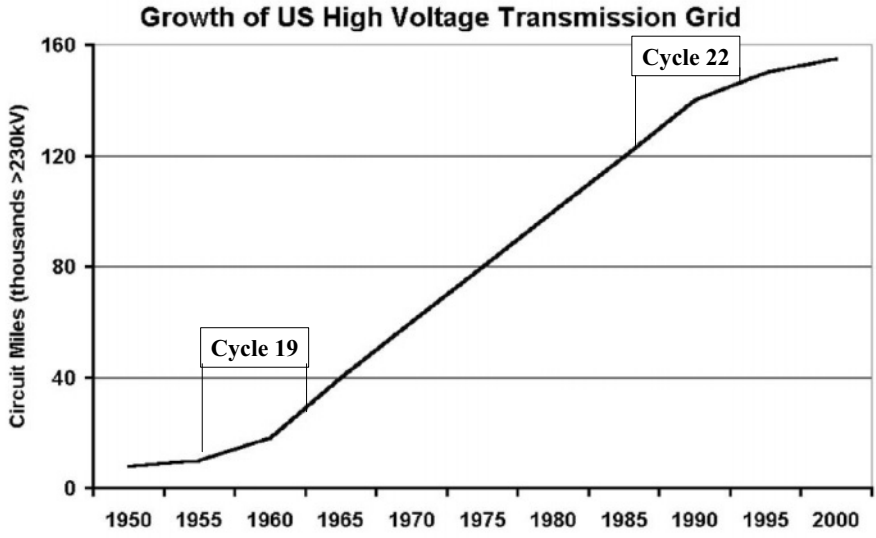

Figure 5**.** Growth of the US High Voltage Transmission Network over the past 50 years.

The resistive impedance of transmission circuits vary significantly with voltage class, the higher the kV rating the lower the resistive impedance per unit distance (i.e. ohms per mile) [Howlett, ECAR, FERC]. Figure 6 provides a plot of the average resistive impedance per transmission line by the major kV Rating classes for the US power grid. The lowest transmission system voltage surveyed was at 69kV, while the highest was at 765kV. As indicated, the average R per unit length decreases by more than a factor of 10 as the voltage level increases over this range. Therefore a 69kV and 765kV transmission lines of equal length will also have factor of ~10 difference in total circuit resistance and if coupled to the same geo-electric field, the level of GIC flow will be ~10 times larger in the 765kV line.

The resistive impedance of transformers exhibits an even larger degree of decrease as the size rating of the transformer increases. Figure 7 provides a plot of transformer R versus the AC Current Rating. As shown in this plot, a sampling of the actual data points for transformers in the US population are shown along with population data, which indicates a factor of 20 reduction in R as the transformer size increases. As shown in this plot, a sampling of the actual data points for transformers in the US population are shown along with population data, which indicates a factor of 20 reduction in R as the transformer size increases. When this resistive element is added to the

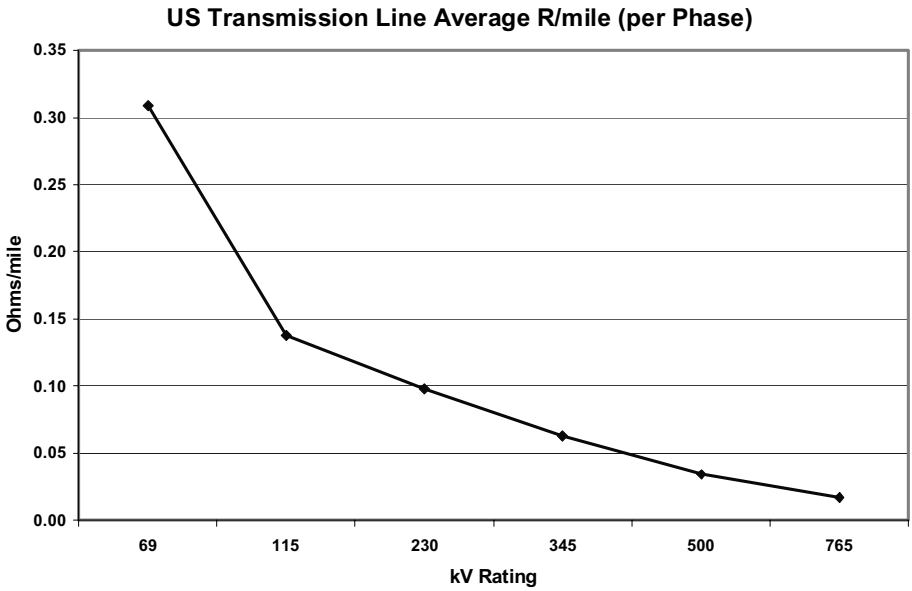

Figure 6**.** Average transmission line resistance per mile in the US by kV rating of the transmission lines.

overall GIC circuit, the expected trend should be a large increase in GIC levels for higher kV ratings.

This trend, of course, has ominous implications in that larger GIC flows will occur on the larger and more important portions of the power grid infrastructure. Most power grids are highly complex networks with numerous circuits or paths and transformers for GIC to flow through. This requires the application of highly sophisticated network and electromagnetic coupling models to determine the magnitude and path of GIC throughout the complex power grid. However for the purposes of illustrating the impact of power system design, a review will be provided using a single transmission line terminated at each end with a single transformer to ground connection. To illustrate the differences that can occur in levels of GIC flow at higher voltage levels, the simple demonstration circuit have also been developed at 138kV, 230kV 345kV, 500kV and 765kV which are common grid voltages used in the US and Canada. In Europe, voltages of 130kV, 275kV and 400kV are commonly used for the bulk power grid infrastructures. For these calculations, a uniform 1.0 volt/km geo-electric field disturbance conditions are used, which means that the change in GIC levels will result from changes

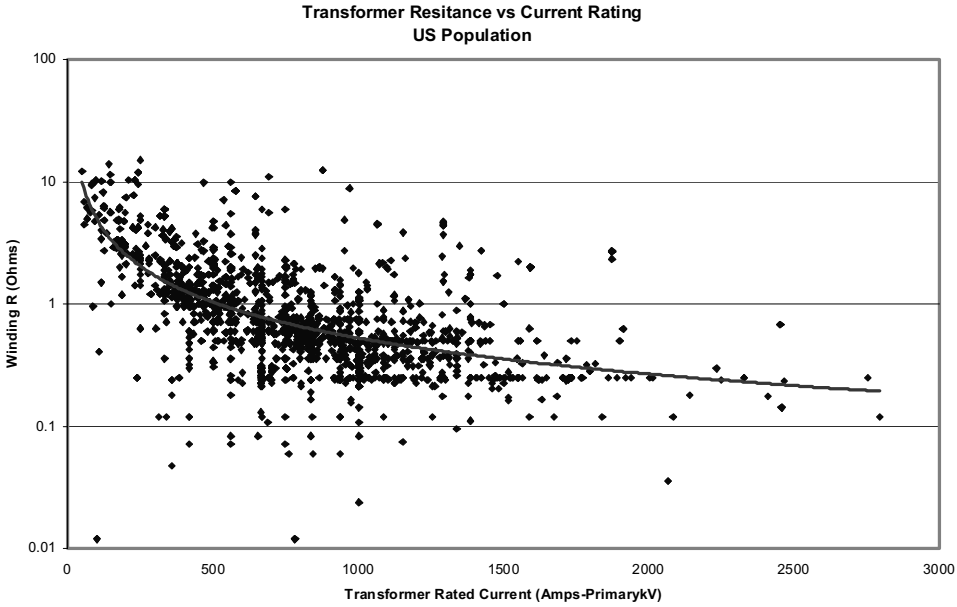

Figure 7. Statistics on average transformer resistive impedance versus AC current rating of the transformers in the US power grid population.

in the power grid resistances alone. Also for uniform comparison purposes, a 100 km long line is used in all kV Rating cases.

Figure 8 illustrates the comparison of GIC flows that would result for various US infrastructure power grid kV Ratings using the simple circuit and a uniform 1.0 volts/km geo-electric field disturbance. In complex networks, such as those in the US, some scatter from this trend line is possible due to normal variations in circuit parameters such as line resistances, etc that can occur in the overall population of infrastructure assets. Further, this was an analysis of simple "one-line" topology network, whereas real power grid networks have highly complex topologies, span large geographic regions, and present numerous paths for GIC flow, all of which tend to increase total GIC flows. Even this limited demonstration tends to illustrate that the power grid infrastructures of large grids in the US and other locations of the world are increasingly exposed to higher GIC flows due to design changes that have resulted in reduced circuit resistance. Compounding this risk further, the higher kV portions of the network handle the largest bulk power flows and form the backbone of the grid. Therefore the increased GIC-risk is being placed at the most vital portions of this critical infrastructure. In the US, 345kV, 500kV, and 765kV transmission systems are widely spread

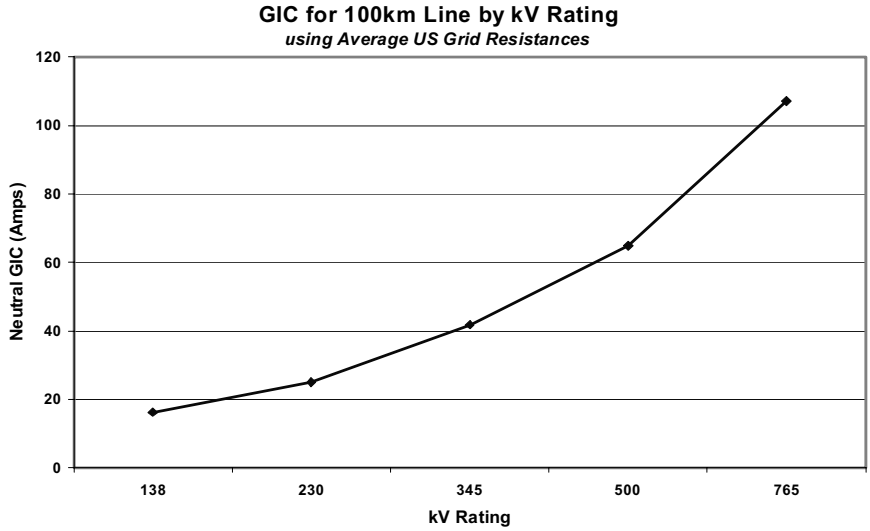

Figure 8**.** Average Neutral GIC Flows vs kV Rating for a 100km demonstration transmission circuit.

throughout the US and especially concentrated in areas of the US with high population densities.

One of the best ways to illustrate the operational impacts of large GIC flows is to review the way in which the GIC can distort the AC output of a large power transformer due to half-cycle saturation. Under severe geomagnetic storm conditions, the levels of Geo-Electric field can be many times larger than the uniform 1 Volt/km used in the prior calculations. Under these conditions even larger GIC flows are possible. For example in Figure 9, the normal AC current waveform in the high voltage winding of a 500kV transformer under normal full load conditions is shown (~300 amps-rms, ~400 amps-peak). With a large GIC flow in the transformer such as 195 amps, the transformer experiences extreme saturation of the magnetic core for one-half of the AC cycle (half-cycle saturation). During this half-cycle of saturation, the magnetic core of the transformer draws an extremely large and distorted AC current from the power grid. This combines with the normal AC load current producing the highly distorted asymmetrically peaky waveform that now flows in the transformer. As shown, AC current peaks that are present are nearly twice as large compared to normal current for the transformer under this mode of operation. This highly distorted waveform is rich in both even and odd harmonics, which are injected into the system and can cause mis-operations of sensors and protective relays throughout the network [Kappenman-1981, Kappenman-1989].

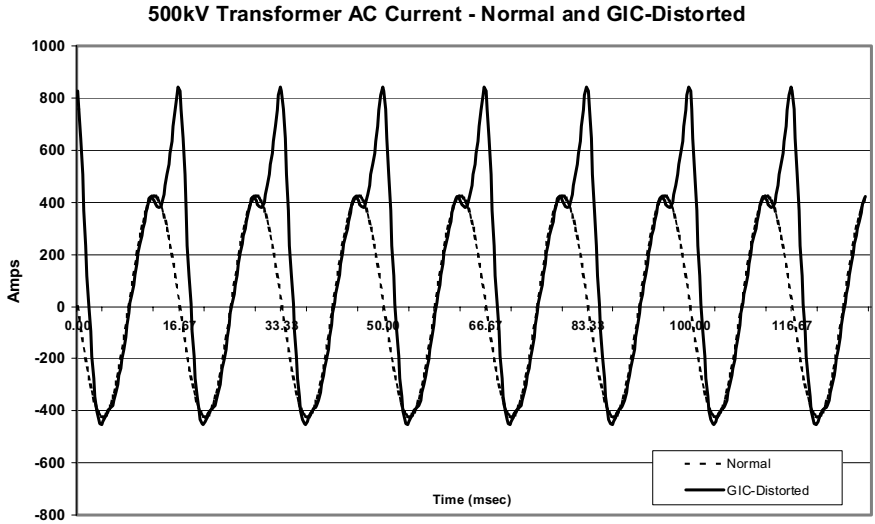

Figure 9. 500 kV Simple Demonstration Circuit Simulation Results –Transformer AC Currents and Distortion due to GIC.

All other transformers on the network can be exposed to similar conditions simultaneously due to the wide geographic extent of most disturbances. This means that the network needs to supply an extremely large amount of reactive power to each of these transformers or voltage collapse of the network could occur. The combination of voltage regulation stress, which occurs simultaneously with the loss of key elements due to relay mis-operations can rapidly escalate to widespread progressive collapse of the exposed interconnected network.

4. POWER GRID THREAT ASSESSMENT FOR GEOMAGNETIC STORM SCENARIOS

Geomagnetic disturbances have caused widespread disruptions to operation of power networks. Recent observations and analysis also indicate that GIC concerns for power grids are not exclusively confined to high-latitude locations. Geomagnetic storms present a number of processes that can drive impulsive or turbulent geomagnetic field variations at ground level. All of these disturbance processes can create conditions of complex and rapid expansions in geographic extent and intensity of impulsive geomagnetic field disturbances. As impacts to power networks can occur on a minute-by-

minute basis, these disturbance environments are not well-characterized by the current 3 hour planetary K, regional K, or any other geomagnetic storm indices that are available to operators of power networks. Rather, it is necessary to fully describe the complex physical manifestation of geomagnetic disturbance environments to model how and to what extent these disturbances impact modern ground-based critical infrastructures. However, the extremes of the impulsive geomagnetic disturbance environment are not well known on a regional basis at low, mid, and high latitude locations [Kappenman-2001].

Because power networks are too large and critical in their operation to easily perform physical tests of their reliability performance for various weather-caused contingencies, the ability to meet these requirements is commonly measured by deterministic study methods to test the power system's ability to withstand probable disturbances. Operators of these critical infrastructures perform extensive modelling and engineering analysis of risks to their systems in evaluating the design and expected performance of their systems for all conceivable operational threat scenarios, with the general exception of threats posed by space weather. These study methods rely extensively upon accurate simulation models of the network and the stress caused by the coupling and reaction to the threat environments. These environmental stress simulations are applied against the network under critical load or system stress conditions to define important system design and operating constraints on the network. System impact assessments for geomagnetic storm scenarios are a pressing need for operators of large complex power systems and if given sufficiently detailed environment data, these simulations can also be readily performed [Kappenman-2001, Albertson-1981, Pirjola-1985]. These advances in modelling have facilitated a number of power grids to begin a process of assessing and quantifying the power grid reliability risks posed by geomagnetic disturbances [Kappenman-2002]. The evaluation of power system vulnerability is, of necessity, a two-stage process. The first stage is one of assessing the exposure to the network posed by the impulsive geomagnetic field disturbances and the long-term climatology of these events specific to the end-user's region of interest. In other words, how large and how frequent can the storm driver be in a particular region? The second stage is one of assessment of the stress that storm events pose to reliability of operation. This is measured through estimates of levels of GIC flow across a network and the manifestation of impacts such as sudden and dramatic increases in reactive power demands and implications such as voltage regulation in the network for power grids. From this analysis effort, meaningful operational procedures can be further identified and refined to better manage the risks resulting from storms of various intensities [Kappenman-2001].

While techniques exist and are quite mature for simulating large-scale power system interaction with the geomagnetic storm environment, the key gap is in the capability to assess the climatology of geomagnetic storms and probable extremes of disturbance conditions in a form needed for systems concerned by GIC impacts. It is not only necessary to provide perspective on the frequency of geomagnetic superstorm events, but also on the extremes in magnitude that are possible. The analysis of historically important geomagnetic disturbance scenarios must take into account the three different and separate geomagnetic disturbance source regions and propagation processes;

i.) ionospheric electrojet intensifications and ground level propagation modes,

ii.) magnetopause/interplanetary boundary shocks and ground level propagation modes,

iii.) ring current intensification and ground level propagation modes.

Even when the geographic scope is limited in application to a country or region, the problem is still complex in that at all latitude locations, at least two of the three disturbance processes will exist.

5. EXTREME GEOMAGNETIC DISTURBANCE EVENTS – OBSERVATIONAL EVIDENCE

A number of new forensic investigations have been undertaken to evaluate storm events over the last 150 years, though modern indices such as Ap only extend back 70 years. These investigations indicate that several storms would far exceed the intensity of all storms over this 70-year period of Ap classification. One such example is a storm from September 1859. Using the Dst index as a measure of storm intensity, a comparison can be provided with the most recent Superstorm of the modern era, which occurred on March 13-14, 1989. For the March 89 Superstorm, the Dst reached a peak of –589. In comparison, the September 1859 storm is estimated to have reached a peak of -1760, a Mega-Storm intensity nearly 3 times larger than the March 1989 Superstorm [Tsurutani-2003]. Other storms, such as on May 1921, have produced measurable geo-electric fields that allow the ability to calibrate against more contemporary storm events. In the example of the May 1921 storm, geo-electric field intensities of ~20 V/km were observed, a level that is again over twice as large as those observed in both the March 1989 and July 1982 Superstorms [Elovaara-1992]. The fact that storms of such intensity as September 1859 and May 1921 have occurred before, indicate that they will eventually occur again. In the examination of these large storms, it is also the conclusion that the source solar event for these

Mega-Storms was not uniquely large and has been observed at intervals as often as once per decade. For example, the very large X22+ solar flare event observed on April 2, 2001 is a contemporary event and is estimated to be larger than the flare that triggered the 1859 storm [Tsurutani-2003]. Rather, what is important is the right convergence of factors from the Sun, to the solar wind and it's interaction with the Earth's magnetic field that set the framework for the Perfect Storm scenario. These Mega-Storms appear to be probable on a 1-in-50 to 1-in-100 year timeframe. Of course, modern critical infrastructures have not as-yet been exposed to storms of this size. Since GIC levels and GIC impacts tend to scale linearly with storm or geo-electric field intensity, it is reasonable to conclude that unprecedented levels of impacts are also likely for power grids and other infrastructures exposed to such extreme environments. More details on the threat to power grid infrastructures from such events will be provided in the following discussion of various storm processes of concern.

5.1 SSC's and Ring Current Intensifications – A New Facet of Space Weather Risk for Power Grids

Large impulsive geomagnetic field disturbances from auroral current systems have always been well understood as a concern for power grids in close proximity to these disturbance regions, predominantly at high-latitude locations. Magnetospheric shocks or SSC's due to large-scale interplanetary pressure pulses, are familiar from a geomagnetic disturbance perspective, but have not been understood in the context as a potential driver for large GIC's. Recent combinations of observational evidence and analysis are determining that such events are capable of producing equivalently large geo-electric fields and associated GIC risks at any latitude, even equatorial locations. Because of the small delta B magnitude observed at low-latitudes, such large GICs pose a paradox. A large SSC disturbance on March 24, 1991 produced some of the largest GIC's ever measured in the US, at mid-latitude locations.

The analysis methods and understanding of electromagnetic coupling processes at that time were unable to fully explain these observations. Figure 10 provides a comparison plot of the impulsive disturbance conditions observed for a typical electrojet-driven disturbance and the geomagnetic field disturbance from a large SSC event at a mid-latitude location. Large electrojet-driven disturbances can cause impulsive disturbances of 2000 nT or greater, while most SSC events are less than 200 nT (only 1/10th as large a disturbance) and could not be conceived as being capable of producing equivalently large geo-electric fields. Yet as shown in Figure 11, the resulting geo-electric fields from these two disturbances produce nearly equivalent intensity geo-electric fields [Kappenman-2003].

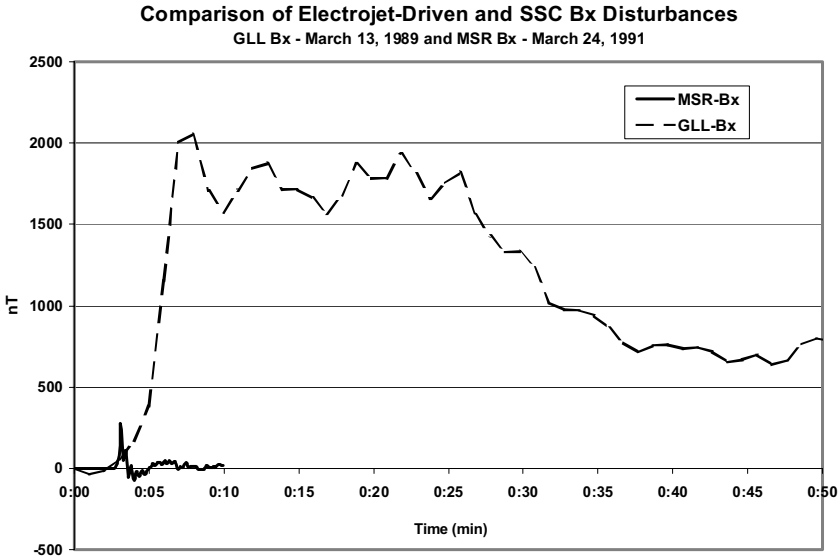

Figure 10. Comparison of delta Bx observed at GLL from electrojet-driven disturbance (March 13, 1989) and at MSR from SSC-event (March 24, 1991).

Electrojet-driven disturbances at high latitudes have large amplitude with relatively lower spectral content, while SSC events are characterized as low amplitude with extraordinarily high spectral content. Disturbance amplitude only accounts for part of the electromagnetic coupling process and the attribute of spectral content of the disturbance is equally important and heretofore had not been well understood and also not well measured unless high-cadence observations were conducted. The deep-earth ground conductivity also provides an important enabling role at higher frequencies. As previously noted, deep-earth ground response to geomagnetic field disturbances is both highly non-linear and highly frequency-dependent. As shown in Figure 12, for nearly all ground conditions, the higher the spectral content of the incident magnetic field disturbance, the higher the relative geo-electric field response. For SSC events, a proportionately smaller

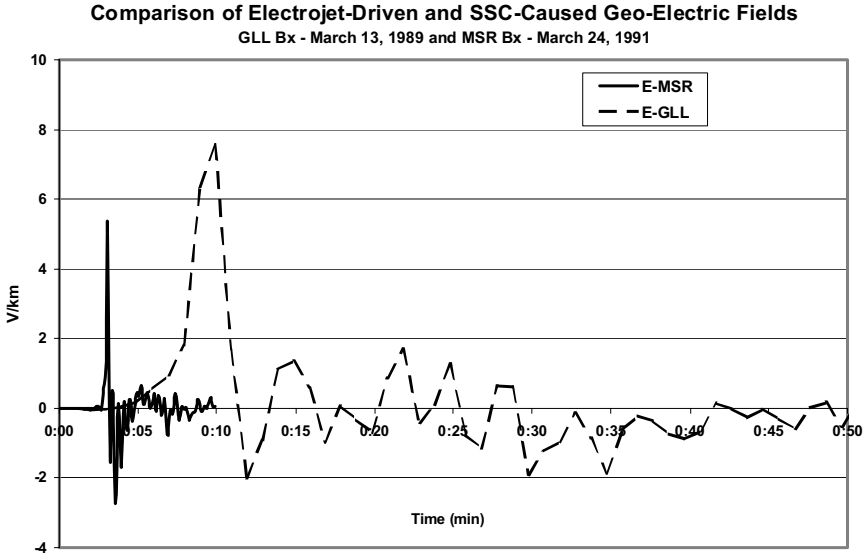

Figure 11**.** Comparison of estimated geo-electric field from Electrojet-driven disturbance as observed at GLL (March 13, 1989) and from SSC-event as observed at MSR (March 244, 1991).

magnitude but higher spectral content B field disturbance is capable of producing equivalent geo-electric fields due to the interaction with the more responsive frequency band of the ground models.

5.2 Ring Current Intensification Scenarios

Recent observations have determined that turbulent ground level geomagnetic field disturbances driven by intensification of the ring current can also create large GIC flows at low latitudes, which were confirmed by observations in central Japan [Erinmez-2002]. These disturbance events have been observed to produce GIC's of unusually long duration as well. These prolonged disturbance processes are driven by intensification of the equatorial ring current which has an equatorial location, as opposed to the electrojet current that has a higher latitude position. Because of the previously mentioned large excursions in Dst that are possible, a series of observations and simulations were conducted to estimate GIC magnitudes that are possible in the exposed 500kV grid of central Japan. Figure 13 provides a trend line projection compared to paired observations and calculations of GIC levels in the regional power grid. This trend line and

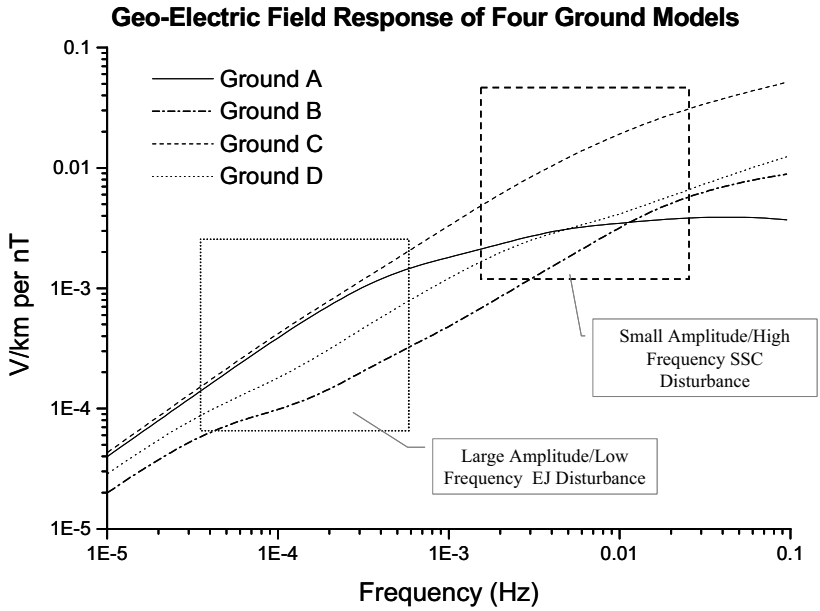

Figure 12. Interaction of large amplitude/low frequency electrojet disturbance and small amplitude/high frequency SSC disturbance with frequency-dependent characteristics of ground models.

companion simulations indicate GIC magnitudes at low latitude location power grids that could reach levels of 100 Amps of GIC for Dst levels reached during the March 13-14, 1989 Superstorm. The simulations utilized 1 second cadence magnetic observatory data, which is not available for a prospective Dst disturbance of ~1700. The high cadence observations provided sufficient spectral content details on the turbulent ground-level horizontal magnetic field disturbances during this storm. However, the observational details are not available for the speculative higher intensity storms such as that of September 1859. Therefore it is difficult to project with any certainty whether the trend line established in Figure 13 will also prevail to significantly higher Dst storm levels.

5.3 Electrojet Intensification Disturbance Scenarios

At high and mid-latitude locations, intensification of auroral or electrojet current systems in the ionosphere can produce very intense impulsive disturbance of the geomagnetic field over wide spread regions. It was predominantly these disturbance processes that triggered most of the power system disturbances over North America during the March 13-14, 1989 Superstorm.

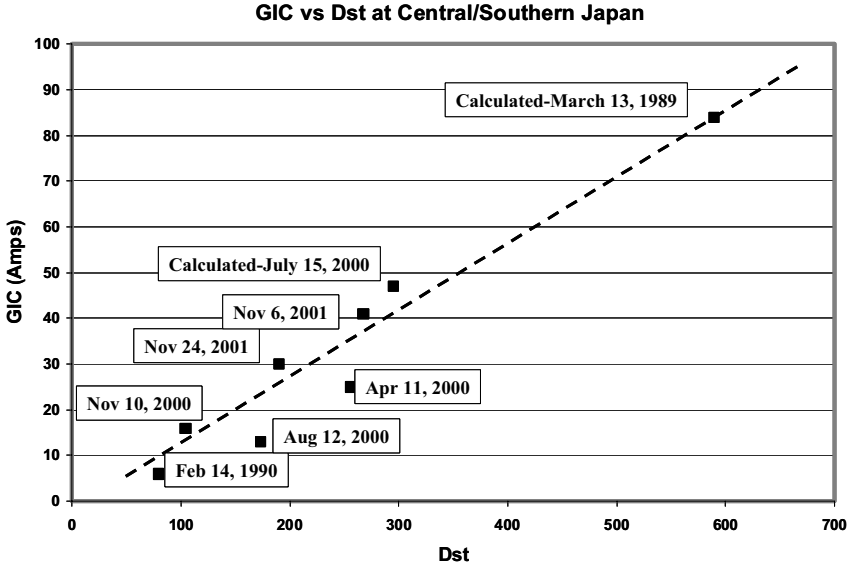

Figure 13. Trend of GIC flows and observed and calculated GIC flows in 500kV transformer in central Japan power grid due to ring current intensification at low latitude locations.

In addition to the Hydro Quebec blackout, the March 13-14, 1989 Superstorm caused numerous and widespread power system problems across North America. The NERC, in their post analysis, attributed over 200 significant anomalies across the continent to this one storm [NERC]. The intensity of the disturbance that triggered the Hydro Quebec collapse was at a level of 400 nT/min, while the most intense disturbance observed in North America was ~900 nT/min at the GLL observatory in southern Manitoba. In further assessing the disturbance intensities produced by this storm, the BFE observatory in Denmark observed the largest dB/dt with an intensity of ~2000 nT/min, a disturbance more than twice as intense than any experienced in North America [Kappenman-2001]. This observatory situated at ~50o geomagnetic latitude is at an equivalent latitude to mid-Atlantic regions across the US. Had this substorm erupted a few hours later, it would have been positioned over North America and could have caused a level of intensity that the power grids in the US have not faced in modern times. The last known disturbance approaching this level of dB/dt was observed over western portions of North America on August 4, 1972 [Anderson]. Less than 40% of the present-day power grid existed at that time and even smaller fractions of the extensive 500kV and 765kV grid that now extends across the US. The empirical experience from smaller intensity disturbances during the March 1989 storm over the US suggest such large disturbance events would

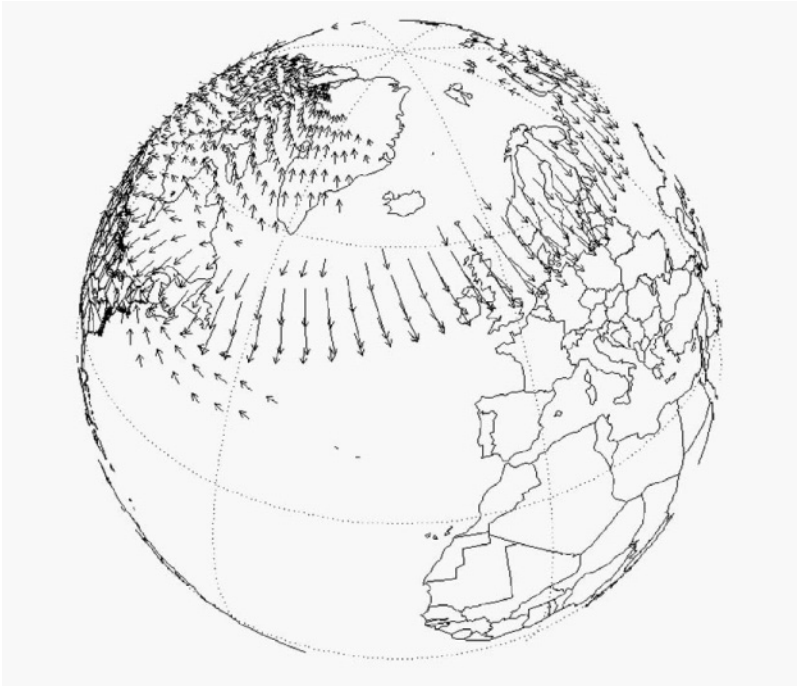

Figure 14. Extensive Westward Electrojet-driven geomagnetic field disturbances at time 22:00UT on March 13, 1989

have the potential to cause far greater power grid impacts in present day grids.

Data assimilation models provide further perspectives on the intensity and geographic extent of the intense dB/dt of the March 1989 Superstorm. Figure 14 provides a synoptic map of the ground-level geomagnetic field disturbance regions observed at time 22:00UT. The previously mentioned BFE observations are embedded in an enormous westward electrojet complex during this period of time. Simultaneously with this intensification of the westward electrojet, an intensification of the eastward electrojet occupies a region across mid-latitude portions of the western US. The features of the westward electrojet extend longitudinally ~120^{o} and have a north-south cross-section ranging as much as 5o to 10^{o} in latitude.

Older storms provide even further guidance on the possible extremes of these specific electrojet-driven disturbance processes. A remarkable set of observations was conducted on rail communication circuits in Sweden that extend back nearly 80 years. These observations provide key evidence that allow for estimation of the geomagnetic disturbance intensity of historically important storms in an era where geomagnetic observatory data is unavailable.

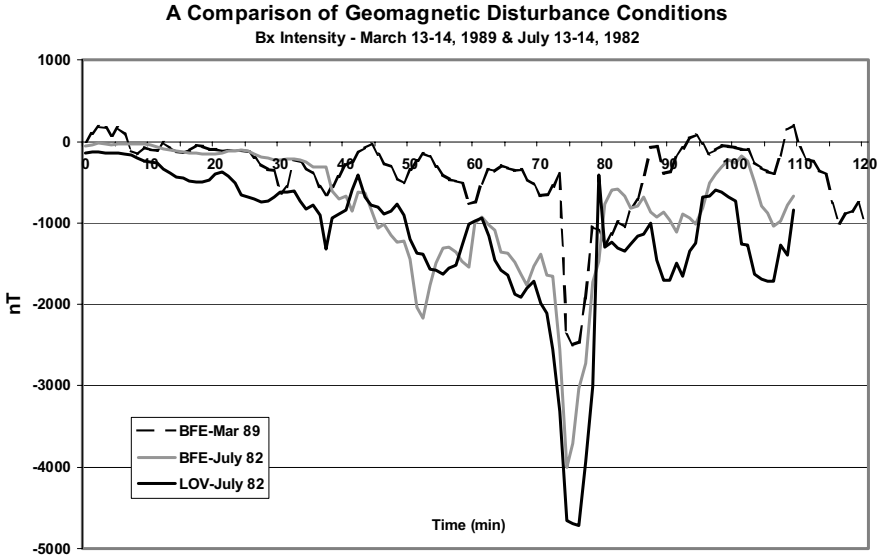

Figure 15. Comparison of observed delta Bx at Lovo and BFE during the July 13-14, 1982 and March 13, 1989 electrojet intensification events.

During a similarly intense westward electrojet disturbance on July 13-14, 1982, a ~100km length communication circuit from Stockholm to Torreboda measured a peak geo-potential of 9.1 V/km [Lindahl]. Simultaneous measurements at nearby Lovo observatory in central Sweden measured a dB/dt intensity of ~2600 nT/min at 24:00 UT on July 13. Figure 15 shows the delta Bx observed at BFE and Lovo during the peak disturbance times on July 13 and for comparison purposes the delta Bx observed at BFE during the large substorm on March 13, 1989. This comparison illustrates that the comparative level of delta Bx is twice as large for the July 13, 1982 event than that observed on March 13, 1989. The large delta Bx of >4000nT for the July 1982 disturbance suggests that these large field deviations are capable of producing even larger dB/dt impulses should faster onset or collapse of the Bx field occur over the region.

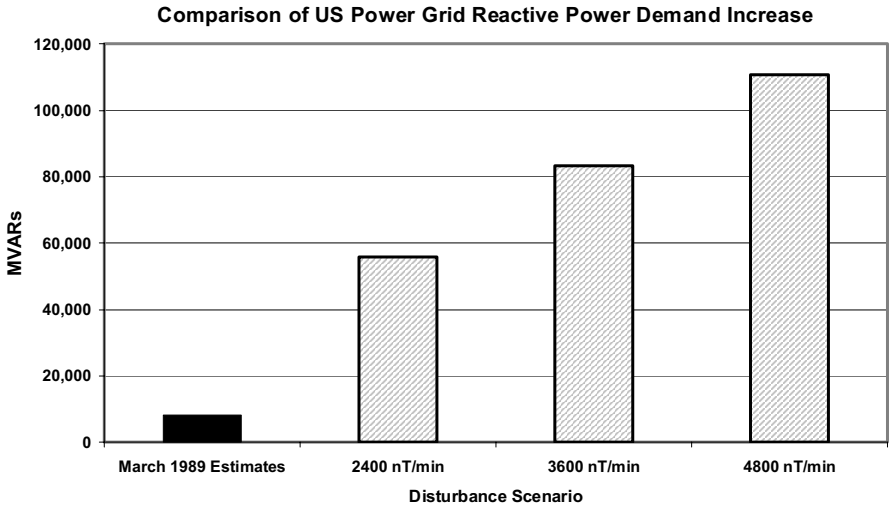

Figure 16. Comparison of estimated US power grid reactive demands during March 13, 1989 Superstorm and 2400, 3600, and 4800 nT/min disturbance scenarios at 50^{o} geomagnetic latitude position over the US.

6. US POWER GRID SIMULATIONS FOR EXTREME DISTURBANCE EVENTS

Based upon these extreme disturbance events, a series of simulations were conducted for the entire US power grid using electrojet-driven disturbance scenarios with the disturbance at 50^{o} geomagnetic latitude and at disturbance strengths of 2400, 3600, and 4800 nT/min. The electrojet disturbance footprint was also positioned over North America with the previously discussed longitudinal dimensions of a large westward electrojet disturbance. This extensive longitudinal structure will simultaneously expose a large portion of the US power grid.

In this analysis of disturbance impacts, the level of cumulative increased reactive demands (MVARs) across the US power grid provides one of the more useful measures of overall stress on the network. This cumulative MVAR stress was also determined for the March 13, 1989 storm for the US power grid, which was estimated using the current system model as reaching levels of ~7000 to 8000 MVARs at times 21:44 to 21:57UT. At these times, corresponding dB/dt levels in mid latitude portions of the US reached 350 to 545 nT/min as measured at observatories such as FRD, OTT and NEW. This provides a comparison benchmark that can be used to either compare absolute MVAR levels or, at a minimum, relative MVAR level increases for the more severe disturbance scenarios. The higher intensity disturbances of

2400 to 4800 nT/min will have a proportionate effect on levels of GIC in the exposed network. GIC levels more than 5 times larger than those observed during the above mentioned periods in the March 1989 storm would be a probable. With the increase in GIC, a linear and proportionate increase in other power system impacts is likely. For example, transformer MVAR demands increase with increases in transformer GIC. As larger GIC's cause greater degrees of transformer saturation, the harmonic order and magnitude of distortion currents increase in a more complex manner with higher GIC exposures. In addition, greater numbers of transformers would experience sufficient GIC exposure to be driven into saturation, as generally higher and more widely experienced GIC levels would occur throughout the extensive exposed power grid infrastructure.

Figure 16 provides a comparison summary of the peak cumulative MVAR demands that are estimated for the US power grid for the March 89 storm, and for the 2400, 3600 and 4800 nT/min disturbances at the different geomagnetic latitudes. As shown, all of these disturbance scenarios are far larger in magnitude than the levels experienced on the US grid during the March 89 Superstorm. All reactive demands for the 2400 to 4800 nT/min disturbance scenarios would produce unprecedented in size reactive demand increases for the US Grid. The comparison with the MVAR demand from the March 89 Superstorm further indicates that even the 2400 nT/min disturbance scenarios would produce reactive demand levels at all of the latitudes that would be ~6 times larger than those estimated in March 1989. At the 4800 nT/min disturbance levels, the reactive demand is estimated, in total, to exceed 100,000 MVARs.

This disturbance environment was further adapted to produce a footprint and onset progression that would be more geo-spatially typical of an electrojet-driven disturbance, using both the March 13, 1989 and July 13, 1982 storms as a template for the electrojet pattern. For this scenario, the intensity of the disturbance is decreased as it progresses from the eastern to western US. The eastern portions of the US are exposed to a 4800 nT/min disturbance intensity, while, west of the Mississippi, the disturbance intensity decreases to only 2400 nT/min. The extensive reactive power increase and extensive geographic boundaries of impact would be expected to trigger large-scale progressive collapse conditions, similar to the mode in which the Hydro Quebec collapse occurred. The most probable regions of expected power system collapse can be estimated based upon the GIC levels and reactive demand increases in combination with the disturbance criteria as it applies to the US power pools. Figure 17 provides a map of the peak GIC flows in the US power grid and estimated boundaries of regions that likely could experience system collapse due to this disturbance scenario.

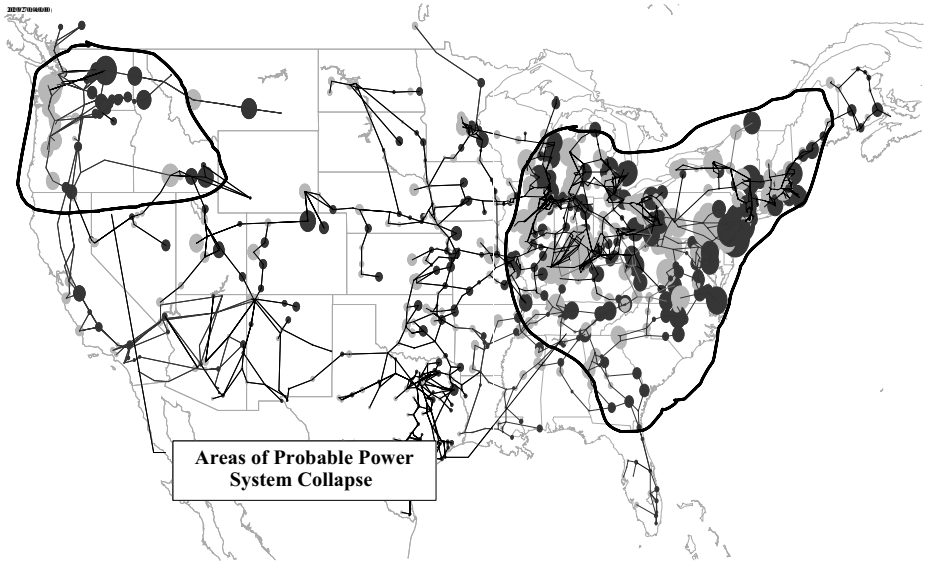

Figure 17. Regions of large GIC flows and possible power system collapse due to a 4800 nT/min disturbance scenario.

In addition to unprecedented levels of reactive power demands, the geographic region of possible power grid collapse is beyond any prior power industry experience with power grid failures, exceeding even the boundaries of the Great Northeast US Blackout of 1965. The 1965 blackout plunged major metropolitan areas of the US northeast (including Boston and New York City) into an extended blackout that lasted in excess of 12 hours duration before load restoration began. Both the size of the projected area of impact and the larger and more complex grid that exists today would present even greater obstacles in the grid restoration process. For an outage of this extent, the process of restoration and recovery could extend days, assuming minimal permanent damage occurs to the power grid infrastructure. The population within the above noted regions of collapse exceeds 100 million. This potential large-scale impact also raises legitimate concerns about the numerous interdependent infrastructures and vital public services that require electric power supplies. For instance in a timeline of only several hours of power supply outage, supplies of potable water for this large population become a concern. Within a day, this concern is further compounded by the probable loss of perishable foods.

7. CONCLUSIONS

Contemporary models of large power grids and the electromagnetic coupling to these infrastructures by the geomagnetic disturbance environment have matured to a level in which it is possible to achieve very accurate benchmarking of storm geomagnetic observations and the resulting GIC. As abilities advance to model the complex interactions of the space environment with the electric power grid infrastructures, the ability to more rigorously quantify the impacts of storms on these critical systems also advances. This quantification of impacts due to extreme space weather events is leading to the recognition that geomagnetic storms are an important threat that has not been well recognized in the past. These capabilities for detailed analysis and have also enabled the development of predictive tools to help the power industry deal with these threats.

New understandings of the complex nature of geomagnetic disturbance environments at low to high latitude locations and the increasing ability of grids of higher kV design to conduct large GIC flows are also changing the view of risks that power grids may face due to the space weather environment. It is no longer the case that power grids at high latitudes which are in close proximity to auroral electrojets are the only power systems that are *at-risk* due to GIC impacts. SSC and ring current intensifications can cause equivalently large GIC's in power grids located even at equatorial latitudes. Ultimately the combination of regional deep-earth ground conditions and the design of the power grid itself will determine the extent of possible GIC risk that will occur for a power system. The geo-electric field responses of regional ground conditions are highly uncertain, but all ground strata exhibit uniformly high degrees of frequency dependency and non-linear response across the frequency range of concern for geomagnetic disturbance environments. While more work is needed to better define the regional risk factors due to ground conductivity conditions, there is near unambiguous evidence that higher kV-Rated power grid designs are likely to experience relatively larger GIC flows for any geomagnetic disturbance condition or grid latitude location. The prevailing design evolution of power grids have greatly escalated this aspect of risk modifier as the power systems have grown in size and kV operating voltages. Because of this, kV rating is a more appropriate initial screening for determining GIC risk for power grids. In other words, power grids with operating voltage levels of 400kV or greater are all potentially at risk no matter where they may be located in the world.

Improving understanding of both storm processes and the interactions with power grid infrastructures are forcing a change in basic assessments of which power grids face risks from geomagnetic storms and for what reasons.

The risk implications extend to power grids that have never considered the risk of GIC previously because they were not at high latitude locations. In contrast to these previous notions, latitude location is not as important a consideration of GIC risk as that due to grid design and related risk factors. Both studies and observation evidence are indicating that power grids even at equatorial locations can have large GIC flows. In initial screening for determining GIC risk for power grids, operating voltage levels are proving to be a more relevant screening criterion. In other words, grids with operating voltage levels of 400kV or greater are all potentially at risk.

8. REFERENCES

Albertson, V.D., J.A. Van Baalen, Electric and Magnetic Fields at the Earth's Surface Due to Auroral Currents, *IEEE Transactions on Power Apparatus and Systems*, PAS-89, pg 578-584, 1970.

Albertson V. D., J. G. Kappenman, N. Mohan, G. A. Skarbakka, "Load-Flow Studies in the Presence of Geomagnetically-Induced Currents," *IEEE PAS Transactions*, Vol. PAS-100, February 1981, pp. 594-607.

Anderson, C.W., L.J. Lanzerrotti, C.G. Maclennan, Outage of the L-4 system and the geomagnetic disturbances of August 4, 1972, *Bell System Technology Journal*, 53, 1817, 1974.

Campbell, W. H., J. E. Zimmerman, "Induced Electric Currents in the Alaska Oil Pipeline Measured by Gradient Fluxgate and SQUID Magnetometers", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-18, No. 3, July 1980, pp. 244-250.

Campbell, W.H., Introduction to the Electrical Properties of the Earth's Mantle, *PAGEOPH*, Vol 125, Nos. 2/3, pages 193-204, 1987.

ECAR, How the Aging of Major Equipment Impacts Reliability, *ECAR Electric Equipment Panel Report*, 99-EEP-61, May 1999.

Elovaara J., et.al, Geomagnetically Induced Currents in the Nordic Power System and their Effects on Equipment, Control, Protection and Operation, *CIGRE* Paper 36-301, 1992 Session, 30 August - 5 September, 1992, Paris, 36-301, 11 pp., 1992.

Erinmez, I. A., S. Majithia, C. Rogers, T. Yasuhiro, S. Ogawa, H. Swahn, J. G. Kappenman, "Application of Modeling Techniques to assess geomagnetically induced current risks on the NGC transmission system", *CIGRE* Paper 39-304/2002-03-26, Paper presented at CIGRE 2002.

FERC, *FERC Form 1 Data*, US Federal Energy Regulatory Agency, Washington DC, 2000.

Howlett, R.D., T.K. Lungren, Bonneville Power Administration, Transformer Impedances, Transmission Impedances (345-138kV), 525 kV Transmission Impedances, *TOP-Network Planning System Electrical Data Book*, Revised 2002.

Kappenman J.G., V. D. Albertson, N. Mohan, "Current Transformer and Relay Performance in the Presence of Geomagnetically-Induced Currents," *IEEE PAS Transactions*, Vol. PAS-100, March 1981, pp. 1078-1088.

Kappenman J.G., D. L. Carlson, G. A. Sweezy, "GIC Effects on Relay and CT Performance," Paper presented at the EPRI Conference on Geomagnetically-Induced Currents, November 8-10, 1989, San Francisco, CA.

Kappenman, J.G., L. J. Zanetti, W. A. Radasky, “Space Weather From a User’s Perspective: Geomagnetic Storm Forecasts and the Power Industry”, *EOS Transactions of the American Geophysics Union*, Vol 78, No. 4, January 28, 1997, pg 37-45.

Kappenman J.G., “Geomagnetic Storm Forecasting Mitigates Power System Impacts,” *IEEE Power Engineering Review*, November 1998, pp 4-7.

Kappenman, J.G., Chapter 13 - “An Introduction to Power Grid Impacts and Vulnerabilities from Space Weather”, *NATO-ASI Book on Space Storms and Space Weather Hazards*, edited by I.A. Daglis, Kluwer Academic Publishers, NATO Science Series, Vol. 38, pg. 335-361, 2001.

Kappenman, J.G., SSC Events and the associated GIC Risks to Ground-Based Systems at Low and Mid Latitude Locations, *Paper submitted to AGU International Journal of Space Weather*, May 2003.

Lanzerotti, L.J., Geomagnetic induction effects in ground-based systems, *Space Sci. Rev.*, 34, pg 347-356, 1983.

Lindahl S, Swedish Railway Authority measurements from July 13-14, 1982. Personal communication from Sture Lindahl, May 22. 2002.

Masse, R.P., Crustal and Upper Mantel Structure of Stable Continental Regions in North America and Northern Europe, *PAGEOPH*, Vol 125, Nos. 2/3, pages 205-239, 1987.

NERC, “The 1989 System Disturbances” *NERC Disturbance Analysis Working Group Report*, “March 13, 1989 Geomagnetic Disturbance”, pp. 8-9, 36-60.

Pirjola, R., Estimation of the Electric Field on the Earth’s Surface during a Geomagnetic Storm, *Geophysica*, Vol. 20, No. 2, pp. 89-103, 1984.

Pirjola R., M. Lehtinen, “Currents produced in the Finnish 400kV power transmission grid and in the Finnish natural gas pipeline by Geomagnetically-induced electric fields”, *Annales Geophysicae* 3, 1985, pp. 485-491.

Rasmussen, T. M., R. G. Roberts, L. B. Pedersen, "Magnetotellurics along the Fennoscandian Long Range Profile," *Geophys. J. R. astr. Soc.*, Vol. 89, pp. 799-820, 1987.

Rasmussen, T. M., "Magnetotellurics in Southwestern Sweden: Evidence for Electrical Anisotropy in the Lower Crust?" *JGR* Vol. 93, No. B7, pp. 7897-7907, July 10, 1988.

Siscoe G.L., A Quasi-Self-Consistent Axially Symmetric Model for the Growth of a Ring Current through Earthward Motion from a Pre-Storm Configuration, *Planet. Space Sci.*, pp 285-295, Vol. 27, 1979.

Tsurutani, B. T., W. D. Gonzalez, G. S. Lakhina, and S. Alex, The extreme magnetic storm of 1–2 September 1859, *J. Geophys. Res.*, 108(A7), 1268, doi:10.1029/2002JA009504, 2003.

Chapter 15

Space Weather Effects on Power Transmission Systems: The Cases of Hydro-Québec and Transpower New Zealand Ltd

Jean Béland[1], Kevin Small[2]

1. *Institut de recherche d'Hydro-Québec (IREQ), Varennes, Québec J3X 1S1, Canada*

2. *Transpower New Zealand Ltd, Wellington, New Zealand*

Abstract Space weather has long been known to effect electric power systems, these effects can range in scale from the barely noticeable to the catastrophic. This paper reviews two events , one the 1989 collapse of the Hydro-Québec system which ranks as probably the most significant power system event tracable to geomagnetically induced currents and two, the 2001 event on the Tranpower system in New Zealand, which while significantly less severe did cause plant failures on a system that had no previously considered geomagnetically induced currents a threat to power quality and security of supply

Keywords Electric Power Transmission, Geomagnetic Storms, Geomagnetically Induced Currents, Transformer Saturation, Hydro-Québec, Transpower New Zealand

1. INTRODUCTION

This paper presents two occurrences of geomagnetically induced current (GIC) that have affected high voltage networks situated almost to the antipodes one from the other. During the first event on March 13 1989, the whole 735 kV network of Hydro-Québec in Canada collapsed, causing a blackout which affected several million people. This case is well known and here we put the emphasis on preventive and corrective actions that have been taken since then to avoid a repetition of such an event in the future.

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 287-299.

The second case took place in New Zealand, where Transpower is the owner and operator of the High Voltage AC and DC power grid. On November 6 2001 Transpower experienced the first known disruption to its grid system caused by GIC. We present the events of the day and attempts to mitigate the risks. We also discuss the damage caused by this storm and the operational procedures put in place to manage this risk.

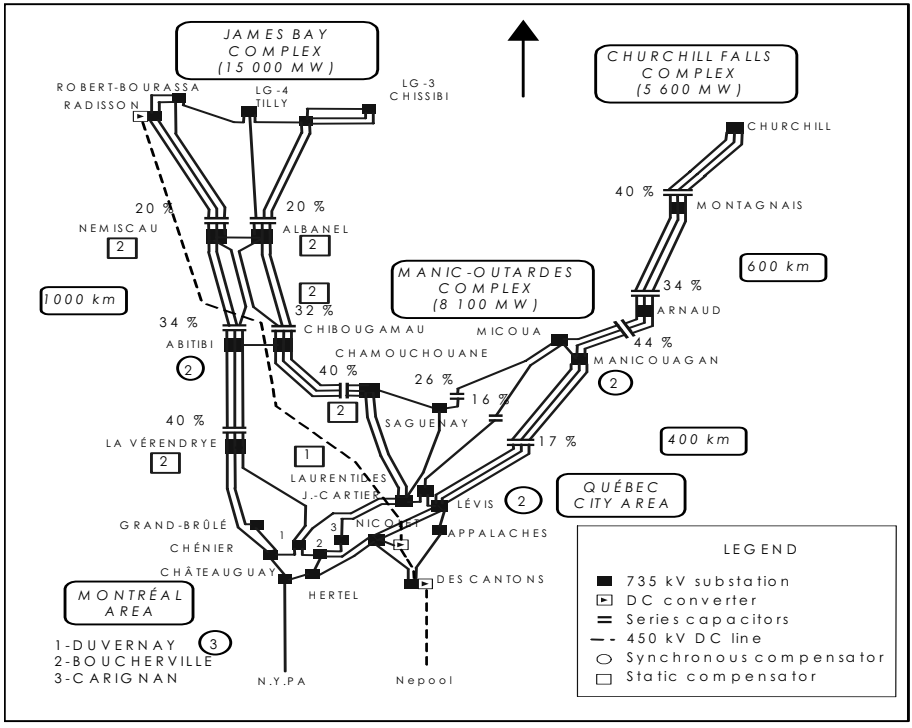

Figure 1. The Hydro-Québec's 735 kV transmission network.

The present Hydro-Québec 735 kV main transmission network is illustrated in Figure 1. The system has 35000 MW of generation with more than 90% from hydro-electric sources. Two major corridors, each 1000 km long separate the major production centers from the main load situated in the Québec and Montréal area. The long lines and poor ground conductivity over most of the northern region make the system susceptible to GIC. System operation is supported by 11 static compensators (7 were in place in 1989), and (since 1995) series compensation on some corridors.

On March 12 1989, voltage instabilities occurred in the evening and had to be corrected by network operators. During the night (March 13) at 02:45 Eastern Standard Time, all 7 static compensators on the James Bay corridor

tripped in less than one minute. Voltage started to collapse and the James Bay complex separated 8 seconds later. Power oscillations caused Churchill Falls to separate also after 6 seconds. Finally, the complete network collapsed less than 20 seconds following the separation of Churchill Falls.

More than a decade later, a very different story happened in another part of the world. The High Voltage electric power transmission system in New Zealand is owned and operated by Transpower New Zealand ltd.

The system is composed of a 220kV and 110 kV AC system on each of the two main Islands joined by a HVDC link. Up until the time of this event Geomagnetically Induced Currents (GIC) had not been recorded on the power system and no damage had been reported. The risks to the power system due to GIC events has been considered to be low due to several factors including:

- The power system has relatively short spans, and covers a short distance when compared to Canada or the USA.
- The natural geography aligns the system on a North /South basis.
- The use of Neutral Earthing Resisters on many transformers.
- New Zealand has a mid latitude position

The utilisation of neutral resistors is justified by a particular need. The South Island of New Zealand has an abundance of hydro generation, which is transmitted, to the more populous and industrial North Island by the use of an HVDC link. At times the return path for the electric current is the ground itself, which can cause problems due to the earth currents rising up through the earthed points of power transformers and onto the AC system as this forms a low resistance path back to the HVDC terminal station.

To prevent this occurrence Neutral Earthing Resistors (NERs) were fitted to the earthed point of the Transformers 'Star' windings. These resistors were fitted with Supervisory Control and Data Acquisition (SCADA) monitoring and alarms. Power system operators monitoring these currents and alarms are instructed to re-configure the HVDC equipment when alarms are received.

Despite these factors on 06 November the 2001 a Kp 8 GIC event occurred causing the loss of voltage support equipment and the destruction of a power system transformer.

In Section 2, this paper discusses several actions that have been undertaken at Hydro-Québec to avoid a similar event in the future. Section 3 is devoted to the Transpower case.

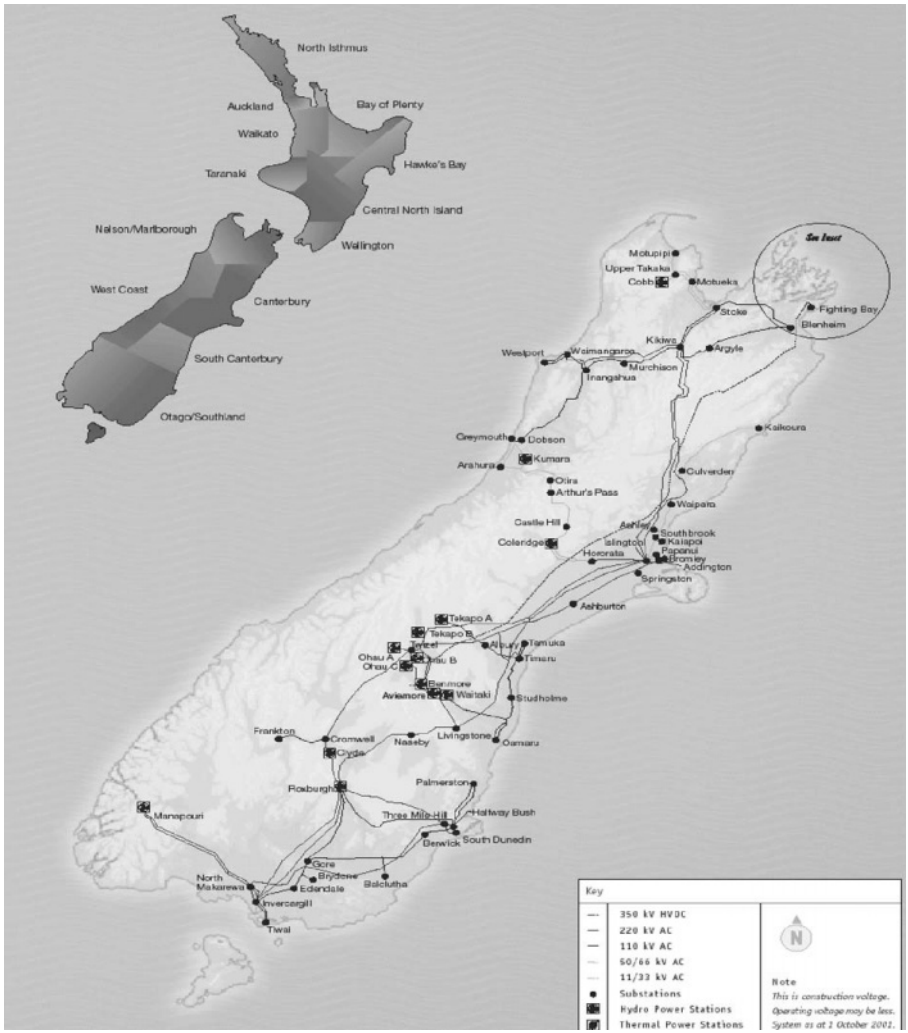

Figure 2. The Transpower New Zealand transmission network.

2. PREVENTIVE AND CORRECTIVE ACTIONS AGAINST GEOMAGNETIC ACTIVITY AT HYDRO-QUEBEC.

Actions have been taken in three different areas: measurement systems and alert services, safe network exploitation rules during geomagnetic activity (measured or predicted) and finally, network modifications.

2.1 Measurement systems and alert service

2.1.1 The voltage phase angle measurement system

The first version of this system has been put in operation about 20 years ago. Over the years several improvements have been implemented. The present version has been running since the end of 1995.

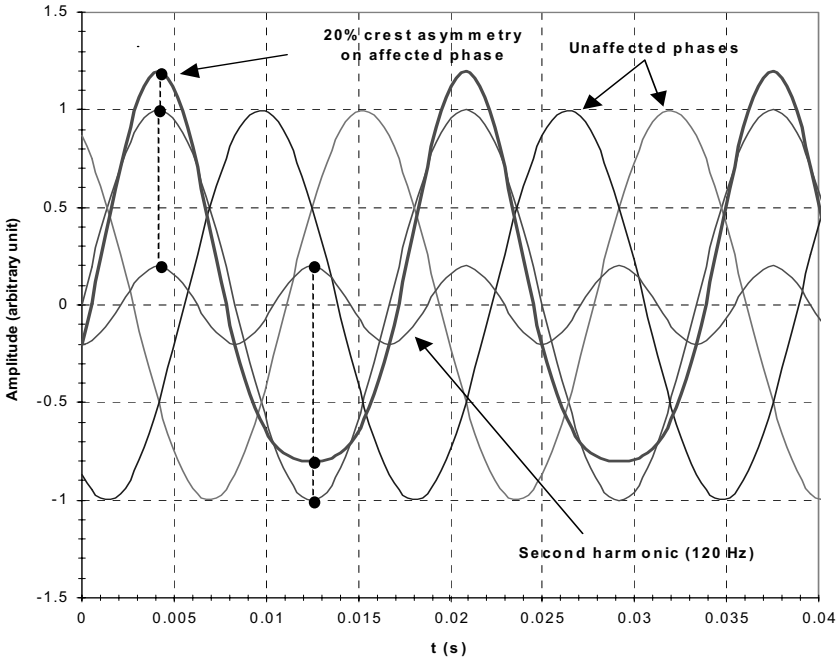

Figure 2. Even harmonics and crest asymmetry.

The system uses 8 measurement units, each of them located in a different 735 kV substation. Each unit measures and transmits the voltage phase angle, the frequency and the positive and negative crest voltage at every cycle, i.e. 60 times/s. Using dedicated phone lines and modems those data are transmitted in real time and on a permanent basis to a central acquisition unit. At this location, some data are recorded for post-mortem studies when specific triggers are exceeded. Other data are processed (to drastically reduce bandwidth and also to eliminate transmission errors) and forwarded in real time to the power system control center.

The crest voltage measurement enables the central acquisition unit to calculate the crest voltage asymmetry at every cycle. Numerical filtering is used to increase signal/noise ratio and resolution of the measurement. As shown in Figure 2, the asymmetry is a way to detect even harmonics which

might be produced if geomagnetically induced currents circulate in the neutral of power transformers. Asymmetry level from each measurement unit is received at the control center every 5 seconds; immediate alert is generated as soon as a minimal level is exceeded (currently between 2.2 and 3.0% depending on the substation).

The whole system is currently being improved. New measurement units will use GPS synchronization. Instead of transmitting crest voltages that could be affected by harmonics, those units will measure the direct sequence voltage phasor, a parameter combining amplitude and phase information that is less sensitive to harmonics and more representative of the "true" voltage in most circumstances. Also, the frequency content up to the 8th harmonic will be calculated using a fast Fourier transform algorithm. Single-phase data coming from up to four lines will be transmitted once every four seconds.

2.1.2 The Geomagnetic storm detection system (DOGME)

This second measurement system has been build to act as a backup for the SMDA. As a consequence its measurement capability are limited and strictly oriented toward low bandwidth, geomagnetically induced current effects.

The system has a measurement unit in four 735 kV substations. Each of them measures the second and fourth harmonics level on the line voltage, and transmits this information in real time to the control center. Data is logged on a permanent basis. An alert is generated when the sum of harmonics 2 and 4 is greater than a given threshold (actual value is 3%).

Data from the DOGME system can be visualized using a web page available to the network operators and other people. This page presents plots of harmonics levels at every measurement point. It also includes pointer to the estimated real-time Kp index produced by the NOAA Space Environment Center. Pointers to the SMDA data acquired by the control center are also included.

2.1.3 Geomagnetic alert from specialized provider

Real time measurement of crest voltage asymmetry and harmonic level enables us to know when geomagnetic effects are present but unfortunately without any lead time.

Solar magnetic disturbance (SMD) alert service used by Hydro-Québec is available on a permanent basis. An expert is *always* validating data at all time. Alert are sent only if a well-identified event exceeding a minimal threshold is detected, or if its probability of occurrence is high. All alert messages are short and clear. They are to be read by network operators, not SMD experts. Alerts transmission is done by e-mail, fax and pager. Finally,

all alerts are followed by an end-of-alert message, to tell the operators when they can resume normal network operation.

2.2. Special Exploitation rules during geomagnetic activity

To maximize the stability margin during geomagnetic activity (predicted by an alert or measured using the SMDA or DOGME systems), Hydro-Québec uses specific exploitation rules to operate its network in such situation.

- During a storm exceeding a given threshold, the maximum transit capacity of all major lines is reduced by 10%, relative to the normal case.
- Maximize the number of lines in service and the spinning reserve.
- Suspend test in progress.
- Minimize maneuvers to avoid instability.
- The threshold is actually set at 2.2% crest voltage asymmetry on SMDA, or storm of intensity Kp = 8 or greater.

The 10% transit reduction is based on a Northeast Power Coordinating Council (NPCC) policy to limit to <90% of their security limit the maximum transit of all "critical" interconnections. NPCC's document "Procedures for Solar Magnetic Disturbances" (document "C-15") basically states what follows:

- Suspend maintenance and restore HV lines.
- Adjust loading of HV dc circuits from 40% to 90% of their nominal rating, and critical transmission lines at <90% of their security limit.
- Reduce the loading of generator operating at full load to increase spinning reserve.
- Maximize the reactive power reserve by using equipment capable of synchronous condenser operation.
- Ensure with personnel that SMD monitoring equipment is in service.

Additional information about the NPCC (members, area covered...) can be found on the web at www.npcc.org.

2.3 Modifications to the network.

2.3.1 Corrective action on static compensators.

Following the 1989 event, examination of static compensators revealed that their protection limit was set too low. For this reason the tripping limit has been significantly raised on a permanent basis. Moreover, during a geomagnetic alert condition, the tripping is blocked and reduced to an alarm signal only. Note that the total number of static compensators is currently 11 (7 in 1989).

2.3.2 Addition of series compensation.

On a transmission line, series compensation means that capacitors are added in *series* with that line. This has two important effects. The first one (and the most important) is that a series compensated line appears to be electrically shorter than an uncompensated one: it becomes more "rigid". The direct consequence is an increased stability, i.e. a network less susceptible to power oscillations. This better stability can be used to increase the power delivered by parts of the network. The second and very fortunate effect is that since a capacitor blocks any DC current, it also blocks GIC since they are quasi DC currents.

The addition of series compensation on several 735 kV lines (completed in 1996) is mostly justified by its positive effects on the stability of the network and was planned before the 1989 event. After the blackout a review of this decision took place and it has been concluded that by stopping the circulation of DC current over many lines, less harmonics would be produced by GIC. The location of blocking capacitors is indicated on Figure 1.

Actual measurement of the crest voltage asymmetry by the SMDA seems to indicate lower values than before, but it is difficult to conclude firmly. Table 1 gives maximum asymmetry levels observed during major between years 1998 and 2001. Those are to be compared with levels in excess of 10% observed in the past.

	Geomagnetic storm		
Date	**Geomag. Intensity**	**Max. Asym. (SMDA)**	**735 kV substation**
20010411	K 8	1,6	Micoua
20010331	K 9	3,9	Micoua
20000715	K 9	4,3	Châteauguay 765
20000406	K 8	7,6	Châteauguay 765
19991022	K 8	2,2	Micoua
19981107	K 8	2,5	Tilly
19980925	K 9	1,9	Micoua
19980827	K 8	1,6	Micoua
19980806	K 8	1,8	La Grande 2P
19980504	K 9	3,1	Micoua

Table 1. Maximum asymmetry level (1998-2001).

2.3.3 Addition of the automated operation of shunt inductors system (MAIS system).

Another modification to the network is the MAIS system, which is operational since 1995. Its main objective is to help in the dynamic control of the voltage following major events.

In each substation the MAIS control unit may use local voltage, frequency and reactive power as decision variables. In general, reactors are switched using only voltage threshold and voltage rate of change.

The system increases the stability of the network and helps the static and synchronous compensators to maintain adequate voltage.

3. GIC EFFECTS AT TRANSPOWER ON NOVEMBER 6 2001; MITIGATION AND RISK MANAGEMENT

3.1 The event

On 6 November 2001 at 14:53 (New Zealand summer time) Alarms from NERs across the South Island were received by the HVDC operator. Simultaneously the Static Var Compensator (Voltage control equipment) for Christchurch city (latitude 43.53) tripped along with a transformer feeding Dunedin city (latitude 45.85).

A fault on the HVDC was quickly eliminated as a cause, and it took another 15 to 20 minutes for people to start suspecting a GIC event.

Confirmation was found by hunting on the Internet and finding a warning on the NOAA Space Environment Center web-site (www.sec.noaa.gov).

While no further damage was sustained NER alarms remained active for many hours after the initial event.

Figure 3 shows the onset of the GIC and the rise in current through the NER on transformer T13 at a hydro power station 'Ohau C' (OHC). The magnetic data is the horizontal component and is recorded at Eyrewell magnetic observatory in Christchurch (latitude 43.25). There is a strong correlation between the change in nT over time and the level of induced current through the transformer NER.

3.2 Equipment damage.

The transformer at Dunedin / Halfway Bush (HWB T4) was of unit construction i.e each of the 3 phases is a separate physical unit. The transformer tripped on red phase Buchholz sudden oil pressure protection and a dissolved gas analysis indicated a major internal flashover.

The remaining units were unaffected and a spare unit replaced the damaged red phase and the transformer was back in service in a matter of hours.

An internal inspection revealed damage to the insulation material. The transformer was beyond repair and it was subsequently written off. Subsequent to this event almost a year later a close in ground fault caused another phase of this transformer to be damaged in a simular way, once again it was beyond repair.

The GIC took affect on all concerned items of equipment at 14:52. The Transformer T4 at Halfway bush failed within one minute of this occurrence. It does not seem possible for a transformer of this sort to fail this quickly due to saturation of the core creating hotspots etc. However, deterioration of transformers is cumulative and caused by events such as power system faults, electrical overloading GICs etc which over time degrades the transformer. It is possible that the transformer was already prone to failure and the GIC was the final contributor, hence failing very soon after the event.

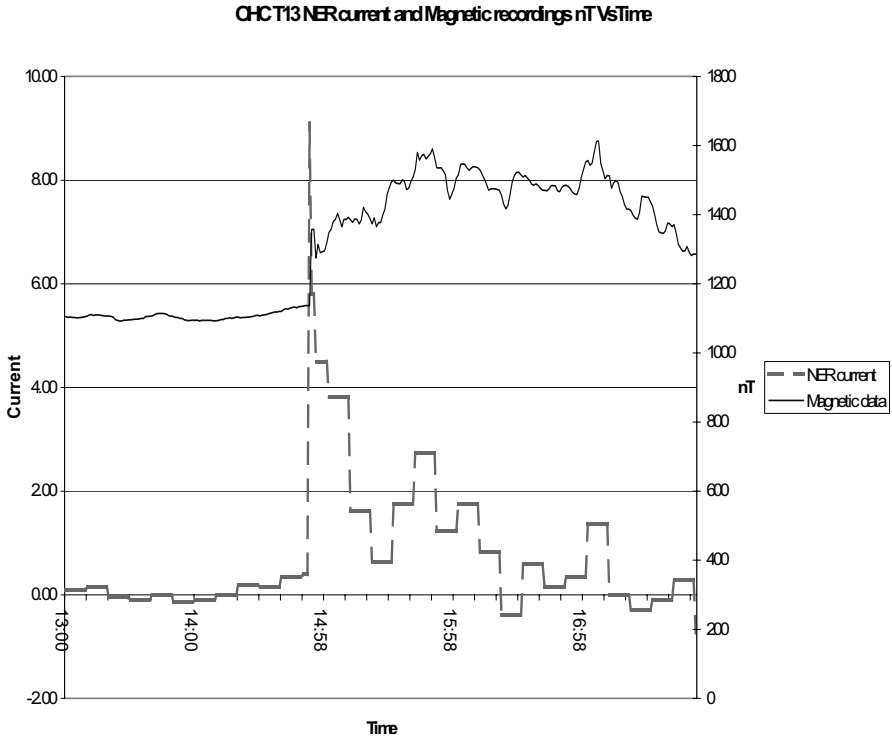

Figure 3. OHC T13 NER current and magnetic recordings (nT) vs time.

The SVC device at Christchurch (Islington) was not damaged at all by the event, the SVC was 'tripped' by the negative sequence over-current relay. This protection has operated previously and in all cases 2nd harmonic (transformer saturation) was present in the waveform.

This tripping is considered to be a fault in the protection system and counter measures have been applied.

3.3 Introduction of contingency plans.

Transpower has subsequently implemented plans in an effort to prevent further events on the power system.

We now subscribe to the NOAA SEC warning system and on receipt of a warning a communications and operational process is put in place. The object of the operational process is to reduce the amount of geomagnetically induced current into the system and manage operations and maintenance work so as to minimise the impact. Power flow studies are undertaken to ascertain whether certain transmission circuits with an east/west aspect can

be removed from service. Likewise for vulnerable interconnecting transformers. At the onset of NER alarms the equipment is removed.

3.4 The New Zealand electricity market

A competitive electricity market operates in New Zealand. This market uses a nodal pricing model. This system is also use in many other Electricity Markets and means that each ‘node or point of supply has a different price representing the cost of losses and constraints involved in supplying electrical energy to that point.. Removing equipment from service can therefore have an effect on the price of electricity. While the equipment we remove is spare capacity it may still affect prices under many generation profiles. This then gives us another direct cost of GIC events. Processes are also in-place to calculate these in the event of the plan being enacted.

4. CONCLUSION

Following the 1989 event, Hydro-Québec took several remedial actions:

- To receive geomagnetic alerts from specialized organizations.
- To operate real time measurement systems able to detect GIC effects and to alert immediately the network operators.
- To apply conservative exploitation rules during strong geomagnetic activity, measured or predicted.
- To modify the protection of its static compensators.
- The addition of series compensation and the automated operation of shunt inductors system both ensure a more robust network, less susceptible to GIC disturbances.

We now think that our main network can operate correctly during any realistic geomagnetic storm.

At Transpower, the contingency plans were used on several occasions during the GIC events in November 2003 there was an immediate reduction in the NER currents which may suggest that the plans were effective, a more detailed analysis is required however before a final conclusion can be made.

5. ACKNOWLEDGEMENTS

The authors would like to thank some people from Hydro-Québec. Denis Larose and Antonio Dutil who gave them access to several data regarding the March 13 1989 events and its consequences. The data on Table 1 have

been compiled by Georges Blais. Finally, special thanks to Léonard Bolduc for his precious collaboration and his long-standing work about geomagnetic storms effects at Hydro-Québec.

In addition thanks also to those below for much valuable information and assistance.

David Boteler from Geological Survey, Canada.

John Kappenman from Metatech.

Murray Dryer from NOAA

6. REFERENCES

Béland J, "SMDA4: Manuel de référence (HP-1000/A-990)", Hydro-Québec internal report, 1998.

Bolduc L, "GIC observations and studies an the Hydro-Québec power system", Journal of Atmospheric and Solar-Terrestrial Physics, 64, p.1793-1802, 2002.

Dutil A, "Impact des orages géomagnétiques sur le réseaude transport de TransÉnergie-État de la situation à l'aube de l'an 2000", Hydro-Québec internal report, 2000.

Larose D, "The Hydro-Québec system blackout of March 13, 1989", Special Panel Session Report, IEEE PES summer meeting, Long Beach, CA, July 1989, p.10-13, 1989.

Northeast Power Coordinating Council, "Procedure for solar magnetic disturbances which affect electric power systems", Document C-15, November 2000. Retrievable at www.npcc.org/procedures.htm

Théorêt M, "Système de mesure du décalage angulaire- Logiciel de l'unité centrale-Guide de l'usager", Hydro-Québec internal report, 1995.

Chapter 16

Neutron Monitor Network in Real Time and Space Weather

H. Mavromichalaki[1], V. Yanke[2], L. Dorman[2,3], N. Iucci[4], A. Chilingaryan[5], O. Kryakunova[6]

1. *Physics Department, Athens University, Greece*
2. *IZMIRAN, Russian Academy of Science, Moscow, Russia*
3. *Cosmic Ray Center, affiliated with Tel Aviv University and Israel Space Agency, Israel*
4. *Dipartimento di Fisica "E. Amaldi", 'Roma-Tre' University, Rome, Italy*
5. *Erevan Physical Institute, Armenia*
6. *Institute of Ionosphere, Alma-Ata, Kazakhstan*

Abstract: Relativistic cosmic rays (galactic and solar) registered by neutron monitors at the Earth, bring valuable information on their interaction with interplanetary disturbances. Therefore, they can play a useful role in forecasting space weather storms and in specifying magnetic properties of CME shocks and ejecta. The reconstruction of pitch-angle distribution of high-energy particles may be derived from ground level cosmic ray (CR) observations well in advance of the onset of geomagnetic storm. This can be used for forecasting. High energy solar particle events during powerful solar flares are registered at the Earth well before the main development of particle profiles recorded onboard GOES. This provides a good chance of a preventive prognosis of dangerous particle flux by ground level observations. To produce real-time prediction of the phenomena, only real time data from Neutron Monitor Network (NMN) should be employed. The increased number of NM stations operating in real-time gives a good basis for using NMN as a single multidirectional tool and for improving the definition of the onset of GLEs in powerful SPEs and to give an immediate forecast of the arrival of the interplanetary disturbance at the Earth. The properties of the Neutron Monitor Network and its possibilities for Space Weather tasks are discussed in this paper. Different real time Neutron Monitor Network topologies, different synchronization methods and the ways of collecting data in a central data server accessible to the users, are also discussed.

Keywords Neutron Monitor Network, cosmic rays, space weather, interplanetary disturbance, space environments and effects, predictors.

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 301-317.

1. INTRODUCTION

What is the meaning of “bad” space weather? It is a situation in which a complex phenomenon at the Sun causes interplanetary perturbations, essentially influencing the Earth environment and different aspects of the human activity. It is difficult to overestimate the extent of possible loss due to the bad space weather. In any case, it is better to try to prevent the effects of bad space weather than to pay for its consequences. For this purpose an operational space weather monitoring has to be necessary to provide a preventative space weather forecast. Space weather signatures appear in many solar-terrestrial and space environmental parameters. At the Earth’s orbit, on average, a heliospheric storm occurs every 4-5 days, leading to significant changes in solar wind characteristics. Disturbances of the solar wind, magnetosphere and cosmic rays (CR) are closely related, since they caused by the same active processes at the Sun. A large heliospheric storm, as indicated by different space weather parameters, is shown in Figure 1, where significant variations in CR density and in the first harmonic of the CR anisotropy, derived from ground level observations, occur simultaneously with dramatic changes in the interplanetary and geomagnetic parameters.

The effect of the solar wind disturbances on cosmic rays may extend to large distances, and, due to their relativistic velocity, cosmic rays bring information on these disturbances well in advance of their arrival at the Earth. Characteristic signatures in CR behaviour may be selected by special methods from neutron monitor network (NMN) data and input to Space Weather applications. Real time data in combination with developed and tested methods, should be used for successful prediction.

The neutron monitor network is already at a new stage of collecting and processing continuously recorded information. This epoch began in 1997, when Moscow neutron monitor data were the first to appear in real time on the Internet (http://helios.izmiran.rssi.ru/cosray/main.htm). Since that time a number of monitors started to operate in real or quasi-real time, including Apatity, Oulu, Athens, Erevan 2000, Erevan 3000, Tixie Bay, Yakutsk, ESOI, Irkutsk, Norilsk, Lomnitsky Stit, Inuvik, McMurdo, Newark, Nain,

Pewanuk, South Pole and Thule. The Athens real-time NM station already records data with 1-second resolution and work has begun to design a system with resolution higher than 1-second. Other neutron monitors plan to be providing real time data in the next few years. This provides a basis for continuous monitoring of Space Weather hazardous effects. For the presentation of data on-line, it is necessary to create special databases, as well as accurate real time synchronization of NMs, This can be achieved through the use of a GPS interface. Preliminary results and models of Space

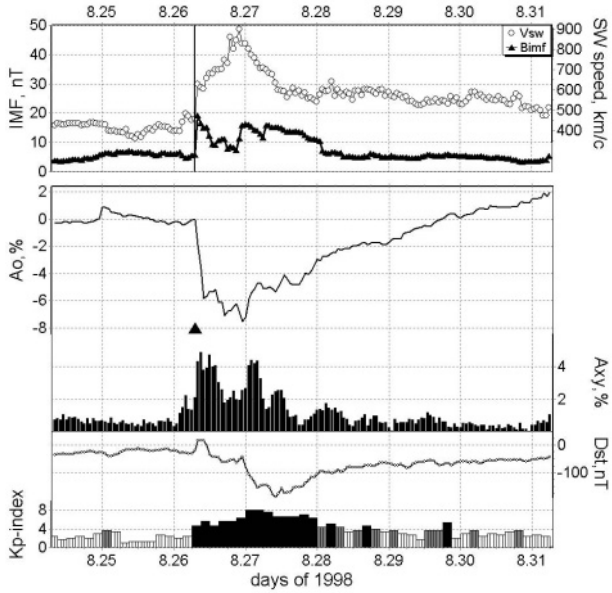

Figure 1. Interplanetary disturbance, a great Forbush effect and a severe magnetic storm during September 1998. A_0 is the 10 GV CR density variations, A_{xy} is the amplitude of the equatorial component of CR vector anisotropy.

Weather forecasting with the use of NMN -measured CRs, have already extensively discussed in a series of reports (Belov et al., 1995; 2001a; b; Bieber et al., 1999; Dorman et al., 2001; Kudela et al., 2000).

2. EXAMPLES OF PRECURSORS IN CR DATA

2.1 Solar Proton Events

One of the hazardous space weather effects is the increase of radiation doses in geo-space during powerful proton events originating on the Sun. Low-energy protons (10-300 MeV) are the most dangerous part of the solar

energetic particle (SEP) spectrum for satellite electronics and crew. Maximal flux of such particles can reach the Earth several (sometimes more) hours after the occurrence of the event on the Sun. High energy (> 1 GeV) particles from the solar proton event reach the Earth with a velocity close to that of light. Their flux cannot be recorded on satellites with enough accuracy because of the small detector square, but it is measured by ground-based neutron monitors (NM) with high statistical accuracy (in average, 0.5% for 5 min) as ground level enhancement (GLEs). In Figure 2 the profiles of particles of different energy during the powerful SEP event of 15 April 2001 are presented. One can see that the high-energy particle profiles registered at the Earth had already ended well before the main development of the low-energy particle profiles as recorded onboard GOES. The early detection of an Earth-directed SEP event by NMs gives a good chance of preventive prognosis of dangerous particle flux and can provide an alert with a very low probability of false alarm. The method developed in Villoresi et al. (2000) and Stoker et al. (2000) using 1-minute NM data (Dorman et al., 2001) from a single observatory, was applied to predict the spectrum of the approaching particles. Alternatively, a more feasible and statistically proven method, using total counts from several stations in real time, should be developed and used. Below a short description of this method is given:

Search for a significant increase (greater than 2.5σ, where σ is the statistical error of measurements) in the total counts of the 1, 2, or 5- minute data from several stations. If it is found, go to a state known as "alert-1". If not, continue to search.

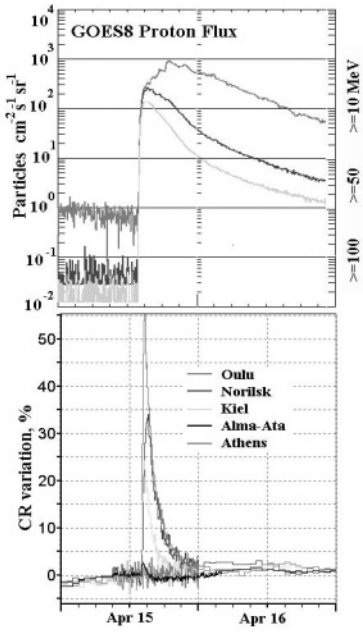

*Figure 2. Be*havior of different energy cosmic rays during the powerful proton event in April 2001.

While in the "alert-1" state:
i) Estimate the rigidity spectrum of the SEP event: $\Delta D(R)/D_0(R) = bR^{-\gamma}$, where $D_0(R)$ is the background galactic cosmic ray (GCR) rigidity (R) spectrum, ΔD is spectra augment at any fixed moment and γ is the spectral index.
ii) Calculate spectra parameters b and γ best fitted to the data from at least three different stations (Dorman et al., 2003)

Evaluate the energy-dependent diffusion constant for energetic particle propagation in inter-planetary space using the rigidity

spectra calculated from the three preceding one-minute data. This allows the determination of the energy spectrum at the source, by solving the inverse problem.

By using the source spectrum and the diffusion constants, predict the near Earth spectrum for a time window of ~ 1 hour. Compare on-line GOES measurements for the last several minutes (if available) with previous predictions, in order to refine subsequent predictions.

If the predicted flux at 100 MeV exceeds a pre-determined threshold, issue an "alert-2".

Repeat steps 2-5, until the total count returns to background level.

2.2 Geomagnetic storms

Another Space Weather effect on the Earth's environment is associated with propagation of interplanetary disturbances and their interaction with the Earth's magnetosphere. There are many direct and indirect data on the origin of coronal transient and the start of their propagation. However, CME observations become rather difficult after the initial stage; cosmic rays can provide an important tool for the study of their structure and the propagation from the source to the Earth.

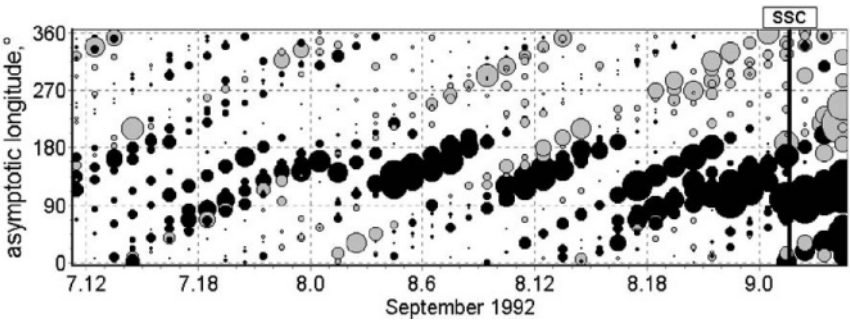

Figure 3. Angle distribution of the cosmic ray intensity on September 7-9, 1992, derived from the high latitude neutron monitors. Vertical incident asymptotic directions calculated for 9.5 GV rigidity, define asymptotic longitudes. 180° correspond to the stations looking Sunward. Gray circles indicate an increase and black ones a decrease of CR intensity. Size of the circle is proportional to the magnitude of variation. Isotropic intensity has been subtracted out.

The most significant near-Earth manifestation of large disturbances is the shock arrival, followed by the passage past the Earth of associated magnetic cloud. CR density and anisotropy vary significantly during these special times, as we can see in Figure 1. These changes may be used for short time prognosis, especially at the earliest moments of the disturbance development. The Forbush effect, as a heliospheric phenomenon, starts simultaneously with the disturbance emergence, well in advance of the geomagnetic storm onset. Due to their high speed and large mean free paths,

relativistic CRs bring information on interactions, such as “loss cone” distribution and shock reflected populations to the Earth in advance of the disturbance itself. In Figure 3 an example of changes in CR pitch-angle distribution before shock arrival is presented. In such cases a precursory flux decrease may result from a “loss cone” effect, where the CR monitoring station is magnetically connected to the cosmic ray depleted region behind the shock. Precursory increases may result from accelerated particles being reflected by the approaching shock. This anomalous pitch-angle distribution has very specific features: a decrease of the CR intensity within a narrow range of pitch-angles (as a rule < 50-60°) close to the IMF direction (usually sunward, more rarely antisunward); a large, sometimes > 1%, difference between the CR intensity from these and from other directions; a sharp transition between the regions with different intensities; a pitch-angle distribution which cannot be fitted by the sum of only the first two spherical harmonics (Belov et al., 1999; 2001b). This distribution, which is unusual for quiet periods, is rather typical of periods just before the Forbush effect. They can therefore be used as an early indicator of an approaching disturbance and as a predictor of a magnetic storm.

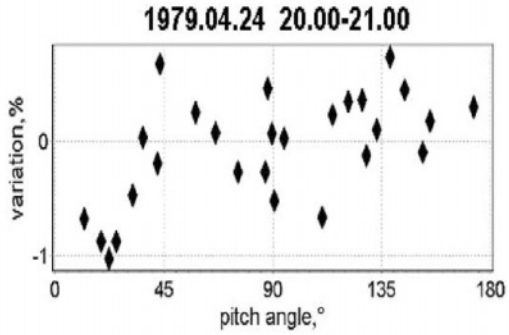

Figure 4a. Pitch-angle distribution of cosmic rays measured by different NMs (each diamond represents a different NMs) at 21:00UT on April 24,1979.

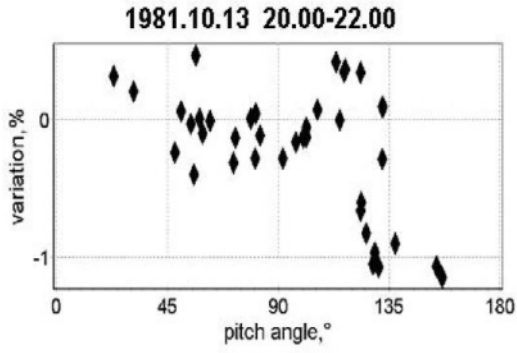

Figure 4b. Pitch-angle distribution of cosmic rays as measured by different NMs (diamonds) at 20-22 UT on 13 October 1981.

Figure 4(a, b) shows examples of such specific CR distributions for different events at the times before a shock arrival. Figure 4a displays an intensity deficit of about 1% near the sunward IMF direction, and a sudden jump from low to high intensity around 35-50°. For this event, the available monitoring stations were distributed uniformly in pitch angle. The shock arrived 3 hours later (23:58 UT), and a severe magnetic storm started immediately. The pitch-angle distribution plotted for the 20:00-22:00 period combined (Figure 4b) is like a mirror image of the distribution discussed above. This seems to be a case where the loss cone is in the antisolar direction. An antisolar predecrease is more rarely observed. Some examples were discussed in Belov et al. (1995; 1999), where such cases were associated with interplanetary disturbances originating from eastern solar longitudes. In this case, despite of leading part of disturbance went out the Earth's orbit at the time of predictor observation, the Earth still can get this disturbance later. Another possible cause of predecrease is that the disturbed interplanetary field is in a complicated loop-like configuration. Cosmic ray preincreases, caused by reflection and acceleration of ambient galactic CR from approaching shock, are also frequently observed prior to Forbush decreases (and magnetic storms). However, the anomalies in CR distributions discussed above (a narrow predecrease) are more unusual. The regularity of this behavior enhances the value of using the preincrease effect for prediction purposes. Therefore, in addition to improving basic knowledge of particle interactions with shocks, studies of precursors also suggest that ground level cosmic ray observations may be useful for space weather forecasting. A critical issue is whether these precursors can be reliably detected sufficiently in advance of the associated geomagnetic disturbance, to furnish a practical benefit. This depends on the reliability of the neutron monitor network and on the use of data in real time.

3. NEUTRON MONITOR NETWORK AND DATA PRESENTATION

Currently, the worldwide network (Shea and Smart, 2000; Moraal et al., 2000) consists of about 45 operational neutron monitors (see Figure 5) with different specific space-energy characteristics and responses to primary CRs.

The monitors are standard devices located at different points on the globe, recording secondary cosmic rays which associated with primaries in the energy range from hundreds MeV to hundreds GeV. These high energies represent an extension of the low-energy ranges measured on spacecraft. On the map of Figure 5 the globe distribution of neutron monitors (NMN) is shown together with isolines of different cut-off rigidities. NMs at different locations are suitable for studying different phenomena. For example, the high latitude network is essential for measuring anisotropies related to transient CR events, such as solar CR and Forbush effects. The homogenous energy response at high latitudes is utilized by placing NMs at high latitudes in such a way as to cover uniformly asymptotic directions on the celestial sphere.

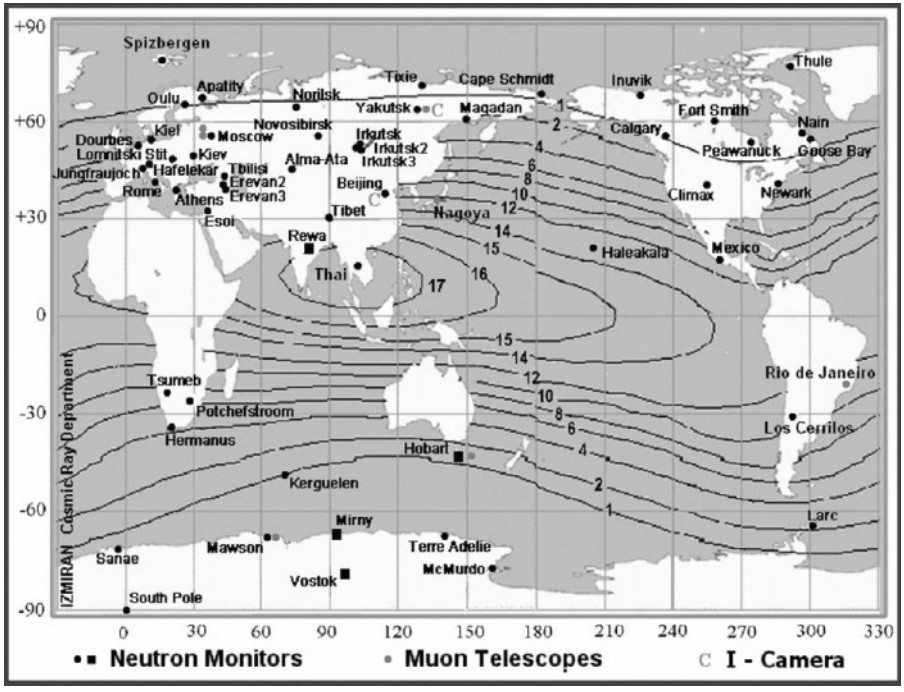

Figure 5. The global distribution of ground level cosmic ray detectors: neutron monitors and muon telescopes. CR cut-off rigidities at different points on the Earth are indicated on the isolines.

This special part of the network, with its separation of asymptotic directions, is named “Space Ship Earth” (Bieber et al., 1995) and it is presented in Figure 6, together with an indication of asymptotic directions for solar CRs at each station. All stations in Figure 6 have very high angular resolution (22o-57o) and receive median rigidity particles within a 22o nearly equatorial range. The regions of median energy directions in space are separated by less than 62o in longitude. Thus, this network provides a high angular resolution of particle equatorial distributions during solar proton events, exceeding the corresponding accuracy of measurements onboard satellites. This set of stations, together with some others at rigidity <4.5 GV, may be successfully used in the “ring station method” to monitor precursors of geomagnetic storms, as shown by Belov et al. (2001a). High rigidity (>5 GV) stations, such as Athens, Beijing, Rome, ESOI, Mexico, Tibet and many others, are necessary for the global survey method, to derive CR density, and anisotropy components, which exhibit anomalous behaviour before the arrival of an interplanetary disturbance at the Earth (Belov and Eroshenko, 2002). These high rigidity stations are also important to estimate spectra of solar cosmic rays during solar proton events. They can give information on the upper energy limit of solar particles in these events. In other words, the era when a single or a few NMs were used to analyse solar-terrestrial phenomenon, has passed. The NM network should be now considered as a unique multidirectional spectrograph. The development of special programs (Global Survey Method-GSM, Ring Station Method-RSM) allow the derivation at any moment of the CR density, anisotropy and CR pitch-angle distribution, by using as many neutron monitor stations as possible. Naturally, parts of the network may be

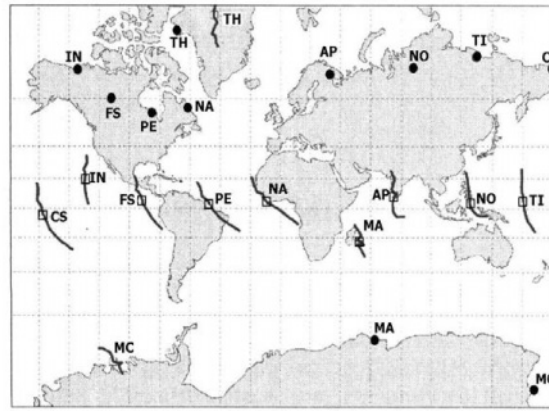

Figure 6. **Spaceship Earth stations** Circles are the geographic positions, squares-asymptotic cones of the stations:
IN–Inuvik **NO**–Norilsk **FS**-Fort Smith **TI**–Tixie Bay
PE–Peawanuck **CS**–Cape Shmidt **NA**–Nain **AP**–Apatity
MC–McMurdo **TH**–Thule **MA**- Mawson

still used for some selected tasks. The evolution of the number of neutron monitors since 1952 is presented in Fig. 7.

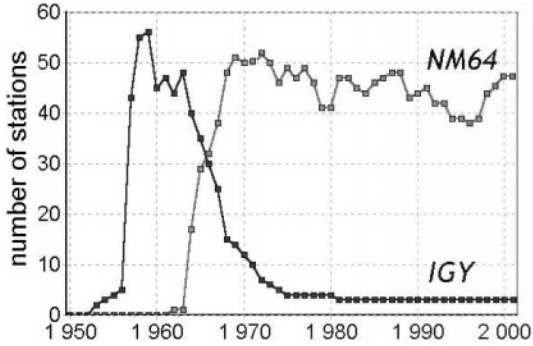

Figure 7. Number of IGY and NM64 neutron monitors over the history of CR observations.

The first version of the neutron monitor (the Simpson NM) was based on small counters, with low statistical accuracy. Their number increased abruptly after the famous flare of 23.02.1956 during the International Geophysical Year 1957 (IGY NM). At the beginning of 60's new counters and a new geometry for NMs were developed. Gradually these super neutron monitors (NM64) of high statistical accuracy replaced the old ones over the whole globe. It is clear that, even for the IGY neutron monitor or a single counter of the NM64, the accuracy is higher than that of CR observations on spacecraft. The use of all stations as a unified multidirectional detector, makes the accuracy substantially higher (<0.1% for hourly data).

All NMs operate continually with 1- or 5- minute intervals of data collection. Several recent results on the use of the prognostic properties of ground level CR observations suggest the need to provide continuous data in real time. Starting in July 1997, the Moscow NM64 was the first in the world to present data on the Internet in real time. After Moscow, several other stations became involved in this process, and now 23 stations present their data in real time, in digital and/or graphical form (Mavromichalaki et al., 2001). The main problem now is to make it possible to get all these data in real time in close sequence from all servers in order to make a real time monitoring of space weather conditions.

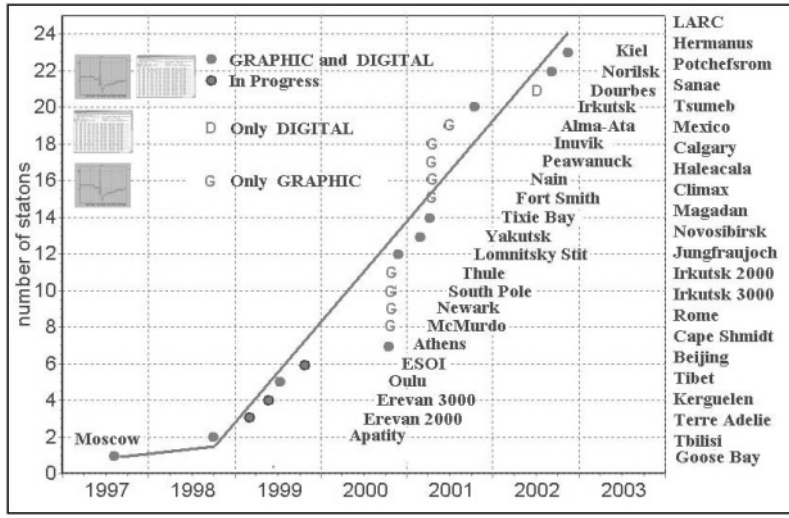

Figure 8. Diagram of the stations presenting data in real time (in the inner frame) and ready to present in a near future (list on the right side).

3.1 Data presentation

The most acceptable way for data publication on the Internet is via a standard interactive WWW-interface, providing access to the local databases of the CR stations, which altogether can be considered as and example of distributed database. The neutron monitor recording system, transfers one minute and hourly data to a server and refreshes this database, every hour. A special program included in a scheduler creates a graphical file once per hour, which is displayed on the web page. Standard access to the database is managed by FTP.

More modern processing systems have been established in Apatity and Oulu stations, in which the database is refreshed every 10 minutes. On request, a graphical file is produced by ISAPI technology on the server, and then displayed on the web page. The result is resent by HTTP protocol both in graphic and ASCII format. Standard access to the database from outside for the one-minute and hourly data (refreshed every 10 minutes) is via http protocol requests.

An improved version of 1-5-15-60-minute refreshed database system is operating at the stations of Athens, Kiel, Moscow and Irkutsk. Standard access to the database is http protocol request as well as FTP. Data may be retrieved both in graphical and digital form.

4. HOW TO GET DATA FROM THE NETWORK

4.1 Synchronization problem

To use data from the network stations in real time it is necessary to have high time precision in the NM registration systems. At present, the NM stations use PC clocks for timekeeping, but its timing uncertainty depends on the stability of the interrupt requests and any change in the interrupt request rate causes the clock to gain or loose time. So another way for the right timekeeping of NM stations must be found to obtain the best synchronization of the stations.

What accuracy is necessary to keep the clock precise? If the statistical accuracy of 1-minute data is about 1%, then the accuracy of 1-minute interval should not be worse than 1 sec. If the maximum change in the CR intensity during a GLE is about 30% per minute, the required time accuracy will be nearly 5 sec. These are not very strong requirements, and may be easily realized. The situation is more complicated when the data accumulation interval is 20 sec (Apatity), 10 sec (South Pole), or 1 sec (Athens). The timekeeping in these cases takes more efforts and some other approaches. We propose two different ways for correct timekeeping and NM stations synchronization.

1) Internet time synchronization

Computers can synchronize their clocks to an Internet timeserver. Special client software is needed in this case, linked to each of the three major Internet timing protocols: Time Protocol, Daytime Protocol, and Network Time Protocol. Timeservers are continuously “listening” for timing requests and will send them by using any of these three protocols. When the server receives a request, it sends the time to the NMN computers in the appropriate format. This timekeeping method has a number of advantages (it is cheap and easy to establish) and disadvantages, such as requirement for special software, dependence of the server-client time latency on the time of day and network health, the occasional loss of synchronization. If the Athens NM server was to act as the central timekeeper, the following time latencies (t_{AV}), observed on March 5/3/2003 at 18:00 U.T., give an indication of the performance:

Table I

Servers	t_{AV} (ms)
Apatity	866
Emilio Segre	225
Erevan	130
Kiel	121
Lomnicky Stit	495
Moscow	261
Roma	136

2) GPS synchronization

GPS can be used to determine precise time, time intervals, and frequency. GPS satellites carry highly accurate atomic clocks on board. For the system to work, GPS receivers on the ground synchronize themselves to this clock, what means every GPS receiver is, in essence, a clock with atomic accuracy. GPS can be used to synchronize clocks to tens of nanoseconds over large distances. The advantages of using GPS are the precision, the same time latency for all NM stations, the reliability and very fast synchronization. The disadvantages are the requirement for modern fast electronics, the need for special software and high prices.

4.2 Bases and ways of the data collection

For data collection by a central system (CS-Client) from all real-time neutron monitor stations (peripheral systems PS), different network configurations may be assumed. The basic network topology must be a “star-type”, but the means of data collection may be different. The network should run independently of the quality level of the operation of each PS; the data collection must be as fast as possible and the data must arrive at the CS simultaneously from all the stations. Here we discuss three data collection models, their benefits and drawbacks. One way, presented in Figure 9, is based on the use of the star topology, in such a way that the central system collects data via FTP from peripheral stations (PS). Each PS records the last measurement locally in an FTP server with read-only access properties, so that the system is more secure.

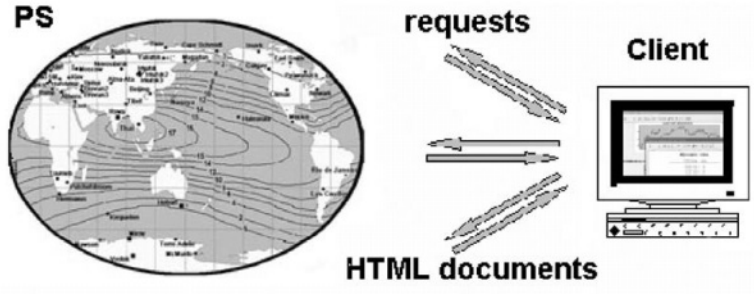

Figure 9. An example of real-time data collecting in the Central database (CS) of the NMN by a HTML request is given.

The CS requests and downloads the data one by one or simultaneously in a multithreading process, and provides the reference time. A problem of this system is that if a peripheral station fails, this will probably affect the central system, which would therefore need a timeout function while processing data from a peripheral station. In this case every PS must be an FTP server serving the CS. The management of so many servers can cause problems on the network.

A second way is based on the same star topology, but the PSs send their data to the CS at a specific time. In this model each PS will have a username and a password in CS and will send their data via the FTP protocol. The data will consist of a file with the latest measurements. In this topology, the system is "strong" because the failure of one PS does not affect the CS. If good enough synchronisation is provided, the data will reach the CS simultaneously. A possible drawback of the above method is the security of the central system, and extra security control will be needed. In any case, the management of the FTP process is easier because it has to be configured on only one FTP server, the central server. The main drawback of this method is that there is no way of the CS collecting data on request from the PS.

The third method is based on the second with some improvements taken from the first. In this case, the PS periodically sends data to the CS, but the CS has also the ability to request data from PS at any time. Such an approach is realized in the project RECORD (Real time Cosmic Ray Database), proposed by the Russian-European collaboration (Yakutsk, Moscow, Oulu, Lomnitsky Stit). Two or several regional databases (RS) should be managed (see Figure 10). One of the regional servers has the status of central data base, all others have a status of "mirrors". Comparison and balancing of regional databases is done by replication. The updating of regional databases uses many different sources, providing data by different ways (in particular, it can occur by the two ways mentioned above).

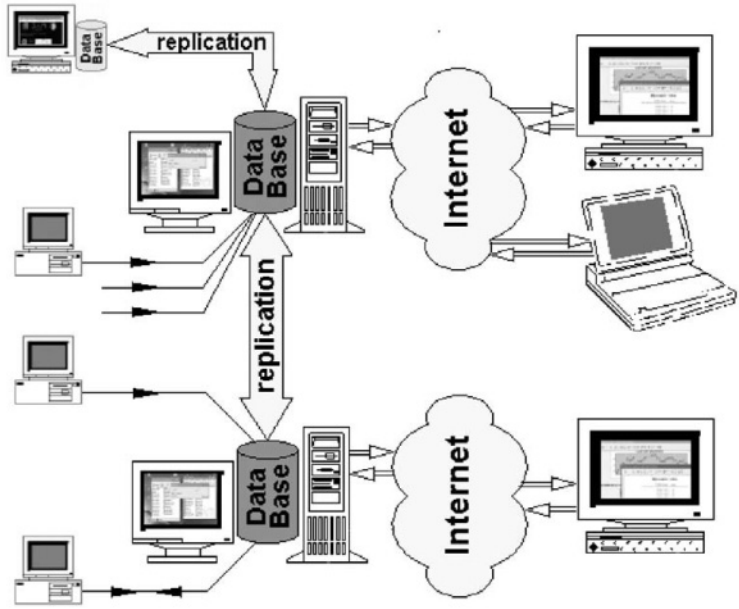

Figure 10. Schematic presentation of RECORD project.

If a station implements its own database on the same platform as in CS, the process of transferring data to the Central Database is simply by replication. In a second method, when a station sends data to the Regional database, the data transfer by TCP/IP Protocol should be established on the server of the station itself. In a third case the station is a passive element and only publishes data on its own server. The Regional Database gets data by a request to this server via http. It provides independence with respect to apparatus and programs. Users have an access to the Central DB via WWW, FTP and SQL interfaces.

5. CONCLUSIONS

Early detection of Earth-directed SEP events by NMs gives a good chance of preventive prognosis of dangerous particle fluxes in space and can provide an alert of an SEP with reasonable accuracy.

The worldwide neutron monitor network is a good tool for detecting anomalies in the CR pitch-angle distribution especially prior to the arrival of the interplanetary disturbance at the Earth. This can be used to give a good space-weather forecast.

At present, data from >15 neutron monitors are accessible in real time, and this provides a good basis for attempts to search for precursors of geomagnetic activity and to organize alerts of SEP events in real time.

The main task in the use of CR variations for space weather forecasting is to manage a real-time data presentation from as many neutron monitors as possible, collecting these data in real time on suitable databases accessible via Internet.

6. ACKNOWLEDGEMENTS

This work is partly supported by RFFR Grants 01-02-17580 and 03-07-90389 and by PENED 01EP87 and ENTER 01EΔ87 grants. We thank a lot for the useful discussion and great help with the preparing of this paper to Drs A. Belov, E. Eroshenko, L. Pustiljnik and also to postgraduate students C. Sarlanis and G. Souvatzoglou from Athens University. We are grateful to the Organizing Committee of ESPRIT for the chance to publish this paper.

7. REFERENCES

Belov A.V., L. Dorman, E. Eroshenko, N. Iucci, G. Villoresi and V.Yanke: Search for predictors of Forbush decreases, Proc. 24-th ICRC, **4**, 888-991, 1995.

Belov A.V., E.A.Eroshenko and V.G.Yanke: Cosmic Ray Effects Caused by Great Disturbances of the Interplanetary Medium in 1990-1996. Proc. 26th ICRC, 6, 431-434, 1999.

Belov A.V., E.A. Eroshenko, V.A. Oleneva and V.G Yanke: Relation of the Forbush effects to the interplanetary and geomagnetic parameters, Proc. 27-h ICRC, **9**, 3552-3355, 2001a.

Belov A.V., J.W. Bieber, E.A. Eroshenko, P. Evenson, R. Pyle and V.G. Yanke: Pitch-angle features in cosmic rays in advance of severe magnetic storms: Neutron monitor observations, Proc. 27-th ICRC, 9, 3507-3510, 2001b.

Belov A.V. and E. A. Eroshenko: Cosmic Ray Observations for Space Weather, Proc. for the 22nd ISTC Japan Workshop on Space Weather Forecast, Nagoya University, Japan, 129-146, 2002.

Bieber J.W., and P. Evenson: Spaceship Earth- An Optimized Network of Neutron Monitors, Proc. 24-th ICRC, v. 4, 1316-1319, 1995.

Dorman L.I., L.A. Pustil'nik, A. Sternlieb, and I. Zukerman: Using ground-level cosmic ray observations for automatically generating predictors of hazardous energetic particles levels. Adv. Space Res., 2001.

Kudela K., M. Storini, M. Hoffer, and A. Belov. Cosmic rays in relation to space weather. Space Sci. Rev., 93, 153-174, 2000.

Mavromichalaki, H., C. Sarlanis, G. Souvatzoglou, S. Tatsis, A. Belov, E. Eroshenko, V. Yanke and A. Pchelkin: Athens Neutron Monitor and its aspects in the cosmic-ray variations studies. Proc. 27th ICRC 2001, 4099.

Shea M. A. and D. Smart: Fifty years of Cosmic ray Radiation Data. Space Sci. Rev., 93, 229-262, 2000.

Stoker P. H., L. I. Dorman and J.M. Clem: Neutron monitor design improvements. Space Sci. Rev., 93, 361-380, 2000.

Villoresi G., L.I. Dorman, N. Iucci, and N.G. Ptitsyna: Cosmic ray survey to Antarctica and coupling functions for neutron component near solar minimum (1996-1997), 1. Methodology and data quality assurance. JGR, 105,. A9, 21025-21034, 2000.

Chapter 17

Space Weather Research and the US Air Force Office of Scientific Research (AFOSR)

Paul J. Bellaire Jr.
Air Force Office of Scientific Research, Arlington, Virginia 22203-1954, USA

Abstract The Air Force Office of Scientific Research recognizes of the importance of space weather to global technological infrastructure, placing a strong emphasis on targeted space physics research and development in partnership with other US federal agencies. This is exemplified by AFOSR participation in the interagency National Space Weather Program led by the National Science Foundation, the Community Coordinated Modeling Center for space weather at NASA Goddard Space Flight Center, international partnerships such as the Air Force Research Laboratory's International Research Initiative, and recent DoD Multidisciplinary University Research Initiatives in space weather.

Keywords Space weather, space physics, space environments and effects.

1. AFOSR AND SPACE WEATHER

The term "space weather" refers to conditions on the Sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems, and that can endanger human life or health. In the United States, the Air Force Office of Scientific Research (AFOSR) recognizes of the importance of space weather, placing a strong emphasis on targeted space physics research and development in partnership with other federal agencies. For further information concerning AFOSR, its mission, and its research interests, please visit http://www.afosr.af.mil/.

The National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), and the Department of Defense (DoD) provide the dominant share of basic research support for space physics in the United States. These mutually synergistic funding activities are coordinated

I.A. Daglis (ed.), Effects of Space Weather on Technology Infrastructure, 319-329.

through the Committee for Space Weather of the National Space Weather Program, or NSWP, which is led by the NSF.

The NSWP strategic plan, provided at http://www.ofcm.gov/nswp-sp/text/a-cover.htm, and the more recent NSWP implementation plan at http://www.ofcm.gov/nswp-ip/tableofcontents.htm, describe these activities in some detail. As part of the new emphasis on space weather, focal points for space physics research have naturally grown through these multi-agency partnerships.

2. AFOSR AND THE COMMUNITY COORDINATED MODELING CENTER

The Community Coordinated Modeling Center (CCMC) for space weather is one such focal point. The CCMC, located at NASA Goddard Space Flight Center in Greenbelt, Maryland, was designed with the philosophy that collaborative, comprehensive space weather model development and testing would benefit both the research and operations communities. In the past, there have been poor linkages between the two communities, even though there is significant convergence in performing space physics research and in improving operational space weather forecasting. Recognizing this opportunity, AFOSR provided critical funding and support to assist in the creation of the CCMC at Goddard.

The United States Air Force and National Oceanographic and Atmospheric Administration (NOAA) have operational responsibility for providing space weather forecasts and support to their respective customers. In the past, improvements to operational capabilities have been slowed by the lack of a comprehensive process to coordinate basic research and model development and to transfer value-added scientific tools to operations. Within the last few years, both NOAA and the Air Force have created Rapid Prototyping Centers (RPCs) in Colorado, which serve to implement space weather tools and models at these agencies' respective operational centers in Boulder, Colorado and Omaha, Nebraska. The CCMC works with both RPCs for the mutual benefit of scientific research and federal operations.

The CCMC concept was initiated in 1998 as a result of efforts by government agencies to enhance space weather research, develop space weather models, and provide a means for a more effective transfer of research models to operations. Thus, the CCMC is an integral element of the NSWP. The charter of the CCMC also emphasizes having close interactions and collaborations with the space science community. The CCMC participating agencies include the Director of Weather at

Headquarters US Air Force Pentagon, Air Force Materiel Command, AFOSR, the Air Force Research Laboratory (AFRL), the Air Force Weather Agency (AFWA), NASA, NSF, NOAA's Space Environment Center, and the Office of Naval Research (ONR).

The CCMC is managed through an interagency Steering Committee reporting directly to the Committee for Space Weather of the NSWP. The Steering Committee has completed a Concept of Operations (CONOPS) for the CCMC, which can be accessed through the CCMC web site found at http://ccmc.gsfc.nasa.gov. The CONOPS describes the functions of the CCMC and how it will be used to benefit space weather research and operations. The CCMC web site also includes information on how to request customized runs of space weather models and how to submit models for validation and testing. Community participation in CCMC activities, to include interested international researchers, is actively encouraged. Input can be submitted through the Steering Committee and Working Group members listed on the web site.

The success of the CCMC depends on active collaboration and coordination. Since the advent of the NSWP in 1995, space weather research funding has increased in the United States, efforts to develop space weather models have intensified, and the stakeholders have proliferated. In recent years, new research initiatives have added to the existing NSF and DoD funding (nominally a combined sum of about US$50M annually) for single investigator awards in space physics. The opportunity and necessity for efficient coordination among the various US agencies involved in funding space weather research is greater than ever.

3. AFOSR AND THE NASA LIVING WITH A STAR PROGRAM

NASA is currently targeting of hundreds of millions of dollars for a new research initiative in their Sun-Earth Connections program, called Living With a Star (LWS). NASA is now partnering with the DoD and NOAA in joint LWS activities, and the CCMC will likely be incorporated within the LWS program. AFOSR has also participated in the process of defining the Targeted Research and Technology effort within LWS.

In the coming years, LWS will be a source of new space weather understanding, as well as physics-based models that demonstrate the feasibility of specifying and forecasting space weather by effectively combining space-based data with state-of-the-art assimilative modeling. LWS seeks to answer three fundamental questions:

How and why does the Sun vary? How does the Earth respond? What are the impacts to humanity?

To answer these questions, LWS proposes interdisciplinary programs to quantify the physics, dynamics, and behavior of the Sun-Earth system over the 11-year solar cycle; improve our understanding of the effects of solar variability and disturbances on terrestrial climate; provide the data and scientific understanding required for advanced warning of energetic particle events that affect the safety of human spaceflight; and provide detailed characterizations of radiation environments useful in the design of more reliable electronic components for air and space transportation systems. More detail can be found at the LWS web site at http://lws.gsfc.nasa.gov/.

NASA, AFRL, the University of Texas at Dallas, and Spectrum Astro are currently collaborating under the LWS program to incorporate the Coupled Ion-Neutral Dynamics Investigation (CINDI) instrument on the Air Force's Communication/Navigation Outage Forecast System satellite (C/NOFS). AFOSR also funds research in support of the C/NOFS mission at AFRL. C/NOFS, to be launched early in 2004 on a Pegasus vehicle, will be the first satellite dedicated to real-time forecasting of ionospheric irregularities and ionospheric scintillation, both of which degrade and disrupt satellite communications and Global Positioning System (GPS) navigation. C/NOFS will be a Low Earth Orbit (LEO) satellite with a 13-degree inclination, a 700 km apogee, and a 400 km perigee. Six on-board instruments will sample and measure the ambient ionospheric plasma and neutral atmosphere, electromagnetic fields, line-of-sight total electron content (TEC), and radio scintillations, in order to make possible space weather forecasts for operational use. More information can be found on the C/NOFS web site at http://www.spectrumastro.com/PDFs/CNOFS-Web.pdf.

4. AFOSR INTERNATIONAL ACTIVITIES

An excellent example of interagency, as well as international, coordination is the satellite program known as the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC). COSMIC partners include NASA, NSF, University Corporation for Atmospheric Research (UCAR), AFWA, AFOSR, AFRL, ONR, the Naval Research Laboratory, the Directorate of Weather at Air Force Headquarters Pentagon, and Taiwan's National Space Program Office. The COSMIC web site can be found at http://www.cosmic.ucar.edu/index.html.

COSMIC is a Taiwan-USA joint project to build six microsatellites for GPS radio occultation observations, with launch anticipated in 2005. Once

on orbit, COSMIC will provide more than 3,000 GPS radio signal occultation limb soundings per day, globally and in all weather. COSMIC will enhance current global observing systems and provide much needed data for improved forecasting of space weather, as well as for basic research in ionospheric physics, meteorology, and climatology.

AFOSR also has been directly involved in other international research efforts. In 1999, an advanced upper atmospheric sodium lidar system was funded by AFOSR at Colorado State University under the DoD's Defense University Research Instrumentation Program. This US lidar system is now installed at the Arctic Lidar Observatory for Middle Atmosphere Research (ALOMAR) in Andøya, Norway. The ALOMAR observatory is located inside the Arctic Circle at 69 N 16 E, well within the auroral zone, and this lidar takes advantage of the opportunities for polar space weather science there, while leveraging many millions of dollars of laboratory infrastructure provided by Norway and its European Union partners. ALOMAR manages all ground-based instrumentation for the Andøya Rocket Range and provides excellent opportunities for collaborative science. ALOMAR instrumentation permits researchers to study the entire atmosphere from the troposphere to the lower thermosphere, as well as measure physical parameters in the ionosphere, magnetosphere, and aurora. ALOMAR is a unique arctic laboratory, providing research opportunities (through the EU Commission's Access to Research Infrastructures, or ARI, project) to scientists from the European Union and associated countries. Details are available on the ALOMAR web page at http://alomar.rocketrange.no/.

It is important to note that AFOSR reaches out to foreign collaborators through many pathways. Two AFOSR liaison offices in London and Tokyo support Air Force research goals by interacting with members of the international scientific and engineering community and encouraging open communication among them and Air Force scientists and engineers. The European Office of Aerospace Research (EOARD) in London, England deals with Europe, the mid-East, Africa, and countries from the former Soviet Union. The Asian Office of Aerospace Research (AOARD) in Tokyo, Japan has responsibility for Asia and Pacific Rim countries, including India and Australia. In addition, AFOSR hosts the Air Force Research Laboratory's International Office (AFOSR/IO). AFOSR/IO monitors international science policy, supports development of AFRL's international R&D strategy, and assists AFRL scientists with international projects by helping develop international agreements and providing training. AFOSR/IO closely monitors the Air Force's international research investments and all AFRL international activities.

AOARD and EOARD provide support to foreign scientists in three ways: the Window On Science program, the Conference Support Program, and

direct research contracts. The Window on Science program allows foreign researchers to visit potential AFRL partners and attend technical conferences in the US, as part of a scientific exchange that may lead to future research collaboration. Foreign scientists can also request up to US$5K in support from AOARD and EOARD for scientific meetings and conferences in their home countries under the Conference Support Program, provided the meetings are open to international and AFRL attendees. Small research contracts, typically not to exceed US$25K, are potentially available to foreign scientists wishing to establish a partnership with an AFRL researcher. Since 1999, AOARD and EOARD have helped AFOSR provide funds for space weather research and conferences in Norway, the UK, Greece, Switzerland, Russia, Ukraine, Australia, and Mongolia. More information about AFOSR/IO, AOARD, and EOARD opportunities can be found at http://afosr-io.afosr.af.mil/content/int_business.asp.

Expanding beyond the nominal AOARD and EOARD programs, AFOSR/IO administers the International Research Initiative (IRI), which may competitively fund more robust projects (in the US$50-100K range). Over the last few years, AFOSR has awarded IRI funds to space weather researchers in Taiwan, Japan, and the UK. AFOSR funds IRI projects to encourage unique foreign research that supports AFOSR science goals and because the potential benefits often go well beyond basic research. A description of the IRI program can be found on the AFOSR/IO web page at http://afosr-io.afosr.af.mil/content/int_IRI.asp.

In 2002, AFOSR initiated partnerships in Australia that should further enhance international collaboration in space weather. The Tasman International Geospace Environment Radar (TIGER) is a part of the Super Dual-Auroral Radar Network (SuperDARN), a collaboration of international polar radars designed to help further our understanding of the geospace environment and supported by the National Science Foundation. TIGER currently consists of a single radar in Tasmania, but AFOSR has recently provided funding to La Trobe University in Australia to initiate the construction of a second TIGER radar on the south island of New Zealand. Space physicists are now using TIGER and SuperDARN to unravel the complex processes involved in transfering energy from the solar wind into the Earth's magnetosphere and ionosphere. The new TIGER radar will improve upon present capabilities by enabling tomographic monitoring of the Antarctic region south of Australia and New Zealand, contributing to enhanced forecasts of the space weather effects that can disrupt global communications and navigation. Information on the SuperDARN network can be found at http://superdarn.jhuapl.edu/, while TIGER is described in more detail at http://www.tiger.latrobe.edu.au/.

The World Institute for Space Environment Research (WISER), with headquarters at Adelaide University in South Australia, is an international network of Centers of Excellence in space physics and high performance computing. WISER is dedicated to promote collaboration in cutting-edge space environment research and in training of first-rate space scientists. Emphasis is placed on theoretical and computational studies of space plasmas and atmospheres, space data analysis, space weather forecasting, and monitoring the impact of space weather on the Earth's environment and technology. AFOSR provided initial seed funding for WISER in 2002, and this international interest encouraged the government of South Australia to contribute financial support through 2005. The WISER web page at http://hermes.physics.adelaide.edu.au/wiser/ provides detail from the 2002 kick-off meeting in Adelaide, as well as detailed descriptions of future plans.

Also in 2002, the Learmonth Solar Observatory at Exmouth, Western Australia received AFOSR funding (through a subcontract from the University of Arizona's Spacewatch program) to deploy a small commercial telescope for the detection and tracking of Near-Earth Objects (NEOs), which are a potential hazard to spacecraft and a source of dust that can damage satellite instruments. This telescope joins another AFOSR-funded NEO observing facility that Spacewatch operates in Mongolia with the support of AOARD. Details concerning the University of Arizona's Spacewatch project are available at http://spacewatch.lpl.arizona.edu/.

For the last several years, AFOSR has funded the University of California at San Diego to develop the Solar Mass Ejection Imager (SMEI), in collaboration with AFRL and the University of Birmingham in the United Kingdom. SMEI was launched into a Sun-synchronous orbit in January 2003, and is currently undergoing on-orbit test and validation. SMEI is a white light, all-sky camera designed to image coronal mass ejections (CMEs) as they propagate from the Sun through the interplanetary medium. Detection and tracking of Earthward-bound CMEs will dramatically improve DoD and NOAA space weather forecasts, perhaps providing days of warning prior to the arrival of damaging energetic particles in geospace. Such forecasts will help protect government and commercial space assets. SMEI data will also be a boon to international astronomers and astrophysicists seeking to understand solar processes, NEOs, and interplanetary dust. The University of California at San Diego provides a SMEI web page at http://cassfos02.ucsd.edu/solar/smei/index.html.

5. AFOSR AND THE NATIONAL SCIENCE FOUNDATION

Dating back to 1995, AFOSR and ONR have been partners with the NSF in the National Space Weather Program, and both DoD agencies continue to contribute funding to the NSF's annual NSWP research competitions. AFOSR and ONR are also both deeply involved with the activities of the NSWP's Committee for Space Weather alongside the NSF, and frequently serve on NSF advisory panels. Since February 2003, an AFOSR program manager has been in place at NSF, working part-time within the agency to help coordinate joint space weather and astronomy-related activities.

In 2002, the National Science Foundation awarded a new Science and Technology Center (STC) for space weather modeling at Boston University, called the Center for Integrated Space Weather Modeling (CISM). AFOSR helped make CISM possible by providing about US$1M in critical financial support for the research team when their first application for STC status was unsuccessful in 1999. AFOSR funding allowed the CISM team to remain productive and unified until their second STC application succeeded in 2002.

Typically, STCs are funded at the level of US$4M annually for 10 years, so AFOSR stands to gain a 40-to-1 "return on investment" in this case. With such funding stability over the next decade, CISM should become a primary focal point for space weather research activities in the United States. During this time, CISM will attempt to develop and continuously improve a comprehensive Sun-to-Earth space weather model. Meanwhile, the NSF also expects CISM to provide community vision and leadership, take prudent technical and management risks, exploit emerging science and engineering opportunities, and forge research collaborations wherever possible. The goal for CISM is to become a national center of excellence in space weather research. CISM will actively involve students (with special emphasis on women and minorities), research scientists and engineers from academic, industrial, and non-profit organizations, as well as various US federal laboratories, in synergistic partnerships for the good of society.

It is important to note that the primary mission of any STC is to perform an integrating leadership function for US research in their chosen field, with an emphasis on education and public outreach. CISM is thus not functionally nor administratively optimized to provide modeling services for the community in the style of the CCMC. To achieve national stature, CISM must ultimately have a positive impact on our society and economy by enabling the transfer of knowledge among academia, industry, and the United States' national laboratories. Sustained test and validation of

community models, while arguably a component of such a goal, is not a primary function that CISM can hope to achieve directly with its limited resources. It is expected that CISM will collaborate closely with the CCMC, taking advantage of the CCMC's strengths as an objective, unaligned, government-run research and modeling facility, while providing value-added partnership and leadership for the entire space physics community. The CISM home page is posted at http://www.bu.edu/cism/, while the NSF web site describing the mission and goals of the STC program can be found at http://www.nsf.gov/od/oia/programs/stc/start.htm.

6. AFOSR AND THE DOD MULTIDISCIPLINARY UNIVERSITY RESEARCH INITIATIVE

In the last few years, the DoD has sponsored three Multidisciplinary University Research Initiatives (MURIs) in space weather topics, each creating a multi-university consortium. MURI projects are funded for five years at levels approaching US$1M annually. The MURI program supports basic science and engineering research of critical importance to national defense. The program is focused on multidisciplinary research efforts that intersect more than one traditional science and engineering discipline. By supporting multidisciplinary teams, these MURIs are complementary to the ongoing DoD space weather programs at AFOSR and ONR that support university research through single-investigator awards.

Since 1999, the Office of Naval Research has administered the joint participation of a Utah State University team and a University of Southern California team in a Global Assimilative Ionospheric Model (GAIM) MURI. The Utah State GAIM team is developing a physics-based ionosphere-plasmasphere model as a basis for assimilating a diverse set of real-time (or near real-time) measurements, while the Southern California team concentrates on Kalman filtering and on perfecting other sophisticated mathematical techniques for data assimilation. GAIM's goal is to provide both ionospheric specifications and forecasts on a global, regional, or local grid. The GAIM team anticipates transferring model products to the NOAA RPC in Boulder, Colorado in the very near future. More information about the Utah State University GAIM model can be found at this web site: http://gaim.cass.usu.edu/GAIM/htdocs/present.htm.

The University of Michigan MURI, administered by AFOSR, began in 2001. This team of six universities uses advanced magnetohydrodynamic models and concentrates on the study of solar eruptive events (such as coronal mass ejections), their propagation through the interplanetary

medium, and their effects on the heliosphere and geospace. The long-term goal of this MURI team, headquartered at Michigan's Center for Space Environment Modeling (CSEM), is to progress toward a predictive "plug-and-play" space weather modeling capability that can eventually be transferred to NOAA and DoD space weather forecasting centers. CSEM is actively involved with the CCMC, having already transferred modeling products to the US Air Force Rapid Prototyping Center in Colorado Springs. Rice University, a member of the Michigan team, is also part of the CISM consortium at Boston University, so this MURI will participate in that ongoing NSF-funded STC effort. The CSEM home page is at http://csem.engin.umich.edu/.

The third MURI, also initiated in 2001 and administered by AFOSR, is focused at the University of California at Berkeley and headquartered at their Space Sciences Laboratory (SSL). This MURI team of nine universities is constructing a series of physically connected and observationally tested models of the Sun and its interplanetary environment. These models will allow the SSL consortium to use observations of the Sun's atmosphere and magnetic configuration to determine when a magnetic eruption is imminent, if that magnetic eruption will impact the Earth's space environment, and whether this will result in a geomagnetic storm and solar energetic particle bombardment at spacecraft orbits. By utilizing ground-based measurements from Mees Solar Observatory in Hawaii and Big Bear Solar Observatory in California (both part of this MURI team), as well as data from SSL-managed spacecraft, the Berkeley consortium is uniquely positioned to develop approaches for data assimilation of solar physics observations into predictive models. SSL is also an integral part of the CISM consortium and this MURI will clearly leverage that ongoing NSF-funded STC effort. The Berkeley MURI web site can be found at http://solarmuri.ssl.berkeley.edu/index.html.

The Michigan and Berkeley MURI programs also play a critical supporting role in the progress of the next-generation Advanced Technology Solar Telescope (ATST). The ATST, to be built within the decade for the US National Solar Observatory (NSO), will require significant advances in forecasting technology and observational techniques to be successful. AFOSR supports the development of the ATST through its direct collaboration with the NSF and through funding of related AFRL activities at the Sacramento Peak Solar Observatory in New Mexico, which is part of the NSO. In addition, the two AFOSR MURIs are producing very promising new results in solar physics, as well as modeling and observational capabilities directly relevant to the ATST project. The ATST web site can be found at http://atst.nso.edu.

7. CONCLUSIONS

AFOSR is willing to collaborate with all parties wishing to pursue the advance of space weather knowledge. The AFOSR mission is to ensure that such advances in our basic understanding effectively address DoD operational requirements. AFOSR is dedicated to working with the global research community to ensure effective participation by all interested scientists, as well as to enable new opportunities and discoveries.

Interagency and international collaboration is essential to achieve the goal of bridging the gap between space weather research and operations. AFOSR will help ensure that excellent scientific research is accomplished as models are developed, coupled, upgraded, and used by the space physics research and operational communities. Most important to AFOSR is that space weather research continues to benefit all participating agencies and stakeholders, while at the same time being responsive to the needs of national space weather customers and researchers.

8. ACKNOWLEDGEMENTS

The author wishes to thank the following individuals for their ongoing efforts to promote space weather research in the United States in partnership with AFOSR: Brigadier General David Johnson, Director of Weather, Headquarters US Air Force Pentagon; Brigadier General Pete Worden, Director of Transformation, US Air Force Space Command Space and Missile Systems Center; Dr. Robert McCoy, Office of Naval Research; Dr. Richard Behnke, National Science Foundation; Dr. Richard Fisher, NASA Office of Space Science; Dr. Ernest Hildner, NOAA Space Environment Center; Dr. Richard Vondrak, NASA Goddard Space Flight Center; Dr. Robert Robinson, National Science Foundation; Dr. Michael Hesse, NASA Goddard Space Flight Center; Dr. Greg Ginet, Air Force Research Laboratory Space Weather Center of Excellence; Lieutenant Colonel Frank Estis, Office of the Federal Coordinator for Meteorology; Lieutenant Colonel David Byers, European Office of Aerospace Research and Development, London; Lieutenant Colonel Mark Nowack, Asian Office of Aerospace Research and Development, Tokyo; Major Bill Olson, Directorate of Weather, Headquarters US Air Force Pentagon.

AUTHOR LIST

Daniel N. Baker
University of Colorado
1234 Innovation Drive
Boulder, CO 80309
USA
daniel.baker@lasp.colorado.edu

Janet L. Barth
Flight Data Systems &
Radiation Effects Branch
NASA/GSFC, Code 561
Greenbelt, MD 20771
USA
Janet.L.Barth@gsfc.nasa.gov

Jean Béland
Institut de recherche d'Hydro-Québec (IREQ)
1800 boul. Lionel-Boulet
Varennes, Qc,. J3X 1S1
Canada
beland.jean@ireq.ca

Paul J. Bellaire
Air Force Office of Scientific Research
801 North Randolph Street
Arlington, VA 22203-1977
USA
paul.bellaire@afosr.af.mil

Anatolii Belov
Institute of Terrestrial Magnetism and
Propagation of Radiowaves
IZMIRAN, Troitsk
142190 Moscow Region
Russian Federation
abelov@izmiran.troitsk.ru

Paal Brekke
European Space Agency
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
USA
pbrekke@esa.nascom.nasa.gov

Paul S. Cannon
QinetiQ
St. Andreus Road
Malvern, Worcs WR1439
United Kingdom
pcannon@qinetiq.com

Ioannis A. Daglis
Institute for Space Applications
National Observatory of Athens
Penteli
GR-15236 Athens
Greece
daglis@space.noa.gr

Eamonn J. Daly
Space Evironments and Effects Analysis Section
ESA/ESTEC
PO Box 299
NL-2200 AG Noordwijk
Netherlands
Eamonn.Daly@esa.int

Dale Gary
Center for Solar-Terrestrial Research
New Jersey Institute of Technology
Newark, NJ 07102-1982
USA
dgary@njit.edu

Bryn Jones
Flight Safety Department
Virgin Atlantic Airways
The Office, Manor Royal
Crawley, West Sussex RH10 2NU
United Kingdom
bryn.jones@fly.virgin.com

John G. Kappenman
Metatech Corporation
Applied Power Solutions Division
5 W. First Street, Suite 301
Duluth, MN 55802
USA
JKappenma@aol.com

Helen Mavromichalaki
Department of Physics
University of Athens
GR-15784 Zografos
Greece
emavromi@cc.uoa.gr

Alexander Nikiforov
Specialized Electronic Systems
115409 Moscow
Russian Federation
aynik@spels.ru

Mikhail I. Panasyuk
Scobeltsyn Institute of Nuclear Physics
Moscow State University
119899 Moscow
Russian Federation
panasyuk@sinp.msu.ru

Risto Pirjola
Finish Meteorological Institute
P.O. Box 503
FIN-00101 Helsinki
Finland
risto.pirjola@fmi.fi

Kevin John Small
Transpower NZ ltd
Box 1021
Wellington
New Zealand
Kevin.Small@transpower.co.nz

Dimitrios Vassiliadis
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
USA
vassi@electra.gsfc.nasa.gov

INDEX

L

M

N

P

R

S

T

W