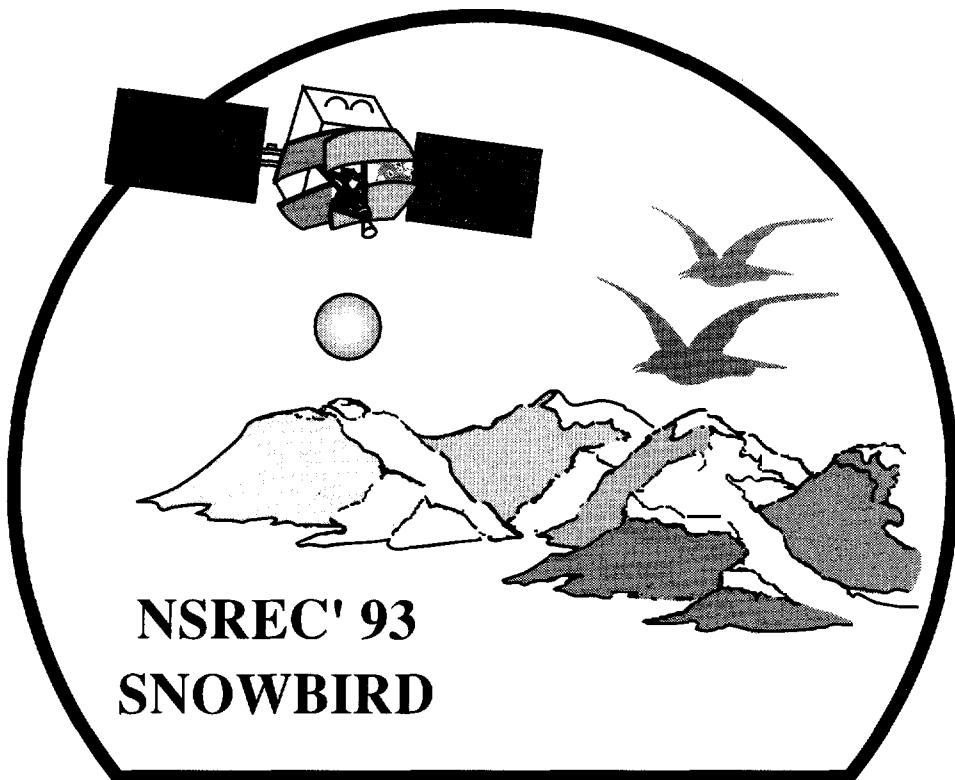


**IEEE Nuclear and Space Radiation Effects
Conference Short Course**

**Practical Considerations in Applying
Electronics to Space Systems**



**July 19, 1993
Snowbird Conference Center
Snowbird, Utah**



**Sponsored by
IEEE NPSS Radiation Effects Committee**

**Cosponsored by
Defense Nuclear Agency
Jet Propulsion Laboratory
Sandia National Laboratories**

**30th International
Nuclear and Space Radiation Effects Conference
SHORT COURSE**

**Practical Considerations in
Applying Electronics to
Space Systems**

**Snowbird, Utah
July 19, 1993**

Copyright © 1993 by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. Instructors are permitted to photocopy isolated articles for noncommercial classroom use without fee. For all other copying, reprint, or replication permission, write to Copyrights and Permissions Department, IEEE Publishing Services, 445 Hoes Lane, Piscataway, NJ 08855-1331.

TABLE OF CONTENTS

Foreword	ii
Biographies	iii
Chapter 1	
Geostationary Communications Satellites	I 1-57
G. D. Gordon, Consultant	
Chapter 2	
Radiation Environments within Satellites	II 1-160
H. B. Garrett, Ballistic Missile Defense Organization	
Chapter 3	
SEU Modeling and Prediction Techniques	III 1-93
L. W. Massengill, Vanderbilt University	
Chapter 4	
Novel Devices and Sensors	IV 1-60
J. C. Pickel, S-Cubed Division of Maxwell Laboratories	

FOREWARD

This is the fourteenth year that a Short Course has been offered on the first day of the Nuclear and Space Radiation Effects Conference. The short-course format provides an opportunity to cover topics in considerably more depth than is possible by individual contributed papers. It is an interesting challenge to select topics that have not been covered in detail in previous years, and also appeal to attendees with wide ranges of background and experience. The theme of this year's Course is "*Practical Considerations in Applying Electronics to Space Systems.*" It includes two topics on fundamental requirements for space systems, and two topics on new areas of research and development.

The first topic, *Geostationary Communications Satellites*, furnishes an in-depth look at the various functions of communications satellites, and relates these functions to specific requirements of electronics and sensors. It also discusses the internal thermal environment, and testing and reliability requirements.

The second topic, *Radiation Environments Within Satellites*, discusses the external and internal radiation environments in space systems. It addresses the environment of interplanetary space probes as well as that of terrestrial satellites.

The third topic, *SEU Modeling and Prediction Techniques*, shows how modeling at the device physics and circuit level can be used to increase the understanding of a number of single-particle effects. It also discusses ways to compare data with modeling predictions.

The fourth topic, *Novel Devices and Sensors*, examines a number of sensor technologies spanning wavelengths from 0.1 to 100 μm . Radiation effects on these devices are compared to more conventional electronic components. This important topic has not been addressed in previous short courses.

It has been a pleasure to work with the four presenters. I wish to thank them for their diligence and hard work, which often conflicted with busy work schedules. I also wish to thank Paul Dressendorfer, the 1993 General Conference Chairman, for giving me the opportunity to organize this year's course, and helping to select the topics. I also would like to thank Lew Cohn of DNA for his able assistance in obtaining classification review, and providing printing services. I hope this year's attendees will find the course interesting and useful, both as a precursor to the technical sessions and as a reference for later work.

Allan Johnston
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

BIOGRAPHIES

Allan Johnston

Allan H. Johnston received the B.S. and M.S. degrees in physics from the University of Washington. He joined the Boeing Aerospace Company in 1965, where he worked on a variety of research topics in radiation effects, including transient radiation and latchup studies in integrated circuits, total dose effects in MOS and bipolar devices, semiconductor device modeling, and radiation effects on optoelectronics. In 1992 he joined the Jet Propulsion Laboratory where he is involved in research studies of total dose effects in advanced integrated circuit technologies, single-particle latchup, and modeling of new single-particle effects, as well as supporting JPL space projects.

Mr. Johnston has authored over 40 publications on radiation effects and semiconductor devices. He has participated in a number of activities associated with the NSREC, including Short-Course Lecturer, Local Arrangements Chairman, Session Chairman, and Associate Guest Editor. He also served as Guest Editor for the HEART Conference, and is a member of the NPSS AdCom.

Gary Gordon

Gary D. Gordon is an Aerospace Consultant in Washington Grove, MD. He is co-author of Communications Satellite Handbook (Wiley, 1989) and of Principles of Communications Satellites (Wiley, 1993). He has taught courses on: Link Budgets, Spacecraft Technology, Satellite Orbits, Satellite Reliability, Spacecraft Thermal Design, and Computer Programming. He received his Ph.D. in physics from Harvard University in 1954.

Prior to his 9 years as a consultant, Dr. Gordon was 14 years at COMSAT Laboratories as Senior Staff Scientist in the Spacecraft Laboratory. Earlier, at GE's Astro Space Division, he contributed to the thermal design of the first weather satellite, TIROS, and one of the first active communications satellites, RELAY. In an education program at the corporate level, he taught managers about semiconductor devices, modern physics, computers, and satellites. He has written technical papers on electric propulsion, geodetic use of satellites, spacecraft thermal design, a proposed 30-kW solar array, and the effect of the moon's shadow on geostationary satellites.

Henry Garrett

Henry B. Garrett received his B.A. in physics (1974), and his M.S.(1973) and Ph.D.(1974) in Space Physics and Astronomy from Rice University. From 1974 to 1980 he served in the United States Air Force as a scientist at the Air Force Geophysics Laboratory, where he was awarded the AF Harold Brown Award, the AFSC Officer of the Year Award, and the AF R&D Award for his work on DCSC III and the AF/NASA SCATHA spacecraft in 1978. He joined the staff of the California Institute of Technology's Jet Propulsion Laboratory in 1980 where he has been the Lead Technologist for the Office of Engineering and Review since 1984. In this capacity, he serves as JPL's chief scientist for radiation environments and radiation effects. Dr. Garrett has been a consultant on space effects for programs ranging from Galileo, INTELSAT, and TDRSS, to Cassini. He is currently on a 2-year assignment from JPL to the DoD Strategic Defense Initiative Organization where he has been appointed SDI chief scientist for the Clementine program - a joint SDI/NASA mission to the moon and a near-earth asteroid.

Dr. Garrett has authored over 75 papers and one book on space environment effects. He served for 3 years as Associate Editor of the Journal of Spacecraft and Rockets. He is a Lt. Col. in the USAF Reserve and a member of the Sigma Xi, AIAA, the American Geophysical Union, the American Astronomical Society, and the American Physics Society.

Lloyd Massengill

Lloyd W. Massengill received the M.S. and Ph.D. degrees in solid-state circuits from North Carolina State University in 1984 and 1987, respectively. He is presently a member of the electrical engineering faculty of Vanderbilt University and is serving as the Microelectronics Group Leader. He has worked in the analysis of dose-rate and single-event radiation effects on integrated circuits, the development of computer-aided design tools, and microelectronics design for ten years, with over 35 technical publications in those fields. He is presently involved in the modeling of SOI devices and circuits, the study of radiation effects in commercial technologies for space applications, and the design of analog circuits for neural network hardware.

Dr. Massengill has served as a session chairman and reviewer for the Nuclear and Radiation Effects Conference, and as a reviewer of the IEEE Electron Device Letters, Trans. on Nuclear Science, Journal of Solid-State Circuits, Trans.on Circuits and Systems, and Trans. on Neural Networks. He is a member of the IEEE.

James Pickel

James C. Pickel received his B.S. degree in Engineering Physics and M.S. degree in Physics from the University of Tennessee in 1966 and 1968, respectively. He joined Rockwell International in 1968 and was a member of the Rockwell Science Center from 1974 to 1984, where he was Principal Investigator and had direct involvement in testing, analysis and modeling of radiation effects in microelectronics devices and infrared detectors. He performed pioneering work in single-event upset (SEU) of microelectronics and in radiation hardening of IR detectors and cryogenic electronics. Mr. Pickel has been a Principal Investigator at S-Cubed for basic mechanisms of radiation effects in IR sensor components since 1984.

He is a member of the IEEE where he has actively participated in the IEEE Nuclear and Space Radiation Effects Conference as a reviewer, Short Course Lecturer, Session Chairman, Awards Chairman, and Guest Editor. He has contributed regularly, as presenter, author, and reviewer, to the IRIS Detector Specialty Conference, the HEART Conference, and the Symposium on Single-Event Effects. He received the Outstanding Paper Award for the 1984 HEART Conference. He has more than 25 technical publications dealing mainly with radiation effects in microelectronics and IR detectors.

Chapter 1

Geostationary Communications Satellites

Gary Gordon
Consultant
Washington Grove, MD 20880

©Copyright IEEE 1993

GEOSTATIONARY COMMUNICATIONS SATELLITES

Configuration and Hardware

The objectives of this paper are:

- Explain how a communications satellite functions
- Describe the location of satellite components, especially those that are sensitive to particle radiation
- Give example of operational experience, especially problems that may be due to particle radiation

Satellite communications include earth stations, a satellite, and terrestrial circuits. The satellite is the space segment, and the other parts are the ground segment. Figure 1 shows a transmit earth station, an uplink, the satellite, a downlink, and a receive earth station. The earth stations and the satellite all operate as microwave relay stations, amplifying the signal and then transmitting it to the next point. Communications with geostationary satellites usually are for distances of around 40,000 km, far more than any terrestrial microwave relay station.

The function of a communications satellite is to receive a signal, amplify it, and retransmit it. This is done by the communications subsystem, which includes antennas, receivers, amplifiers, etc. Other subsystems must provide electrical power, maintain the pointing of antennas, control the temperature, send telemetry, receive commands, and include a structure for mounting all the equipment.

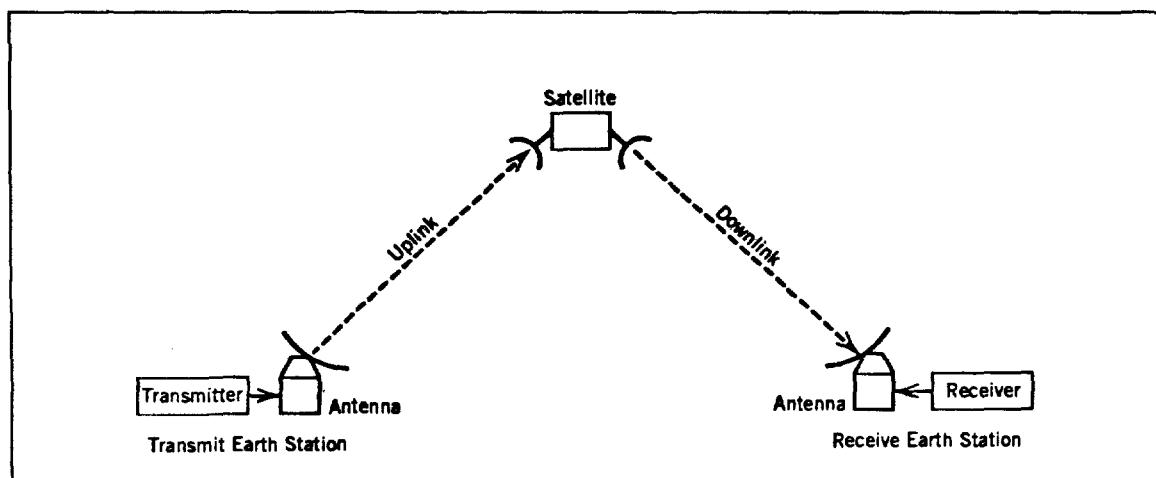


Figure 1 A communications satellite and its links to earth stations (Copyright, Gordon and Morgan, 1993. Reprinted by permission of John Wiley & Sons, Inc.).

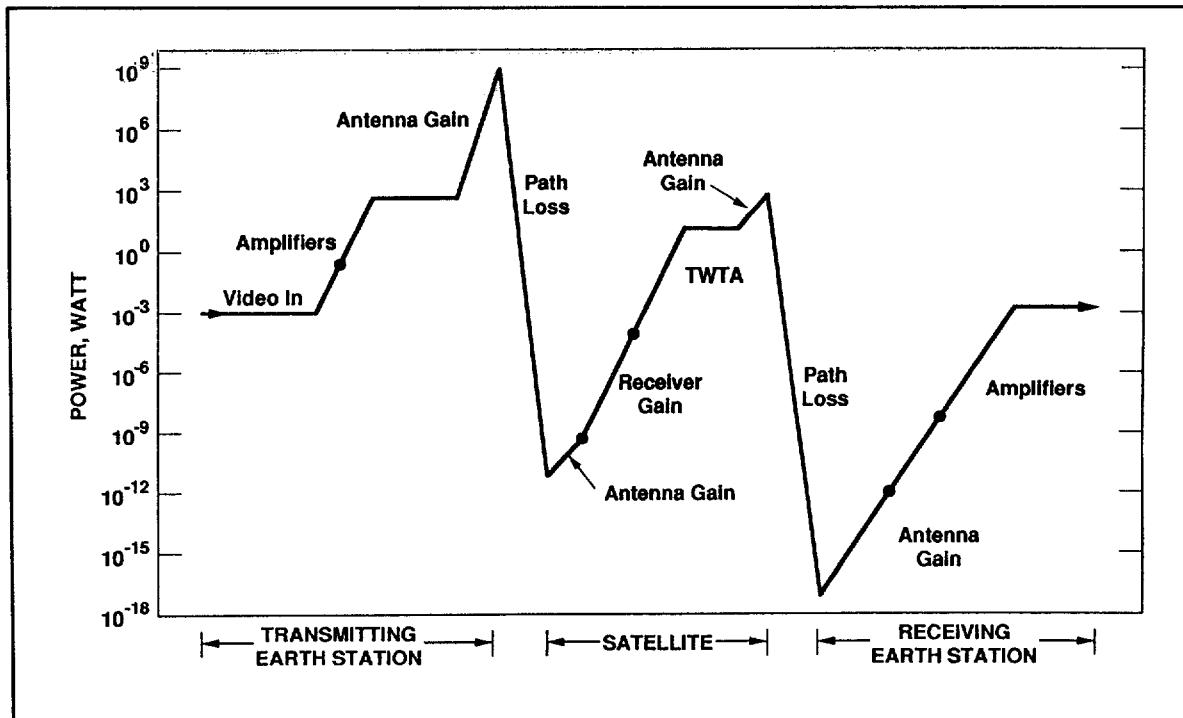


Figure 2 Power levels in satellite communications

The system and subsystem requirements of communications satellites are described in this report. This includes the location, function, and requirements of each satellite subsystem, with emphasis on the components that are most sensitive to space particle radiation effects. Included is the internal thermal environment and the factors that determine the temperature of internal components. This concludes with a discussion of space-craft testing and reliability, and their impact on requirements and specifications for electronic components.

This report concentrates on commercial communications satellites. Most of these satellites are in a special circular equatorial orbit called a geostationary orbit. They appear to hover 35,786 km over an assigned spot on the earth's equator. Thrusters on the satellite keep it over that spot within certain limits. When viewed by an earth station antenna, such satellites appear stationary in the sky. Therefore, many earth stations can use fixed-pointed antennas. Much of the discussion is applicable to satellites in other orbits or with different functions.

An example of the RF power through a communications satellite is shown in Figure 2. In this example the earth station transmits almost 1000 W, but only 10^{-9} W is received by the satellite. The satellite amplifies this signal to more than 10 W, but the earth station receives only 10^{-11} W. The graph shows the effect of several amplifiers, including the effect of antenna gain. The performance of a communications satellite is measured by the ratio of power to the noise present, calculated in a link budget.

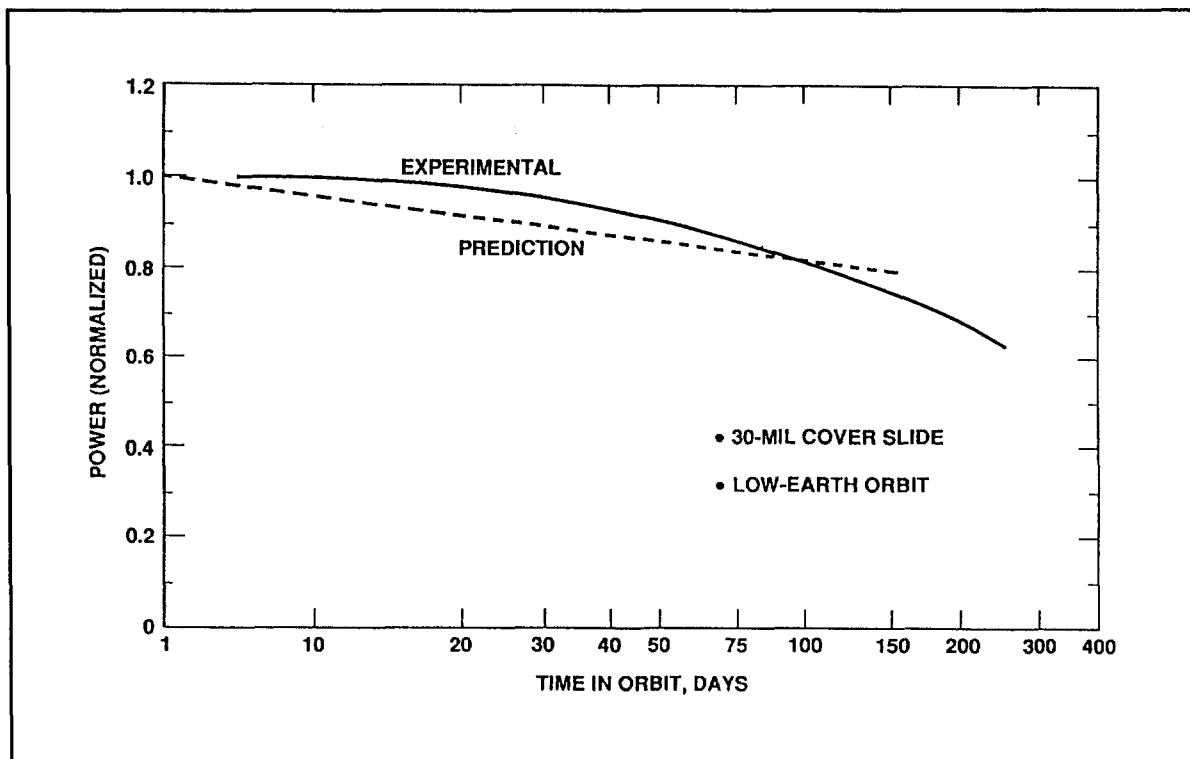


Figure 3 Radiation damage of silicon solar cells on RELAY communications satellite (1962).

Some effects of radiation on communications satellites were well known when the first ones were launched. The dotted line in Figure 3 shows the decrease in the DC solar array power that I predicted before the launch of the low orbiting RELAY satellite launched in 1962. The solid line shows actual measurements made after launch. In spite of a 30-mil cover slide, the power had dropped 40% in less than a year.

1. Spacecraft Configuration

Before discussing each of the subsystems of a communications satellite, a discussion on satellite configurations is necessary. There are many factors in choosing a configuration, not all technical. The selected configuration affects many spacecraft subsystems. Satellites of several configurations have been built, launched, and operated.

The two main configurations used in geostationary communications satellites are body stabilization and spin stabilization. Body-stabilized satellites have an orientation fixed with respect to the earth. They rotate once a day as the satellite goes around the earth. Body-stabilized satellites (see Figure 4), also called three-axis stabilized satellites, are simple in concept, but need hardware to control three axes. Body-stabilized satellites may have a momentum wheel that always rotates (thousands of rpm) in one

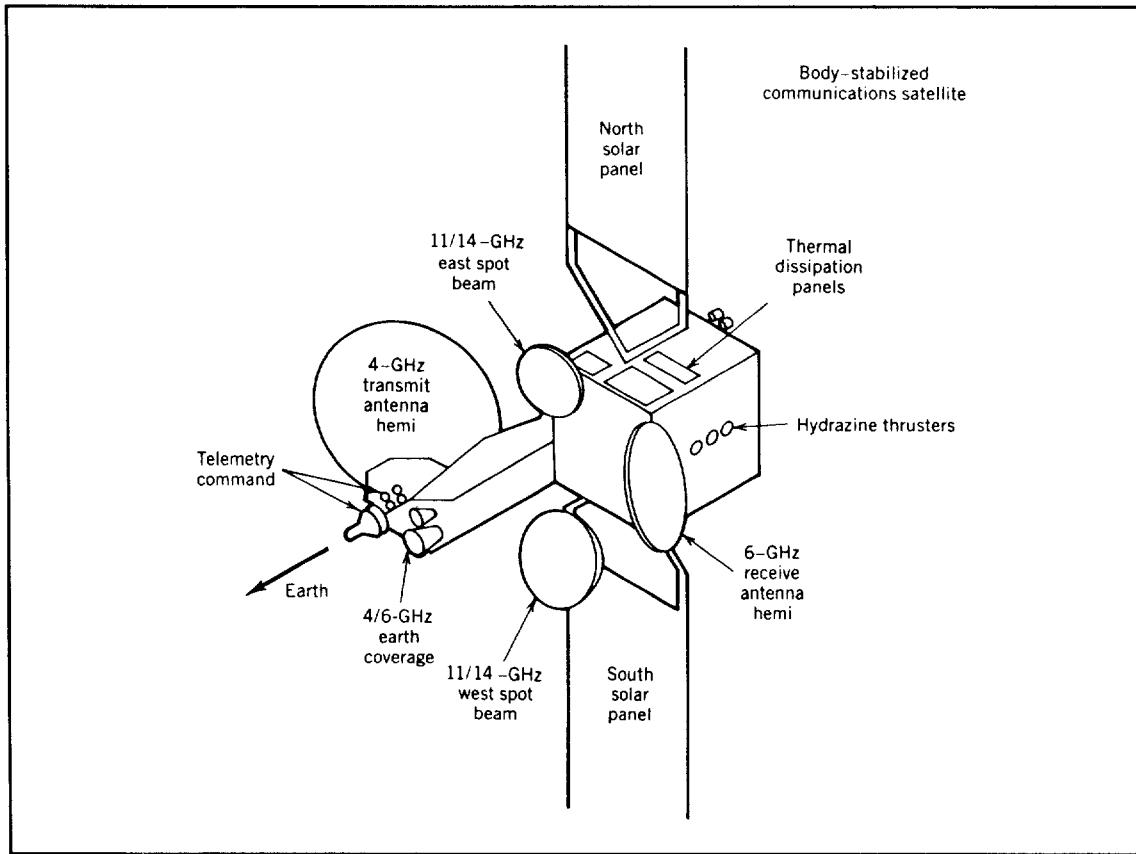


Figure 4 Body-stabilized communications satellite (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

direction, or reaction wheels that rotate either way.

A spin-stabilized satellite has a significant part of the satellite that rotates, on the order of one revolution a second. Small, simple communications satellites may be single bodies that rotate around the maximum moment of inertia axis. They may have an omni or toroidal antenna. Often the antennas and communications subsystem are despun, that is, do not rotate. Many larger communications satellites are "dual spinners" and rotate around a minimum moment of inertia (see Figure 5).

The location of satellite components affects their sensitivity to particle radiation. Solar arrays must be exposed to the sunlight. The power amplifiers dissipate the most heat; this is easier to do on a north or south face. Antennas must point at the earth. Thrusters must point north/south and east/west; they are frequently located on east/west faces, since there is more room on those faces. Batteries, fuel tanks, and the rest of the electronics are located anywhere in the interior, subject to constraints of center of mass and moments of inertia.

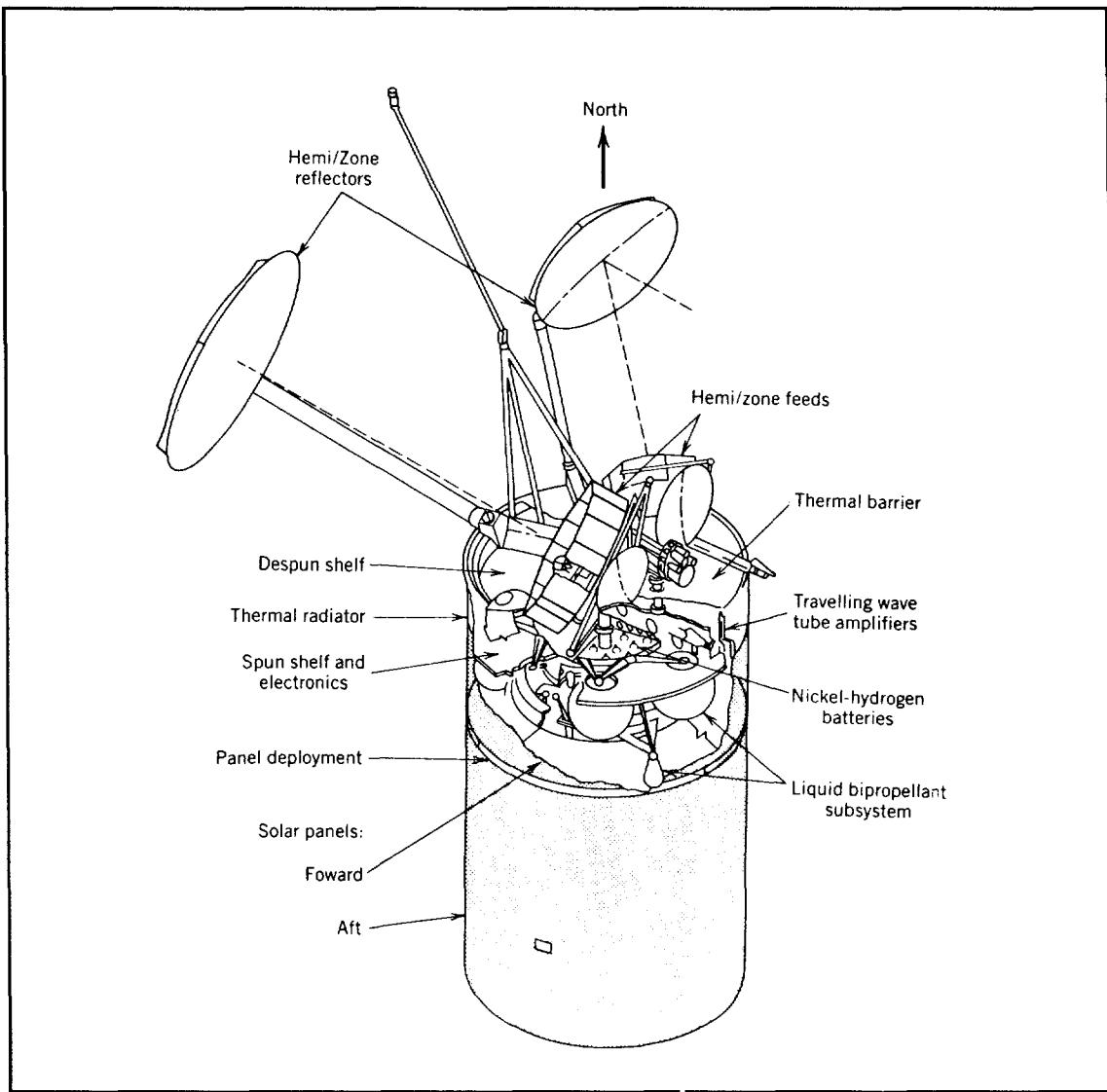


Figure 5 Spin-stabilized satellite (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

In a body-stabilized communications satellite, solar arrays are usually deployed panels, and rotate around a north-south axis. An antenna farm faces the earth, with parabolic reflectors for most communications, and horn antennas for some functions. The power amplifiers are mounted so heat can be easily dissipated, usually on the north- and south-facing panels. Thrusters are shown on the west face, and are also on the east face. They provide thrust along both east-west and north-south directions. Pairs of thrusters also control attitude.

A spin-stabilized satellite is cylindrical, with the solar cells mounted on the outside. The rotation axis is a north-south axis. Much of the mass is in the spinning part, which includes the

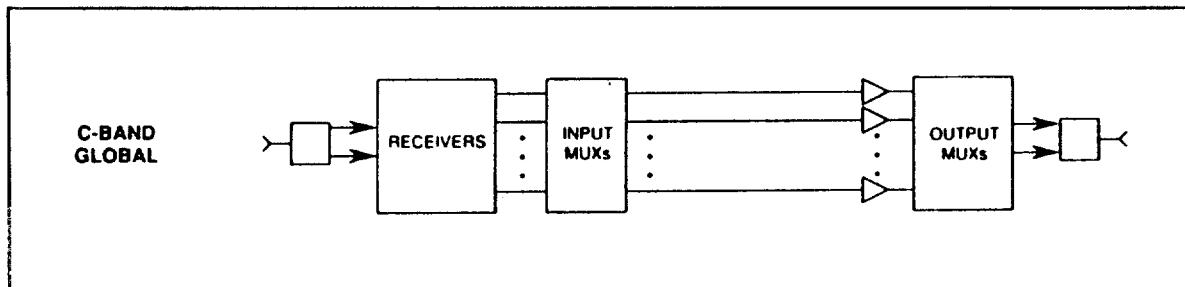


Figure 6 Main communications components for the global beam of the INTELSAT VI satellite (Gupta et al., 1991; reprinted by permission from COMSAT Tech. Rev.).

solar array, batteries, fuel tanks, and thrusters. The electric power and signal wires connect to the communications payload through slip rings. The antennas are mounted on a despun shelf on the north face, and point at the earth. All the communications payload is also on the despun shelf. The telemetry and command functions are divided, since both parts of the spacecraft require these functions.

2. Communications Subsystem

The function of a communications satellite is to receive radio frequency signals from the ground, amplify, and transmit them. The main parts of a simple communications subsystem are shown in Figure 6. The basic parts are a receive antenna, amplifiers, and a transmit antenna. Amplification is usually in two stages: a receiver that includes a low noise amplifier (LNA) and a frequency converter, and a power amplifier that provides the maximum power required.

The signal is received by an antenna, consisting of a parabolic reflector and a feed. The receiver amplifies the received signal (about a nanowatt) to the order of a milliwatt, and is designed to add a minimum amount of noise. The receiver also includes a down converter, which changes the frequency from 6 GHz to 4 GHz (for the usual C-band frequencies).

The power amplifier increases the signal power by many orders of magnitude. Usually the last amplification is done by several transponders, each amplifying a given band of frequencies. Each transponder provides of the order of 10 watts of RF output. An input multiplexer (MUX) distributes the receiver signal to the various transponders, and an output multiplexer combines all the signals for transmission. The transmit antenna is similar to the receive antenna, and consists of another parabolic reflector and feed.

Communications satellites can be more complex. Another part of the INTELSAT VI communications subsystem is shown in Figure 7. This satellite uses both C-band (6/4 GHz) and Ku-band (14/11 GHz).

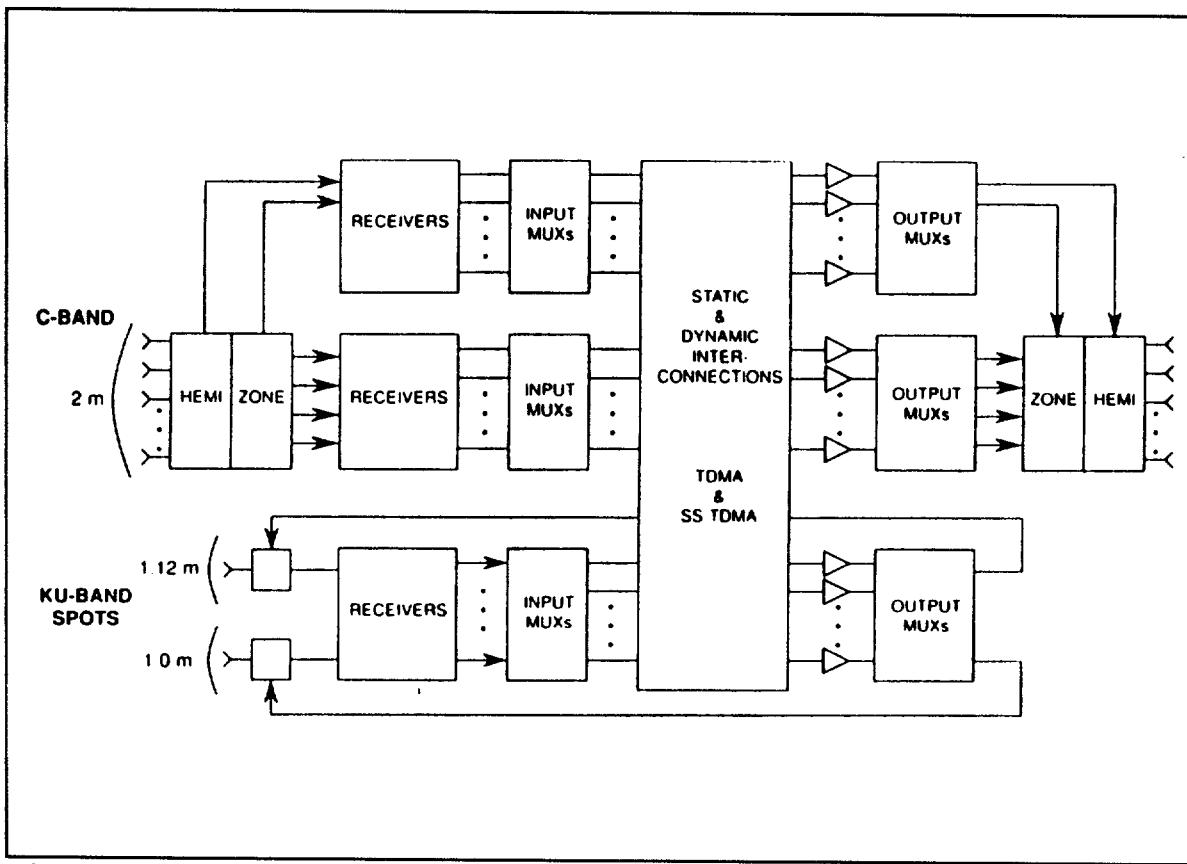


Figure 7 Interconnected C-band and Ku-band components of the INTELSAT VI communications subsystem (Gupta et al., 1991; reprinted by permission from Comsat Tech. Rev.)

The antenna feeds can generate several beams so the same frequencies are used more than once. And the signals inside the satellite are switched by static and dynamic interconnections. This provides flexibility so signals on each input beam can be switched to any output beam. This satellite can be used for time division multiple access (TDMA), and also for satellite switched TDMA (SS/TDMA).

2.1 Receiver

The uplink signal enters the satellite receive antenna and is amplified by the satellite receiver. Since the received RF signal is usually at a very low power level, the first amplifier stages must amplify the signal without introducing much additional noise. These initial stages are often called a low-noise amplifier (LNA).

The receiver also shifts the frequency of the received signal. In a satellite it must be shifted to a different frequency before it is transmitted. It is easier to shift the frequency of the signal at low power levels. After the initial amplification stages, the frequency is changed, followed by additional amplification.

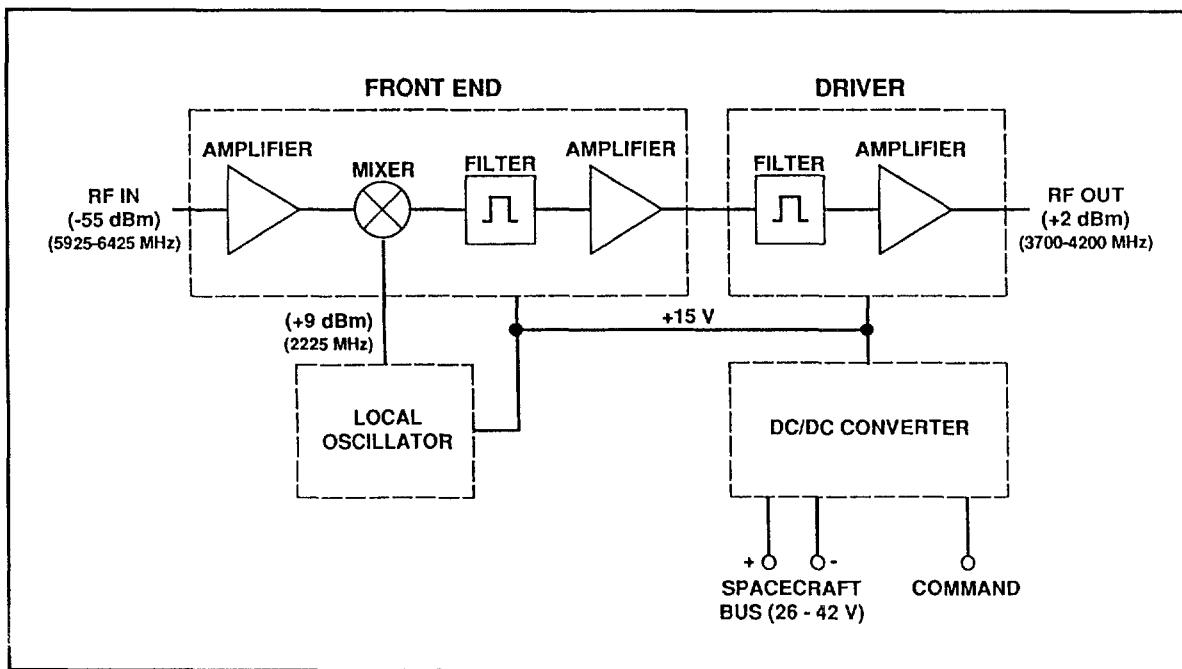


Figure 8 Block diagram of 6-GHz receiver.

The major parts of a 6-GHz satellite receiver are shown in Figure 8. The downconverter consists of a local oscillator source (in this case 2225 MHz) and a mixer. There are amplification stages before and after the frequency shift. The signal is amplified in several stages, which may consist of gallium arsenide (GaAs) field effect transistors (FETs). Each stage typically has about 10-dB gain. The figure shows an amplification from about 3 nW (-55 dBm) to 1.5 mW (2 dBm). Satellite receivers are broadband devices, and can operate with a 500-MHz bandwidth.

The first amplifier consists of several stages. Attenuators are often added to control the total receiver gain. These may be adjusted during initial tests, may provide temperature compensation, or may be commanded from the ground.

To change the frequency, the input RF signal is mixed with the frequency from a local oscillator. This generates both the sum and the difference in the two frequencies. A filter is used to select the desired frequency, and suppress the unwanted ones, including the original input frequency. For example, a satellite C-band receiver uses a local oscillator frequency of 2225 MHz. Signals in the band from 5.925 to 6.425 GHz are changed to a band from 3.700 to 4.200 GHz.

The local oscillator consists of a temperature-controlled crystal oscillator operating at a lower frequency, with multipliers and filters to generate the desired signal at 2225 MHz. For example, the desired frequency can be generated by starting at 23.177083 MHz, and multiplying successively by factors of 6, 4, and

4. Bandpass filters are used to get a pure frequency, without harmonics or spurious signals. After the mixer, a notch filter is used to reduce the level of the second harmonic (4.45 GHz) of the local oscillator. This harmonic is generated by the mixing process.

Other receivers work with different frequencies. A Ku-band satellite receiver shifts from the uplink frequency of 14 to 14.5 GHz to the downlink frequency of 11.7 to 12.2 GHz, so the local oscillator frequency is 2.3 GHz. Satellites that operate in both frequency bands may downconvert all frequencies to the 3.7- to 4.2-GHz band, perform all the necessary switching, and then upconvert some signals back to the 11.7- to 12.2-GHz band.

2.2 Power Amplifier

The function of the power amplifier is to increase the power of the signal for transmission into space. It is the last link in a series of amplification stages. Satellite transmitters may be of various bandwidths (such as 27, 36, 43, 54, 72, or 241 MHz). The bandwidth depends on the access method, service provided, and most important, mass and power requirements.

The output amplifier may be a solid-state power amplifier (SSPA) or a traveling wave tube amplifier (TWTA). Both operate at microwave frequencies, and are used in satellite communications. They differ in the maximum RF power delivered, as shown in Figure 9. The SSPA delivers less power and the TWT more power.

Efficiencies of TWTS are in the range of 30% to 50%, while SSPAs are more limited and are around 30%. The inefficiency of these amplifiers leads to requirements to dissipate the heat. This is a challenge for all power amplifiers, especially those operating in a satellite.

The traveling wave tube (TWT) was once the dominant power amplifier for communications satellites. It is still widely used. It delivers significant amounts of power (around 50 W), is efficient (around 50%), and operates over a wide bandwidth (40 to 80 MHz). In comparison to an SSPA, it is heavy and has a limited life. Solid-state devices are inher-

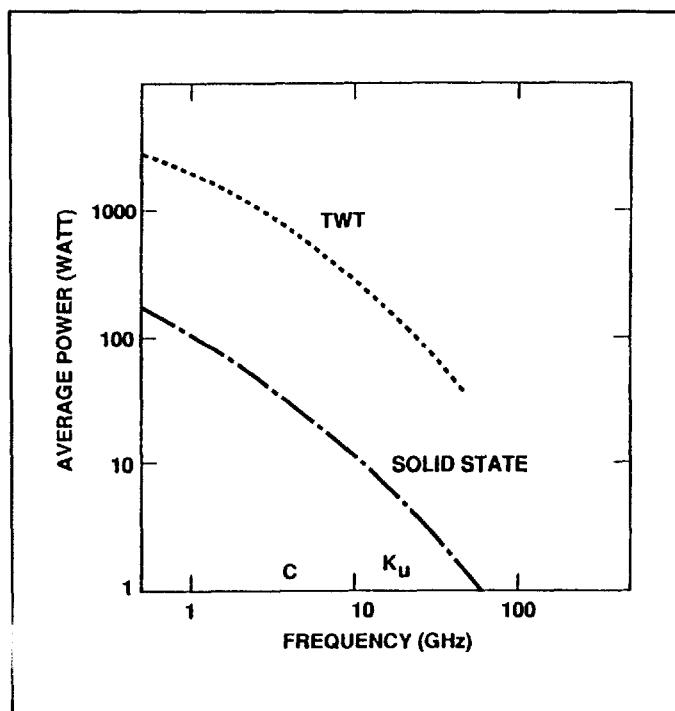


Figure 9 Maximum power for satellite power amplifiers.

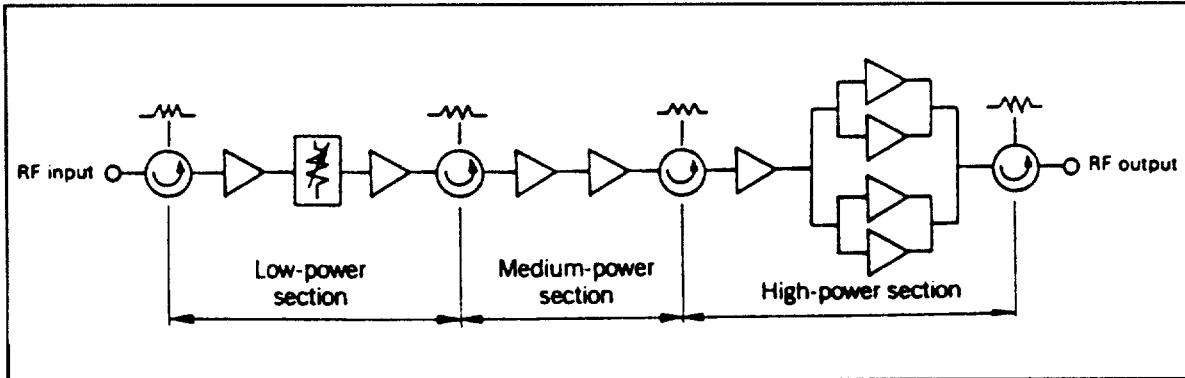


Figure 10 Typical solid-state power amplifier, SSPA (Copyright, Gordon and Morgan, 1993. Reprinted by permission of John Wiley & Sons, Inc.).

ently small, and this makes it difficult to dissipate the heat generated in power amplifiers. In applications where the heat can be dissipated, the SSPA provides low mass and long lifetime. SSPAs are less efficient, and require more DC power to produce the same amount of RF output.

Solid-State Power Amplifiers (SSPA). The SSPA consists of several cascaded amplifier stages. An example is shown in Figure 10. A typical stage is a gallium arsenide field effect transistor (GaAs FET). Each stage has a gain around 10 dB, so the total gain with six stages can be about 60 dB. The maximum power is generated in the last stage by four matched final power amplifiers summed by a power combiner. Amplifiers are separated by microwave circulators to isolate the earlier low-level amplifier from the higher power that follows. A variable attenuator is used to adjust the gain and provide temperature compensation.

Reliability and long service life are important requirements for satellite equipment. These are inherent in solid-state devices, but satellite requirements impose additional steps in design, selection, qualification, and testing. Usually there is not enough time to test for the desired service life, which may be 10 yr or more. Accelerated testing is possible at higher temperatures. For example, an SSPA that is designed for an operating channel temperature of 90°C may be tested at elevated temperatures of 190° and 215°C. The test results can be extrapolated for the operating temperature. Failure rates for SSPAs may be 500 failures in 10^{-9} hours (FITS), compared to 1500 to 2000 FITS for comparable TWTs.

Traveling Wave Tube (TWT). The TWT is a specially designed vacuum tube device. It makes special use of a long stream of electrons. The operating principle of a TWT may be understood by referring to Figure 11. An electron gun, consisting of a heated cathode and focusing electrodes, produces a stream of electrons. The electrons travel along the axis of the tube, constrained by

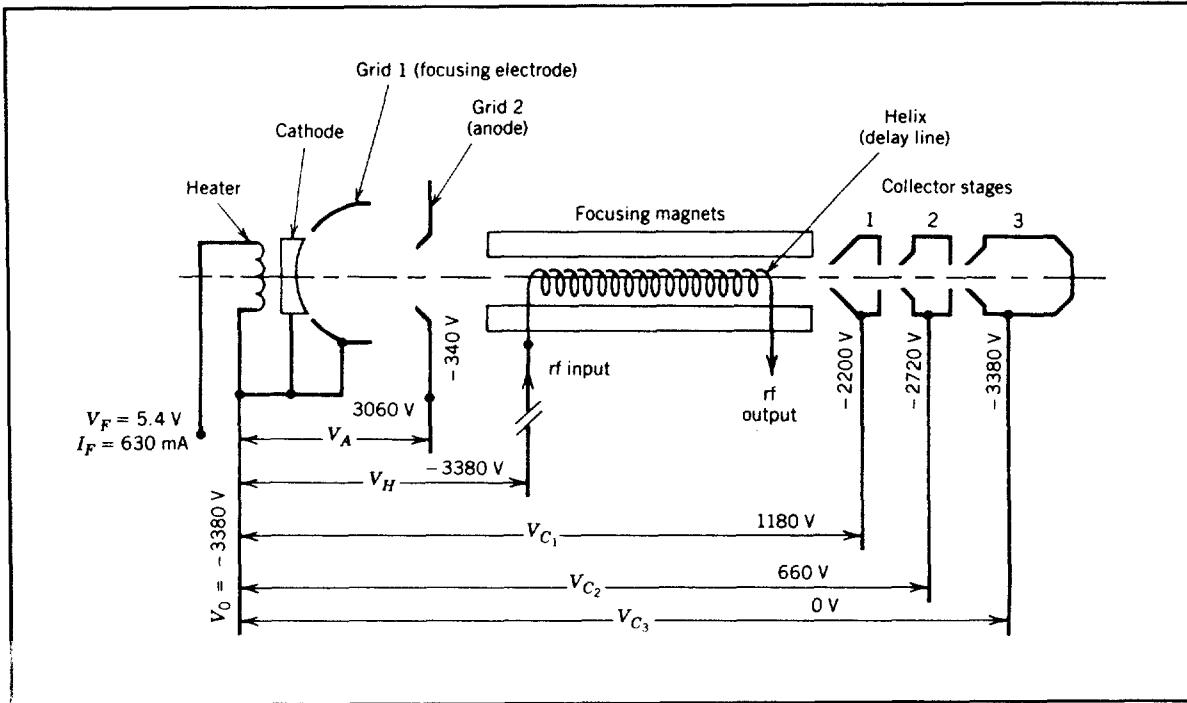


Figure 11 Main components of a traveling wave tube (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

focusing magnets, until they reach the collector. Spaced closely around the beam is a helix, capable of propagating a slow electromagnetic wave. The input RF signal enters one end of the helix, and the RF output comes out the other end.

The RF signal travels along the helix wire at velocities close to the speed of light. The signal must travel around each helix coil, so the phase velocity of the wave along the tube axis is small compared to the velocity of light. In typical low-power tubes this may be about one-tenth of the velocity of light. For higher power tubes it is two or three times higher. The high voltage accelerates the electrons to a velocity similar to the axial phase velocity of the RF wave.

When the RF signal enters the helix, the longitudinal part of its field interacts with the electrons traveling down the tube. Some electrons are accelerated, and some decelerated. The result is a progressive rearrangement in phase of the electrons with respect to the wave. In turn, the modulated electron stream induces additional waves on the helix. This process of mutual interaction continues along the length of the tube. The net result is that the DC energy of the electron stream is transformed to RF energy in the helix, and the wave is amplified.

Each traveling wave tube has a DC/DC converter to provide the different voltages required (see Figure 12). When the tube is

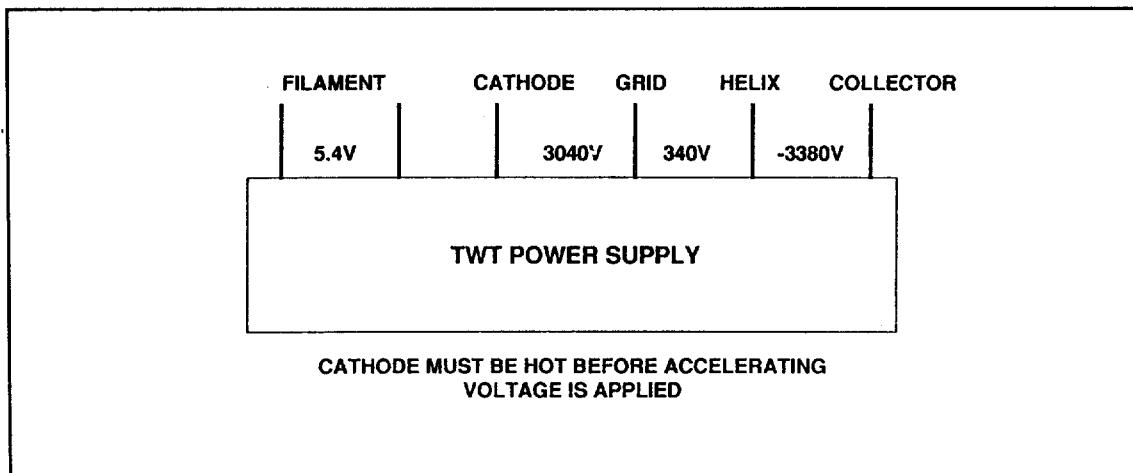


Figure 12 Voltages supplied to a traveling wave tube. Sometimes switches off spontaneously.

turned on, the high voltage must be delayed a minute or two until the filament has heated the cathode. On some satellites a transponder may shut down for no apparent reason. This may not happen very often (like once a year), but still represents a serious communications outage. Usually it can be turned on again by ground command. This spontaneous switch-off may be due to radiation effects, or could be some other unknown cause.

2.3 Sharing Satellite Capacity

Most communications satellites handle many signals. The satellite resources may be shared in different frequencies, at different times, by different antenna beams, or other methods. One of these sharing methods, known as SS/TDMA, needs considerable computer memory, and may be sensitive to radiation effects.

Frequency-Division Multiplexing (FDM). At an earth station, multiplexing may start by combining several telephone channels, each in a different frequency band. Or the multiplexing may have been done before the signals reached the earth station. Other earth stations may use the same transponder in the same satellite, but transmit a shifted frequency band that doesn't overlap with the first earth station.

In most communications satellites there are multiplexers that operated with large bandwidths. A wideband (for example, 500 MHz) receiver is connected to an input multiplexer, which separates the 500 MHz into various channels. A common configuration uses 12 filters, each with 36 MHz of usable bandwidth and 4 MHz of guard bands between filters. After amplification by separate transponders, the 12 outputs are combined with an output multiplexer.

Antenna Beams. Satellite antenna patterns continue to evolve toward more customized shapes to cover certain countries or areas,

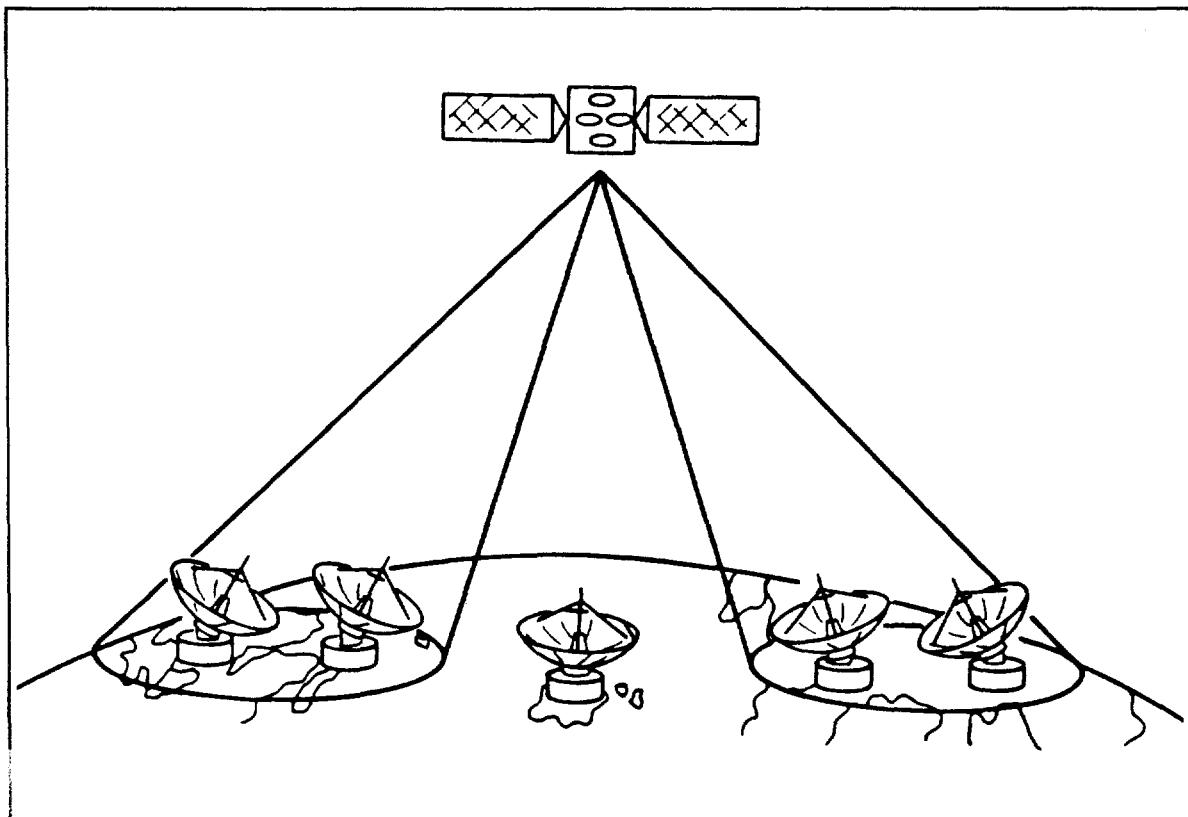


Figure 13 Frequency reuse by using separated beams (Copyright, Gordon and Morgan, 1993. Reprinted by permission of John Wiley & Sons, Inc.).

while reducing the power in other areas. A satellite designed for the fixed satellite service may not want to waste power on the oceans. A domestic satellite may have one beam on its homeland and other beams on isolated territories.

If these beams do not overlap and the sidelobe energy from one beam into another is low, they become candidates for frequency reuse. Figure 13 shows an example. Signals are received from and sent to only the two areas shown. The earth station in the center is either not covered or must use another frequency that has a beam covering the island. Beam selections may be done on a transponder by transponder basis.

One set of beams used by INTELSAT satellites are zone beams. In the Atlantic Ocean Region there is one beam focused on North America, a second on South America, a third on Africa and a fourth on Europe. Figure 14 shows the beams for a satellite at 332.5°E longitude (27.5°W). Note how widely the beams are separated. Thus, the same frequency assignments may be used four times.

Time Division Multiple Access (TDMA). Instead of using the frequency domain of a transponder, it is possible to time-share the entire bandwidth. As viewed at the satellite's power amplifier,

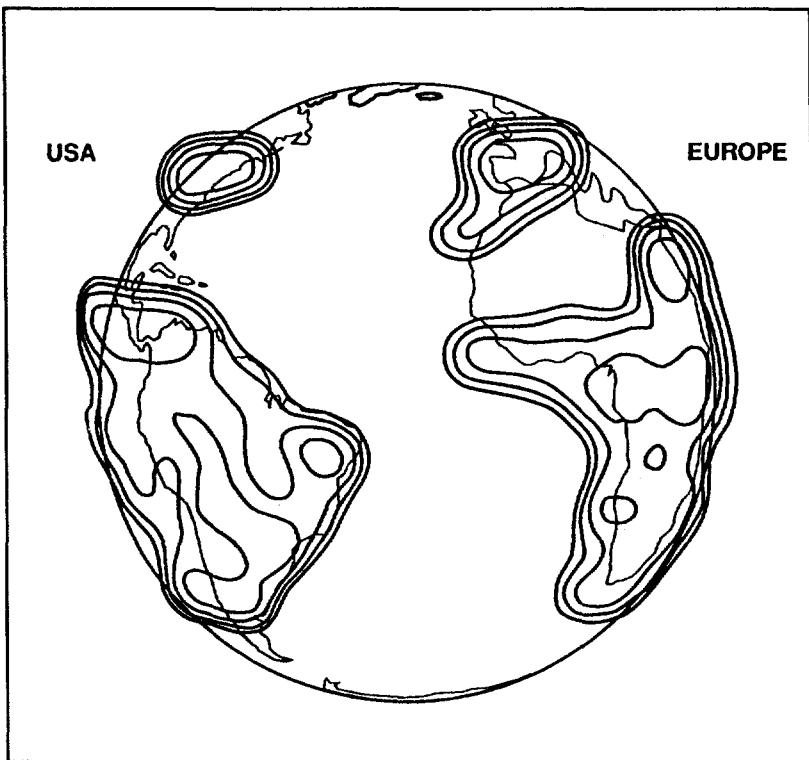


Figure 14 Zone beams for a geostationary satellite over the Atlantic Ocean.

assigned interval. TDMA systems are flexible, and accommodate traffic variations by increasing or decreasing the duration of each station's bursts to match the traffic flow. As soon as one station's burst is completed, another station gets a time slot.

TDMA is inherently compatible with digital signals, since it uses the maximum power of a transponder and is bandwidth efficient. It also results in more complex (and often more costly) earth stations than FDMA, because of the need to maintain precise network timing. Integrated circuits have reduced the cost of these stations. Network synchronization requires a master clock, either on the satellite or at a master earth station. Satellite motion changes the time delay for different stations, and this is included in the synchronization plan.

Switching. For a simple TDMA system there is no need for additional satellite equipment. Such a system requires only one input beam, one transponder, and one output beam. For a multi-beam satellite, this may be a severe constraint. For a TDMA station to transmit signals in different downlink beams, switching in the satellite is required. Such a system is known as satellite-switched TDMA (SS/TDMA).

Dynamic switching is at the heart of SS/TDMA. Each uplink beam can be connected to a specified downlink beam, at a certain

there is only one signal present at any time. Each earth station in the network assigned to the transponder takes its turn using the transponder. A prearranged schedule avoids overlapping transmissions. Guard times are established between each earth station's bursts.

Time-division multiple access provides access to the available transponder spectrum on a time-shared orthogonality basis. In a simple example, one carrier fully occupies the transponder at a time. Each station transmits a burst of digitized voice, video, or data in its exclusively

time, for a certain duration. Suppose the frame duration is 2 ms, and the frame unit is 1 μ s. The same combinations occur every 2 ms. A station in spot beam 1 may transmit a burst to beam 5 between 23 and 28 μ s after frame synchronization. The return transmission does not have to be simultaneous, and may not be the same length of time. A certain pattern is set up, and repeats every frame duration. The pattern can be changed by sending and storing a different switching pattern.

Dynamic switches provide rapid switching of microwave circuits. Figure 15 shows a 6×6 microwave switch matrix (MSM) or crossbar switch. This switch provides dynamic connectivity for six beams, two hemispheric and four zone beams. To switch 6 channels, a 6×6 grid is needed, with a solid-state switch at each intersection.

The diode switches marked with a heavy dot are "on." Only one switch is on for any one column and for one active row. In Figure 15, input channel 1 is connected to output channel 1, and input channel 2 is connected to output channel 4. Any of the six input channels can be connected to any of the six output channels, in any desired combination. In this example there are 36 switches at the intersections. Each switch includes two PIN diodes that have a low impedance in the ON state, and a high impedance in the OFF state. The switches are individually controlled by a driver circuit and a timing source.

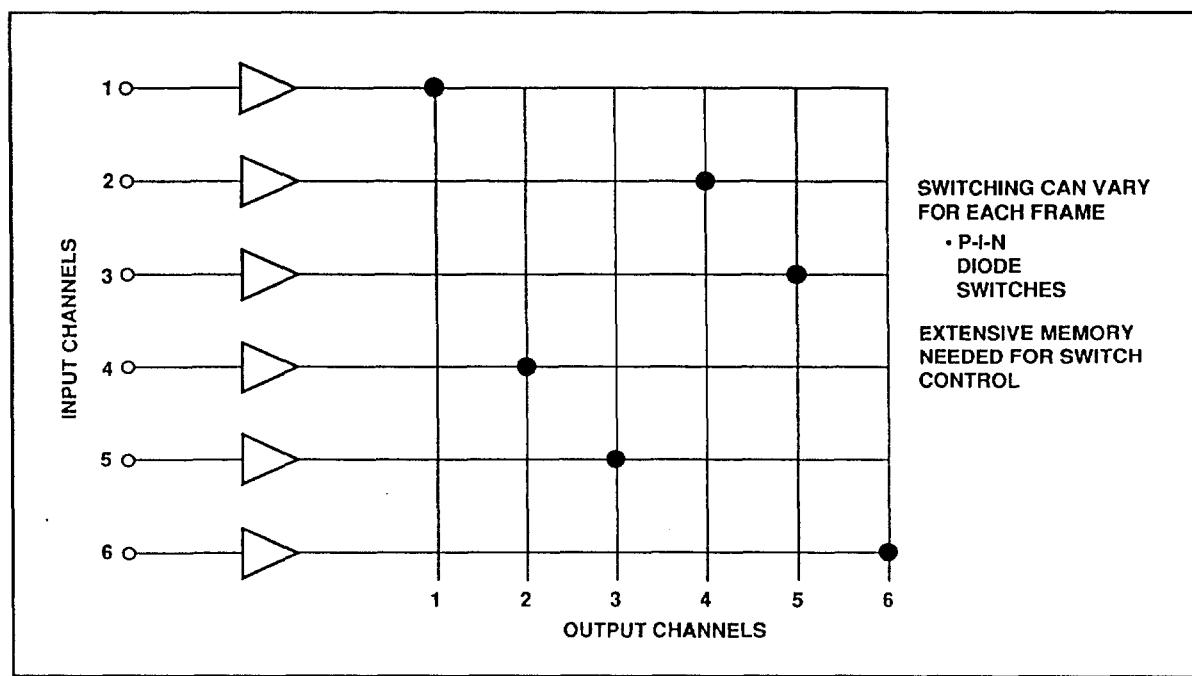


Figure 15 Microwave switch matrix for a satellite-switched time division multiple access (SS/TDMA).

Each pattern on a 6 by 6 switch matrix requires the setting of 36 diodes. If a frame has several hundred patterns used in rapid succession, then there may be more than 10,000 diode settings to remember. This requires a memory to remember all these settings, and a microprocessor to set each diode to its proper position at the proper time. Careful design is needed to reduce the effects of radiation damage, and redundant memory is one way.

3. Telemetry and Command

The state and health of the various subsystems are transmitted by the telemetry system. A command system transmits the desired commands to the satellite. A signal sent through the command link and returned by the telemetry transmitter is used for ranging. A function diagram of the TT&C subsystem is shown in Figure 16.

3.1 Command

A command starts at the satellite control center, goes to an earth station, and is then transmitted to the satellite. The command receiver in the spacecraft receives the signal, demodulates it, and processes it. The command is decoded, stored, and verified. Commands are critical, and verification is used to make sure that incorrect commands are not executed. The signal is sent to the proper location in the spacecraft only after an execute signal is received from the satellite control center.

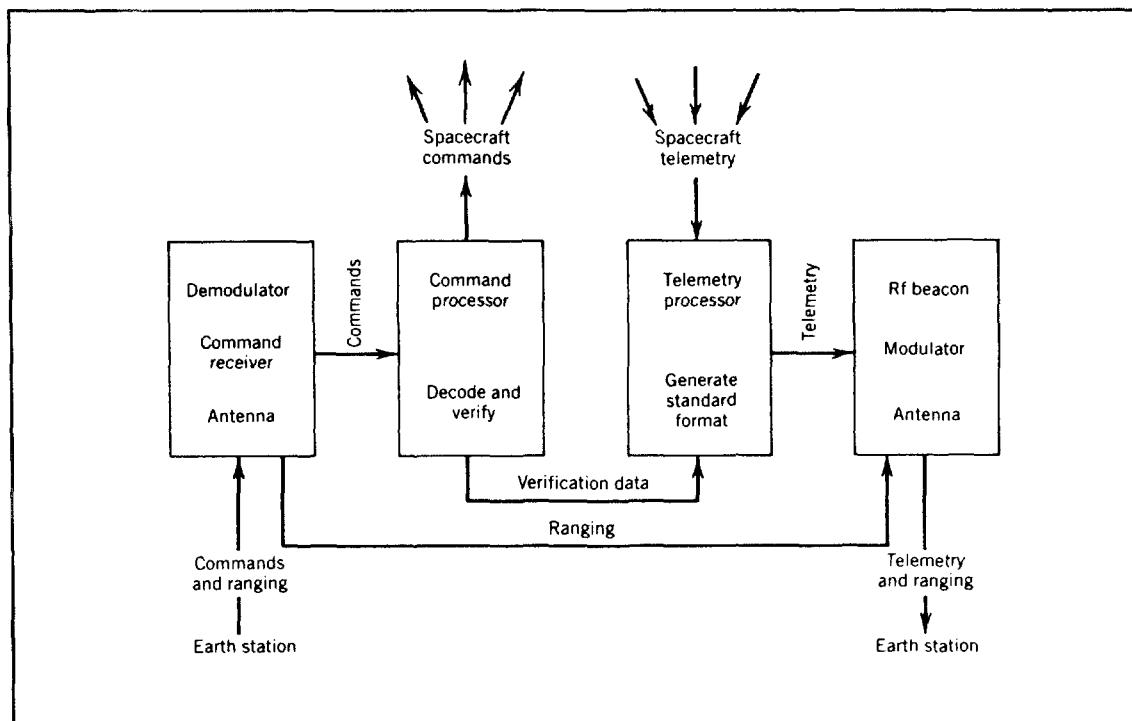


Figure 16 Telemetry, command, and ranging (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

Communications
Receiver, transponders -- on/off
Switch matrices -- switch status
Electric Power
Array drive -- normal, fast fwd, reverse
Battery, bus -- charge, recondition, heater
Telemetry and Command
Beacon, encoder -- on/off
Command -- execute
Propulsion
Valves -- open/shut
Thrusters -- on/off
Fuel lines -- heater
Orientation
Sensors, wheels -- on/off, mode

Table 1 Typical items on a list of satellite commands.

Commands from the ground control the satellite's on-board equipment status, operational modes, position, and orientation. Discrete commands use signals that are fixed-duration pulses to turn on or off various components or to change operational modes. Analog signals are used to generate variable-length pulses which turn equipment on for specific intervals. Digital data blocks can be transmitted to a satellite for reprogramming on-board processors.

Table 1 is a typical command list. Each satellite will have a different list depending on its own construction. The communications subsystem has receivers, transponders, and other switches to turn on and off. The electric power subsystem has commands to charge or discharge each battery. The telemetry and command system has redundant beacons and encoders to turn on or off. The propulsion system has valves to open and close, and thrusters to fire. The attitude control system has various sensors and different operational modes. Most satellites have electric heaters that may turn on automatically, or may be turned on by ground command.

3.2 Telemetry

The telemetry is on the right side of Figure 16. It gathers data from various subsystems, processes them into a desired form, modulates the RF beacon signal, and transmits it to the ground. The telemetry collects, formats, and transmits data of various types. Digital data show the condition of on-off equipment or the position of various switches. Many sensors check the performance

Communications
Receiver, transmitter -- switch status, T, I
Antenna -- T, actuator status
Electric Power
Solar array -- T, V, I
Array drive -- Motor-I, position, deployment, T
Battery, bus -- V, I, T, heater & switch status
Telemetry and Command
T, on/off status
Propulsion
Tanks -- pressure, T, heater status
Valves & lines -- T, valve and heater status
Thrusters -- T, heater status, safe/arm (AKM)
Orientation
Sensors -- T, output
Electronics -- error, torque, modes
Wheels -- speed, error, T, I

Table 2 Typical items on a telemetry list.

of on-board systems. They measure temperature **T**, voltage **V**, current **I**, or pressure. Signals from attitude sensors and accelerometers are transmitted for interpretation on the ground. Some scientific and experimental data may be gathered to aid the design of future satellites. This could be radiation, magnetic field, electrical discharges, motion of liquid fuel, etc.

Table 2 lists possible telemetry items on a communications satellite. The list is different for each satellite. The major goal is to provide information for both normal and abnormal operation. For each item the benefits of the information is compared to the weight and cost of including the monitoring point in the telemetry list. A few parameters are difficult to measure; some are curves with many points and a considerable amount of information to be transmitted.

There is much redundancy in the telemetry. As long as most of the telemetry is working most of the time, occasional errors can be tolerated or corrected. In most systems no action is taken due to a single telemetry point, but only after it's confirmed by additional telemetry. Telemetry is often monitored by a computer, and significant changes trigger an alarm in the satellite control center. In some systems the status of various switches is monitored, and a spontaneous change will trigger an automatic command to change it back to its previous value.

3.3 Ranging

Ranging is used to determine the satellite position in space. This data is used to predict future orbital motion. Thrusters are fired by command to control the position of the satellite. This is called stationkeeping. Ranging measures the distance (slant range) from an earth station to the satellite, as a function of time. The distance is measured by timing a signal from the earth station to the satellite and back to the ground. Position determination is also done by measuring the azimuth and elevation angle from the earth station to the satellite. This only requires a RF beacon in the satellite.

4. Electric Power

The electric power subsystem furnishes power to operate the communications payload, especially the power amplifiers. The latter often take three quarters of the total power, and even more in a direct broadcast satellite.

Silicon solar cells provide the primary power of a communications satellite. The power P_s generated by a solar array is

$$P_s = \eta S_c A \quad (W) \quad (1)$$

where η is the solar cell efficiency, S_c is the incident solar power (average value = 1370 W/m^2), and A is the projected area of the solar array.

Occasionally the earth or moon block the sun. Then rechargeable nickel-hydrogen (NiH) batteries provide secondary power. A typical power subsystem is shown in Figure 17. A body-stabilized

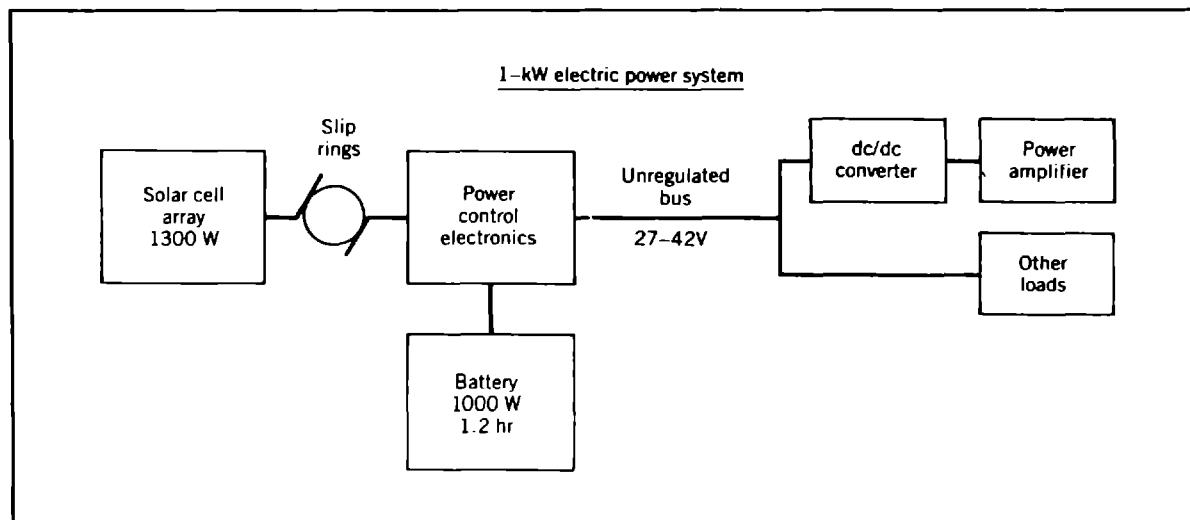


Figure 17 Electric power system (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

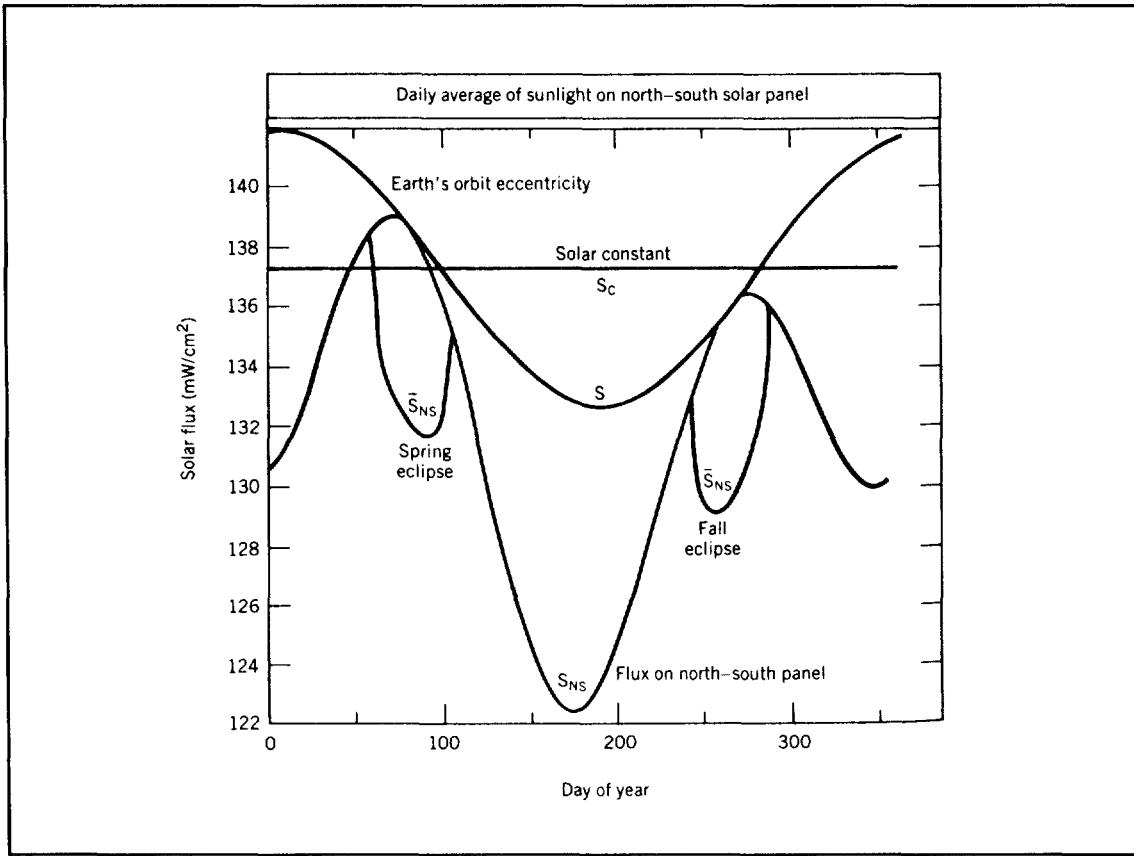


Figure 18 Intensity of solar radiation incident on satellite solar array (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

satellite requires slip rings between the rotating solar array and the rest of the power subsystem. A spinner satellite needs slip rings between the power subsystem, on the rotating part, and the stationary communications platform.

Control electronics connect the solar array and batteries to the various loads. This control regulates the bus voltage, and the charging and discharging of the batteries. Different satellites have varying degrees of bus voltage regulation. Most of the power in a communications satellite is for the power amplifiers. It is more efficient to do the voltage regulation at the dc/dc converter for the power amplifier, rather than in the control electronics.

4.1 Solar Radiation

The average solar intensity above the earth's atmosphere (air mass zero) is 1370 W/m^2 . This is shown as a horizontal line in Figure 18. This is often called the "solar constant", but the intensity is not really constant. The maximum intensity occurs at perihelion (closest approach to the sun) in early January. This is about two weeks after the winter solstice (around December 22),

when the sun has its most southern declination. The curve labeled "S" in the figure shows the maximum solar intensity as about 1415 W/m^2 in January and a minimum of 1325 W/m^2 in July.

Many communications satellites have solar arrays that are sun oriented about a north-south axis, but are always perpendicular to the equatorial plane. When the sun is far from the equatorial plane (December and June) the solar intensity is decreased by the cosine of 23° . The available intensity for most solar panels is the curve in the figure labeled S_{NS} . It shows that solar arrays on most communications satellites deliver the minimum power in June.

The average intensity on satellite solar arrays is also reduced by eclipses, when the satellite is in the earth's shadow for part of its orbit. For geostationary satellites these eclipses occur at local midnight, in the days within three weeks of the vernal and autumnal eclipses. In the spring, they occur in March and the first half of April; in the fall the eclipses occur in September and the first half of October. The maximum duration of an eclipse is 72 minutes. The average intensity is shown in Figure 18 as the curves labeled "spring eclipse" and "fall eclipse."

4.2 Solar Cell

Silicon solar cells are used to convert the power in solar radiation to electricity. A cross section of a typical cell is shown in Figure 19. Each cell is a single crystal of pure silicon, doped so most of it is p-type, and positive electron holes carry most of the electricity. The top layer is doped differently, so it is n-type, where free electrons are the majority carrier. Free electrons are collected by a grid on top of the cell. This grid forms the negative electrode, and positive current is collected by

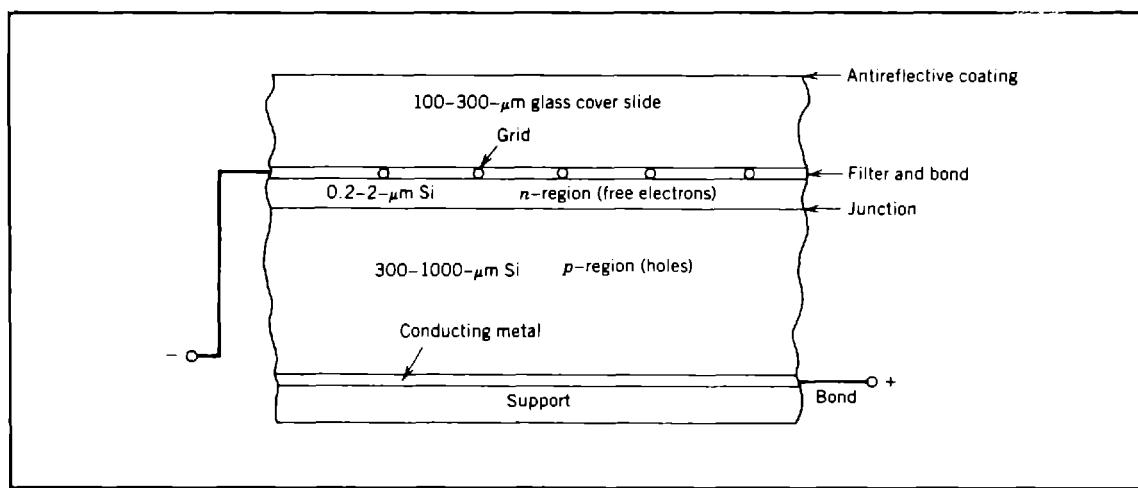


Figure 19 Typical single crystal n-on-p silicon solar cell (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

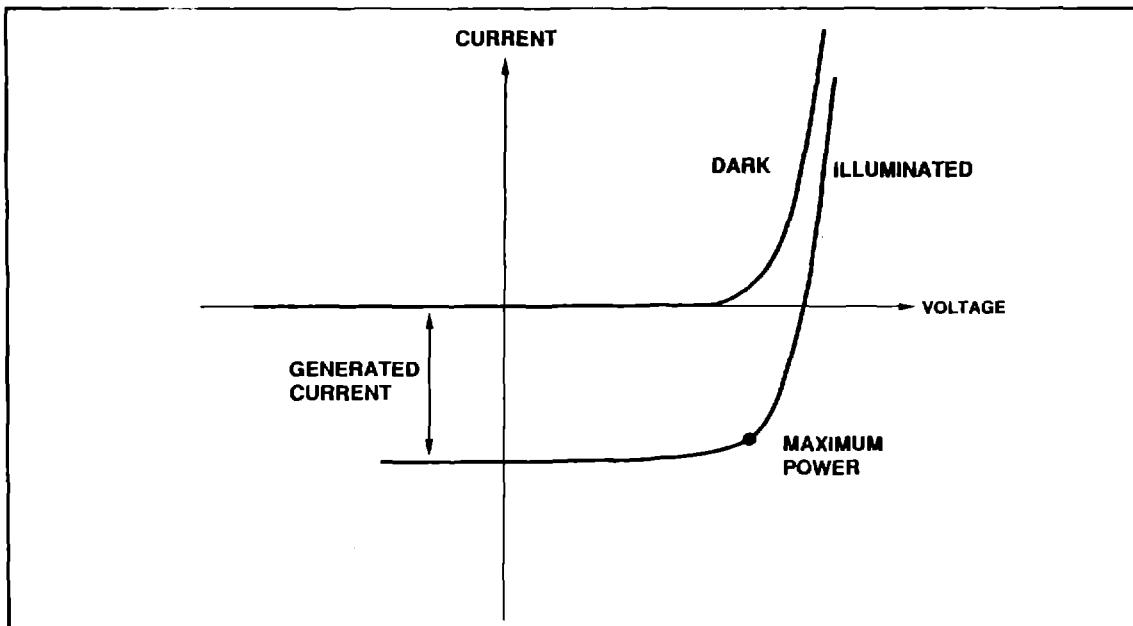


Figure 20 I-V curve for a solar cell, with and without illumination.

conducting metal on the bottom of the cell (positive electrode). A cover slide is placed over the cell to protect it from particle radiation in space; an antireflective coating reduces the amount of light lost by reflection.

When sunlight is incident near the junction, electrons are knocked out of the lattice. This creates free electrons that drift into the n-type material, and holes in the lattice that drift to the p-type material. An intrinsic electric field separates the free electrons from the electron holes, and inhibits their recombining. The current generated is proportional to the number of photons absorbed, and the voltage is a function of the band gap of the material.

When a solar cell has no incident radiation, its electrical characteristics are that of a diode. This is shown as the I-V curve labeled "dark" in Figure 20. When a forward bias voltage is applied (to the right), the current increases exponentially. When a reverse bias voltage is applied (to the left), the reverse saturation current is so small it doesn't show in the figure. Without incident radiation a solar cell is a passive device, and the current is always in the direction of the applied voltage.

Incident solar radiation produces an additional current, which shifts the curve downwards in the figure. The current is now in the opposite direction to the voltage, that is, producing power. The maximum power is generated when the product of the current and voltage is a maximum; this is labeled the maximum power point. When the voltage changes from the maximum power voltage the power produced is decreased, sometimes by a significant amount.

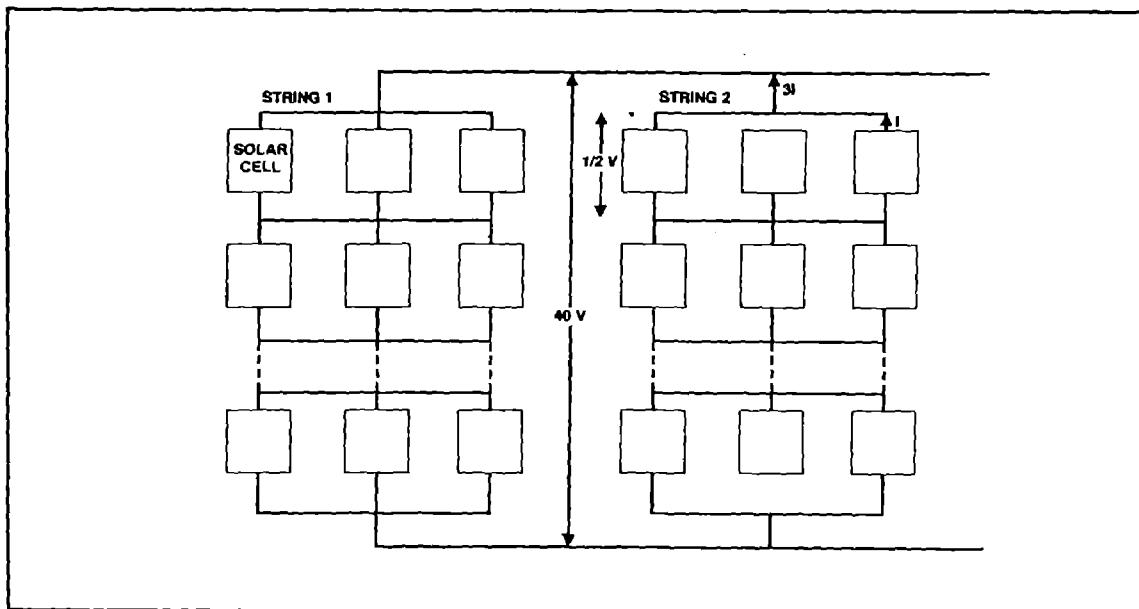


Figure 21 Connections of strings of solar cells.

A simple equation for the current I of a solar cell is

$$I = I_L - I_o(e^{qV/aKT} - 1) \quad (\text{mA}) \quad (2)$$

where I_L = current generated by solar radiation

I_o = reverse saturation current

q = charge of an electron (1.602×10^{-19} coulomb)

V = voltage

k = Boltzmann's constant (1.38×10^{-23} J/K)

T = absolute temperature

a = adjustable constant of the order of unity

While not exact, this is a useful equation. Some parameters vary with temperature, so the equation can be misleading in studying temperature dependence.

4.3 Solar Array

A complete solar array requires thousands of solar cells. The number of cells in series determines the desired array voltage. For increased reliability, strings of cells are often configured as shown in Figure 21. Three or four cells are connected in parallel, and then these sets connected in series. For maximum power, the cells are matched for compatible I-V characteristics.

Damage to a single cell often results in a power loss greater than the power lost from that cell. A short circuit in one cell will short circuit the other two cells in parallel. The loss in voltage must be made up by the other cells in the string, which may no longer be delivering maximum power.

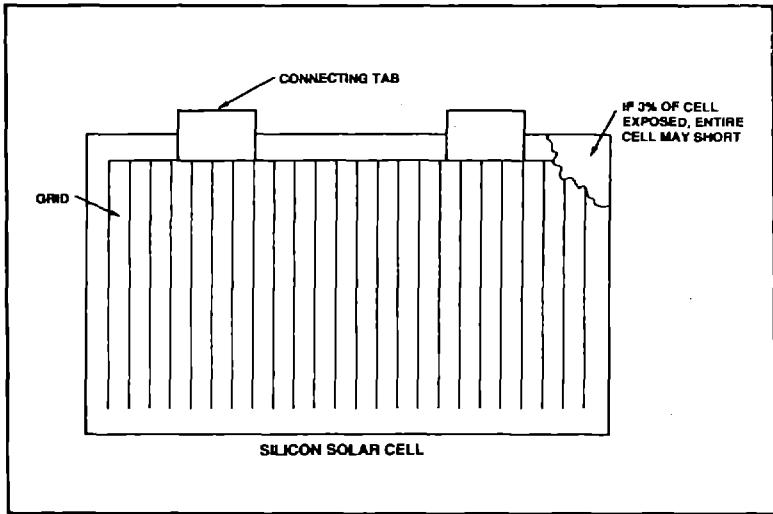


Figure 22 Incomplete cover slide on solar cell caused loss of power.

An open circuit in one cell is less common than a short, but the power loss is greater. All the current must go through the other cells in parallel. If there are three cells in parallel and one fails open, the string current is reduced from three times the maximum power current to about twice the short-circuit current. The string tries to force more current through the two remaining cells. However, the I-V curve in Figure 20 shows that the current

can increase only slightly. The voltage across these two cells drops, and may even go below zero (reverse voltage). The power generated by the entire string depends on the current, and drops to roughly two-thirds (65 to 75%). The power loss in such a string may be 25%, even though only one cell in several hundred is defective.

An example of unexpected radiation damage occurred on an early INTELSAT satellite. A mistake was made, and 3% of the solar cell was not protected by the cover slide (see Figure 22). At first this seemed to be an unimportant detail, only a 25% loss of 3% of the power. After the satellite was launched, the degradation in power was unexpectedly high. In that 3% area the effect was to introduce a parallel resistance that tended to short out the cell. Fortunately, there was enough power margin so the extra degradation did not affect the communications capability.

4.4 Voltage Variations

When the sun is eclipsed by the earth's shadow (and sometimes by the moon's shadow) the solar array stops working. Power is then delivered by nickel-hydrogen batteries. This causes additional voltage variations, as shown in Figure 23. For a constant load the battery voltage decreases during discharge. After the eclipse the solar cells are cold, especially for deployed solar panels. Cold solar cells are more efficient, and deliver a higher voltage until they return to normal temperatures. Unfortunately, this extra power is unwelcome, and precautions must be taken to handle it. The power control electronics and the DC/DC converters must handle all these variations in voltage, and deliver the required voltages to the power amplifiers and other loads in the satellite.

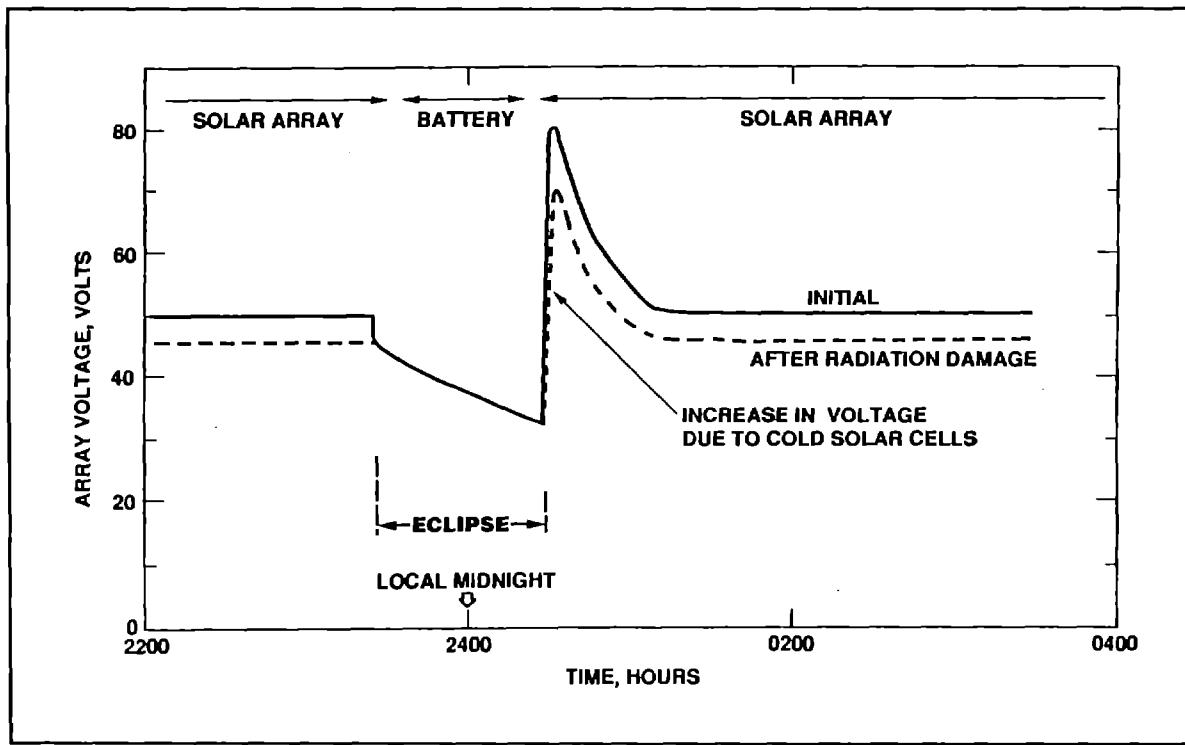


Figure 23 Variations in available voltage due to eclipse and radiation damage.

4.5 Radiation Effects

The solar array is the most radiation sensitive component of the electric power supply. Some radiation damage is expected, and the beginning-of-life power is purposely made larger than the expected end-of-life power. The sensitive portion is the pn-junction, located very close to the cell surface. The cells must be on the external surface of the satellite. A thin, transparent, radiation shield protects the cells from low energy protons. In some cases the fractional power loss may be much larger than the fractional damage produced.

When particle radiation is incident on a silicon semiconductor crystal, two effects are possible. First, electron-hole pairs may be produced in great abundance. In some circuits this may be disastrous by generating an extraneous signal or resetting a register. In solar cells these extra electron-hole pairs are unimportant. They produce extra power, but the effect is transient, and usually negligible.

The second effect is that an energetic particle can knock a silicon atom out of the lattice. This creates a vacancy in the lattice and an interstitial atom. This effect permanently damages the solar cell, and reduces its efficiency. The maximum power output is shown in Figure 24 as a function of 1-Mev electron irradiation. The actual particle radiation in space includes both

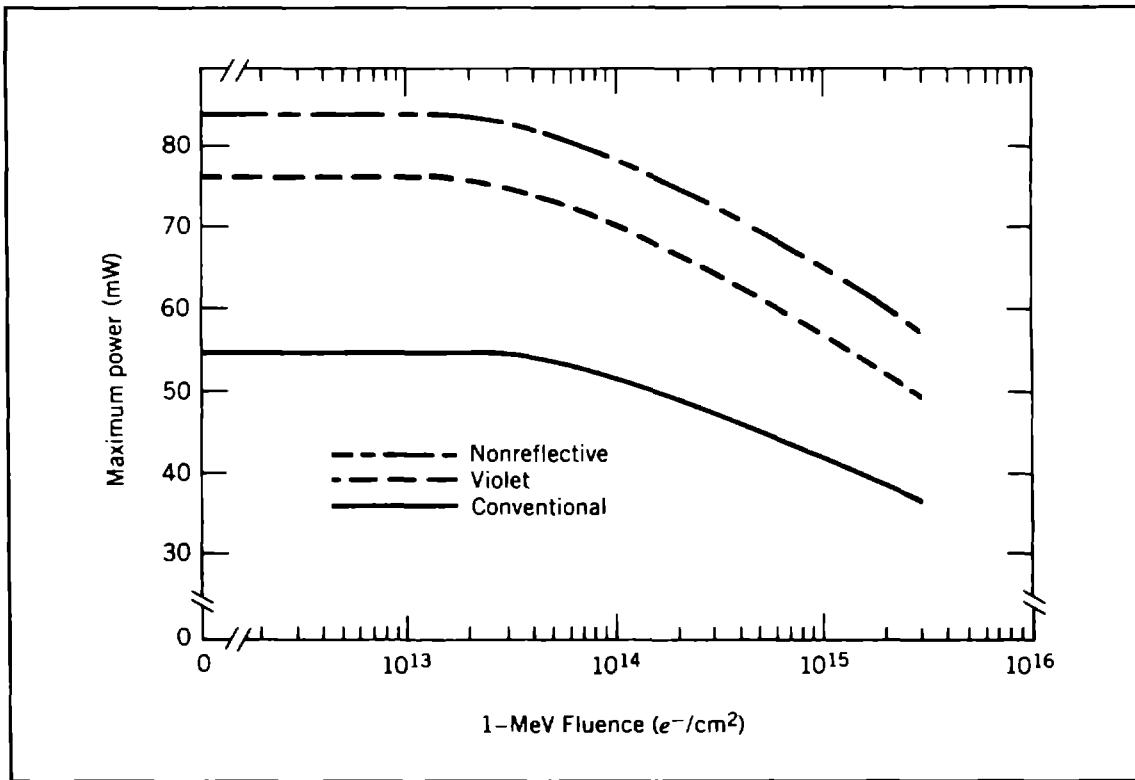


Figure 24 Power output degradation due to 1-Mev electron irradiation (Allison et al., 1975; reprinted by permission from COMSAT Tech. Rev.).

protons and electrons of many different energies, but solar cells are usually tested with 1-Mev electrons.

A well-protected cell (on a spinning satellite, with a coverslide about $300\ \mu\text{m}$ thick) might receive the equivalent of $3 \times 10^{14}\ e^-/\text{cm}^2$ in 7 years at geostationary orbit. On a lightweight, deployed solar array, a cell with a thin coverslide and exposed to radiation from both sides may receive the same amount of radiation in 1 year. The intensity of solar flare protons also varies from year to year, depending on the solar activity. Usually the radiation degradation due to the electron fluence is plotted on a logarithmic scale. This can be misleading to an individual not used to this graph. If it takes 7 years to reach $3 \times 10^{14}\ e^-/\text{cm}^2$, it takes 70 years to reach $3 \times 10^{15}\ e^-/\text{cm}^2$. With time, cells become more resistant to particle radiation, and the changes in later years are much less than in the first year.

5. Spacecraft Attitude

Communications satellites require antennas pointed at the earth, and that focus the radiated power into narrow beams. The attitude control system maintains the satellite orientation.

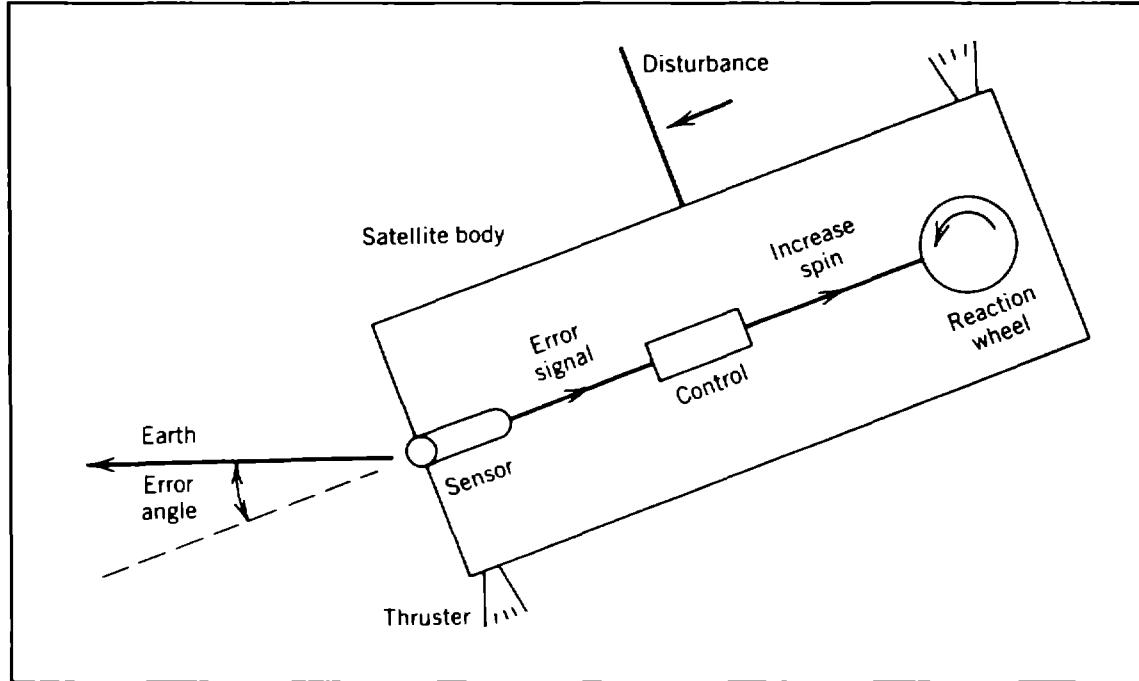


Figure 25 Concept and parts of attitude control system in body-stabilized satellite (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

The concept and parts of a simple attitude control system are shown in Figure 25. A sensor detects an error in pointing by a sensor, and this is corrected by changing the speed (or axis) of a rotating wheel. If the wheels have reached their limit, thrusters are used to increase (or decrease) their speed. A practical attitude control system must operate with three axes, and have a variety of controls.

The antennas require an attitude control system that will keep them pointed at the earth, frequently within 0.1° or 0.01° . Figure 26 shows a block diagram of an attitude control system. A sensor detects any error in pointing, which is corrected by changing the speed (or axis direction) of a rotating wheel. If the wheels have reached a limit, firing thrusters change the total angular momentum. The performance specifications of an attitude control system are set by the disturbance torques and the required pointing accuracy. The major disturbances on a geostationary satellite are torques due to solar radiation pressure and misalignment of thrusters.

A control system takes the information from the sensors, and provides commands to the torque generators. Wheels can change the satellite orientation, but do not change the total angular momentum. Thrusters are always available to generate a desired external torque. To save fuel, magnetic torquing coils, or solar flaps, can provide a weak, but steady, torque.

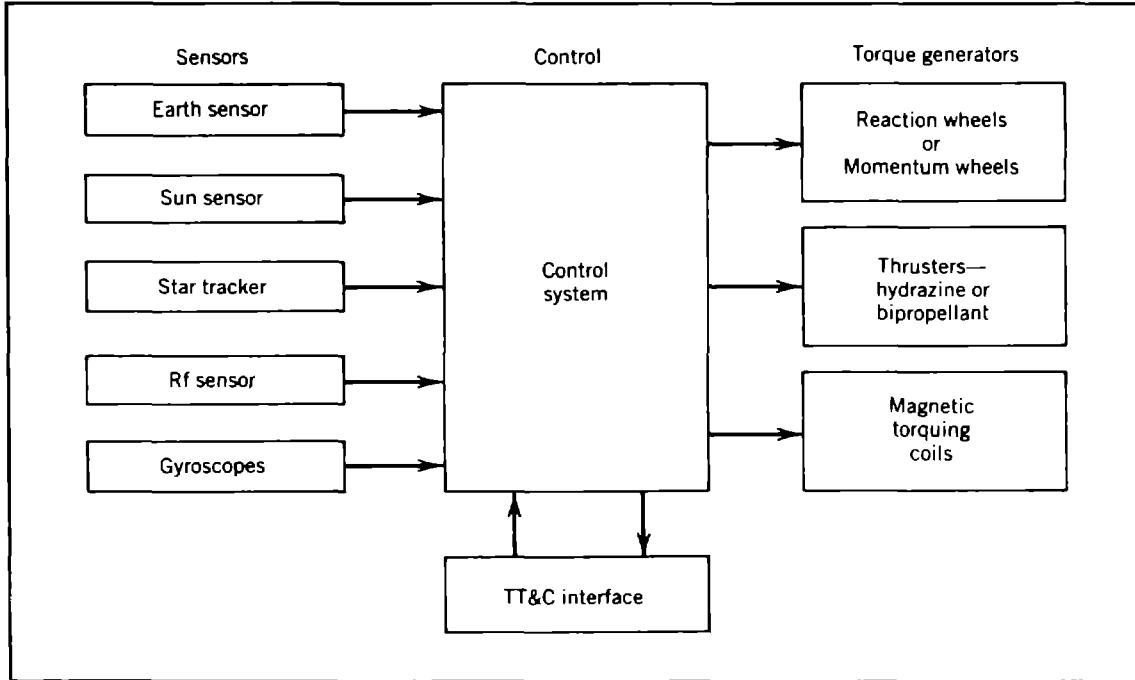


Figure 26 Attitude control system. Satellites have only some of the sensors shown (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

An attitude control system is a balance between various requirements, hardware available, and methods to use the hardware to fulfill the requirements. The attitude control elements for the INTELSAT VI satellite are shown in Figure 27. Sun sensors, earth sensors, and nutation accelerometers provide the error signals. Sun sensors are light and accurate, but obviously are no good during an eclipse. Earth sensors directly measure the direction for pointing. Star trackers are not used, primarily because of their weight and expense. RF sensors (not shown) measure the direction of a RF beacon on earth. Spinning satellites have a bearing and power transfer assembly (BAPTA) to connect the spinning and non-spinning portions. The electronics consists of a microprocessor with many safety features.

The attitude control electronics operate in several different modes. It takes information from different sensors and actuates wheels, motors, or thrusters. Because of its complexity it is subject to transient problems, or permanent damage.

The safety features of the INTELSAT VI attitude control system are highlighted in Figure 28. The earth sensor detects errors in east-west pointing, which is controlled by changing the BAPTA motor speed. The accelerometer detects nutation, which is controlled by varying the motor torque or by firing thrusters. Spin axis misalignment is detected by measuring the daily roll error with the earth sensor. This is corrected by firing thrusters. The attitude control processor has several fault protection features. The

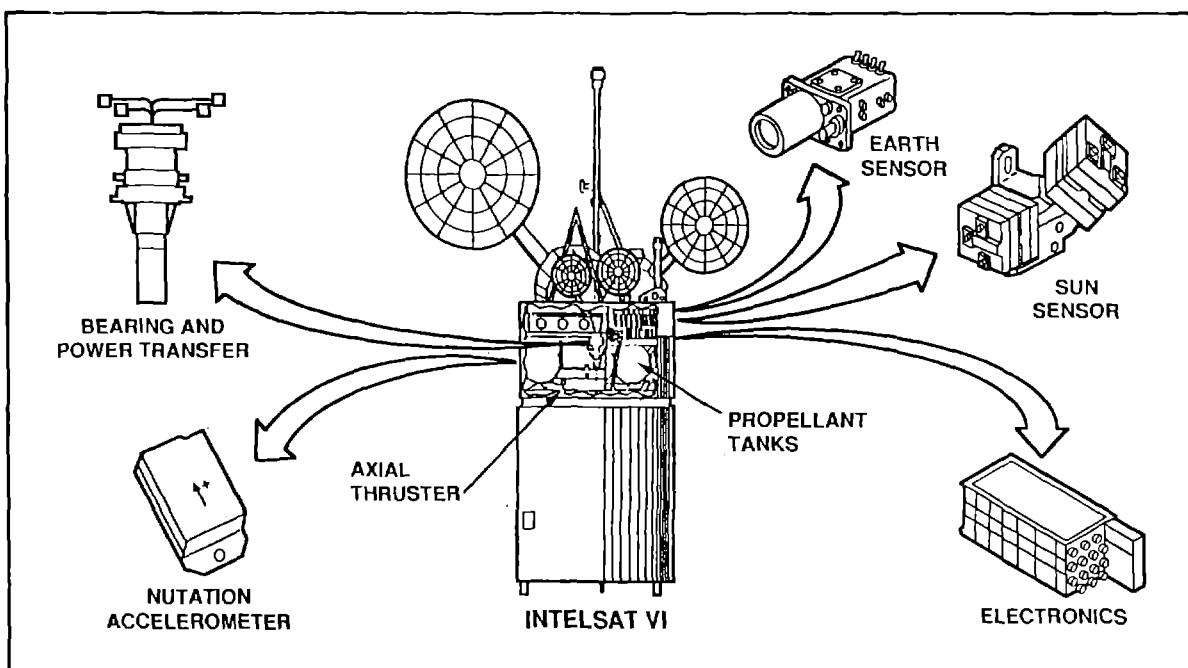


Figure 27 Elements of the attitude control system for INTELSAT VI, a spin-stabilized satellite.

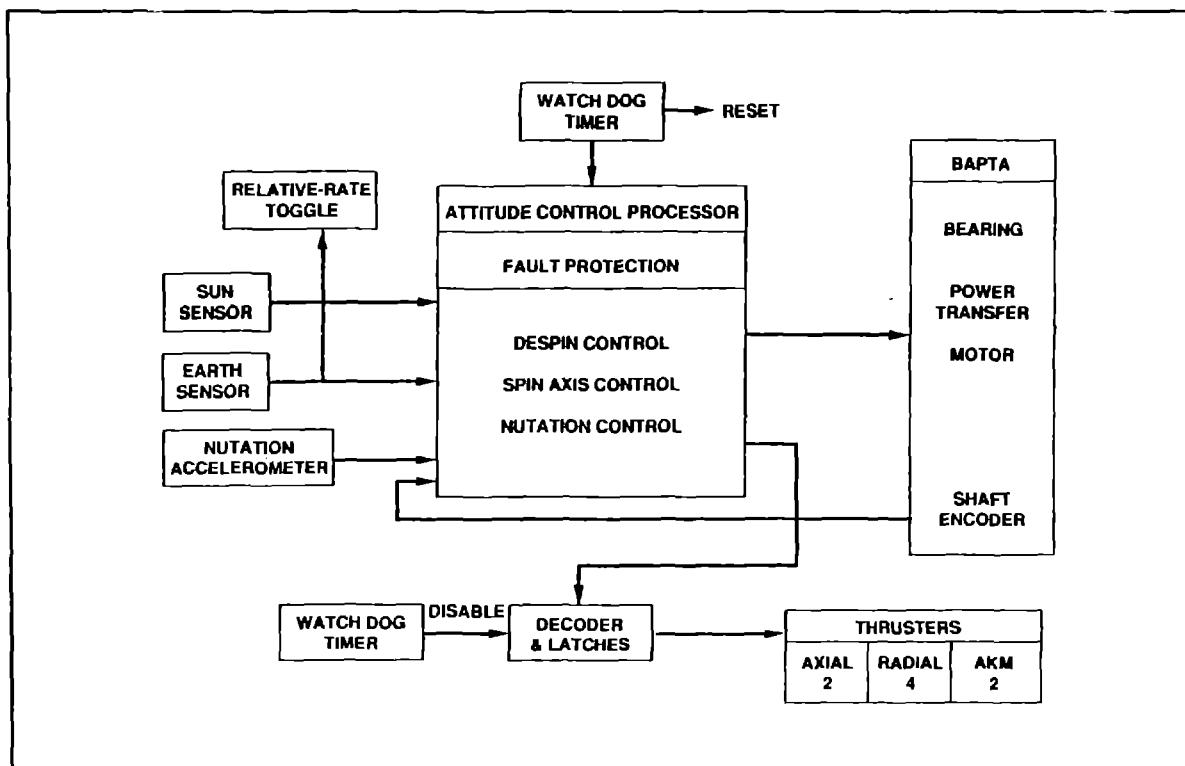


Figure 28 Safety features in the INTELSAT VI attitude control system.

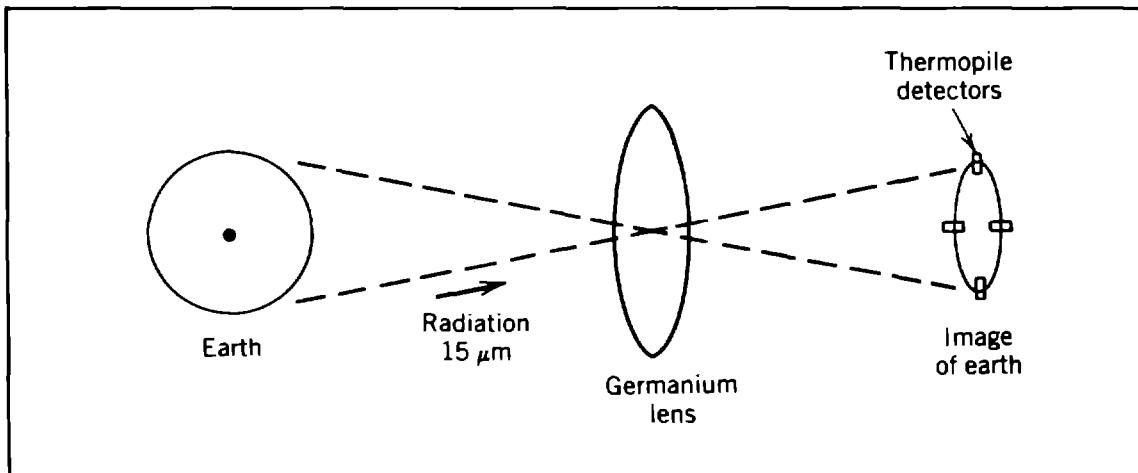


Figure 29 Static earth sensor detects the earth's horizon by infrared radiation (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.)

watchdog timer detects a stoppage in the processor, and resets it automatically. It also watches the amount of thruster firing, and disables them if too much firing is occurring. Finally, there is a relative rate toggle completely independent of the rest of the system. If all else fails, the toggle prevents the satellite from going into a flat spin.

5.1 Attitude Sensors

Earth sensors can detect the disk of the earth against the background of space by measuring the infrared radiation, usually at wavelengths near $15 \mu\text{m}$. Three types of commonly used horizon sensors are: The static horizon sensor generates an error signal until it is aligned precisely with the earth. The scanning horizon sensor uses an oscillating mirror to pass the earth's image over a sensor, and generates a signal for the leading edge and the trailing edge. A spin infrared sensor follows the same principle, but is used in a spinning satellite, which generates its own scan as it rotates. Many other variations and combinations are possible.

The basic components of a static horizon sensor are shown in Figure 29. The earth is imaged by a germanium lens, which transmits the infrared radiation. Four detectors are located around the edge of the image. If the image moves toward detector, made of a thermopile, the additional radiation heats the detector and generates a stronger signal. Sensitive error signals can be generated by taking the difference between opposite detectors.

A horizon sensor used on spinning satellites is shown in Figure 30. A single bolometer measures the presence or absence of the earth's image. As the satellite rotates, the earth's image sweeps across the detector. The midpoint of the signal provides

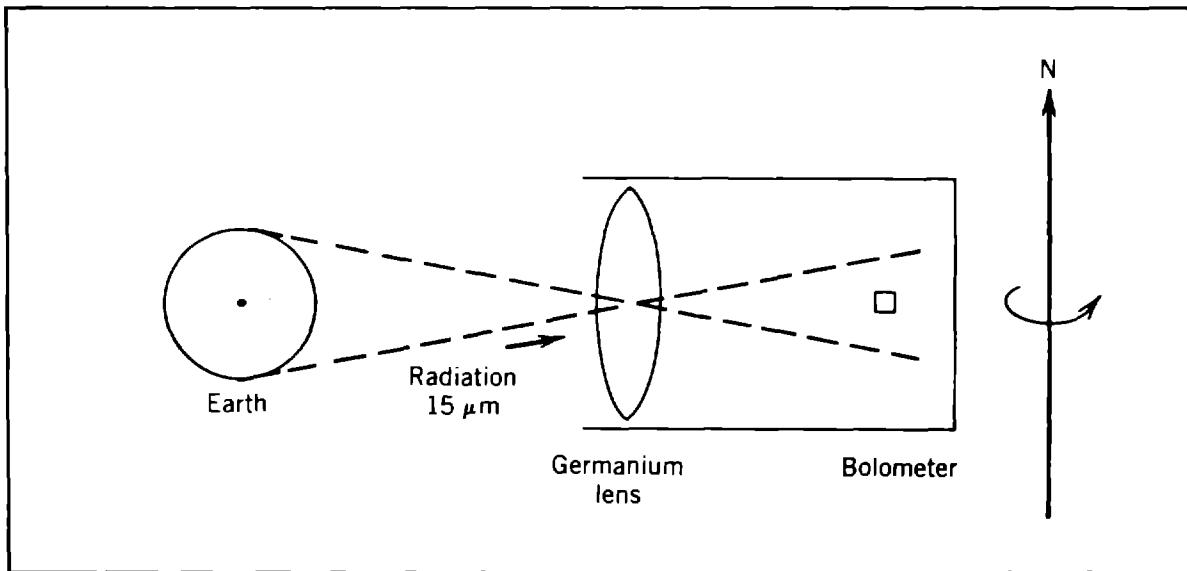


Figure 30 Earth sensor for a spinning satellite (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

east-west information, when the detector is centered on the earth's meridian going through the subsatellite point. The length of the signal provides roll information; the nominal scan is offset from the equator. If the pulse length is longer than nominal, the scan is closer to the equator. If the pulse length is smaller, the scan is farther from the equator. A scanning horizon sensor operates in the same general way for a body-stabilized satellite; a special oscillating mirror moves the earth's image across a detector.

A sun sensor, shown in Figure 31, can be small (a few cm) because the sun is such a bright source. When the sun shines through a narrow slit a bright line is formed that can be detected by small solar cells or other light detectors. Frequently

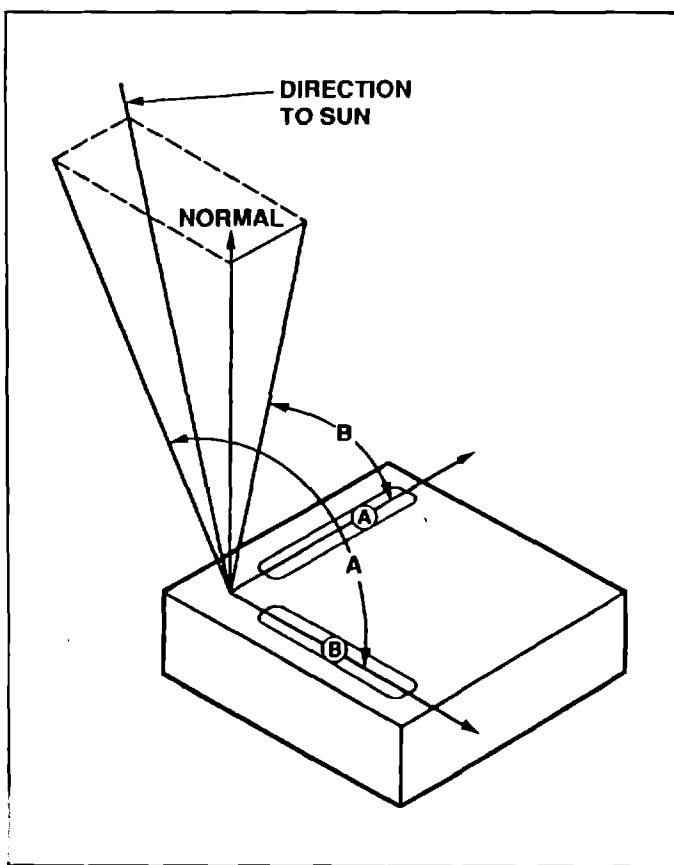


Figure 31 Sun sensor has two slit detectors at right angles.

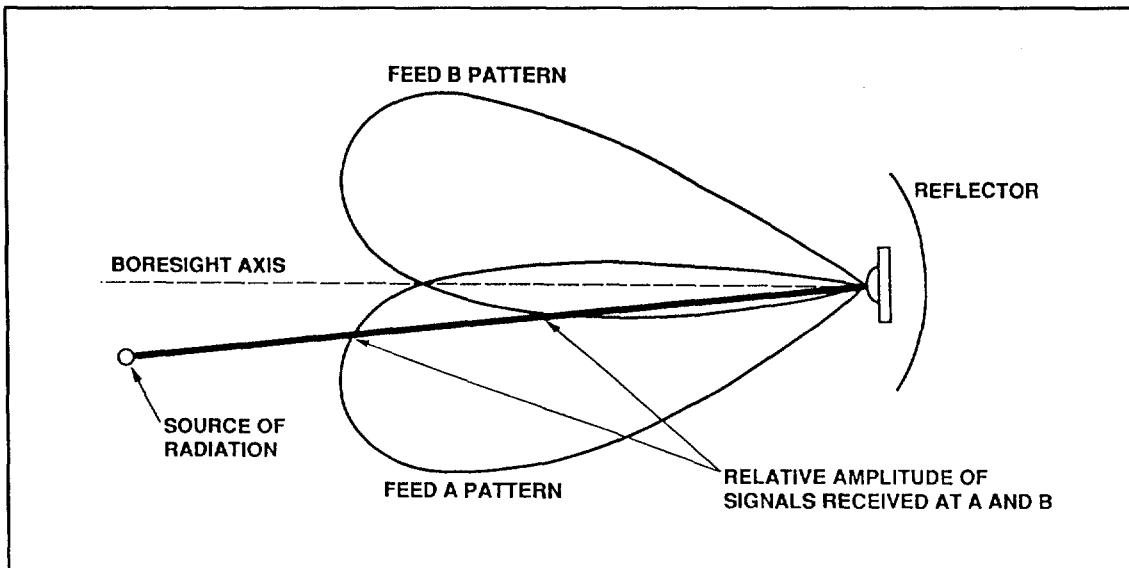


Figure 32 The amplitude-sensing pattern of an RF sensor.

several detectors are coded with masks, so a different combination of detectors is illuminated for each angle. The information can thus be provided directly in digital form. Figure 31 shows two slit detectors (A and B) at right angles which provide the direction of the sun precisely and over a wide field of view. Detectors under the A slit determine angle A, and detectors under the B slit determine angle B.

Another method of determining the direction of the earth is to detect a radio frequency from the earth with an RF sensor. The principle for a single plane is shown in Figure 32. Two feeds with a single reflector produce two antenna patterns in slightly different directions. If a source of radiation is more in the A pattern, the amplitude of the signal received in feed A is larger than that in feed B. By changing the satellite attitude until the two amplitudes are equal, the boresight axis will then be pointed at the RF signal. To measure both pitch and roll, four feeds are normally used, and two error signals are generated.

6. Temperature of Satellite Components

There is no air in geostationary orbit, so heat is transferred only by conduction through solids and radiation through space. An electrical engineer has to learn that a 1-W resistor is rated as 1 W for an air environment, and it may overheat and burnout in a space environment.

The temperature of a satellite is determined by incident solar energy, internal electrical dissipation, and thermal radiation into space. The heat q radiated by a black body is proportional to the fourth power of the absolute temperature T and is given by the Stefan-Boltzmann law of radiation as

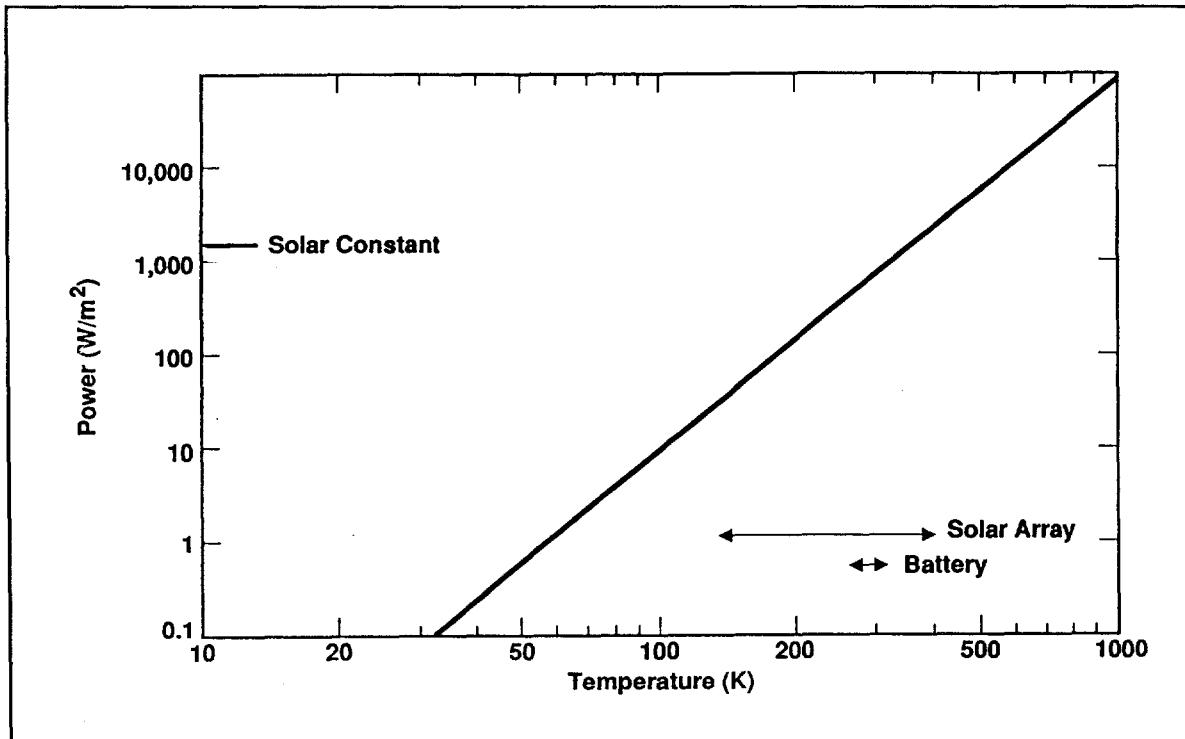


Figure 33 Power radiated by a black body, according to the Stefan-Boltzmann Law.

$$q = A\sigma T^4 \quad (W) \quad (3)$$

where A is the surface area and σ the Stefan-Boltzmann constant ($= 5.6703 \times 10^{-8} \text{ W/m}^2\text{K}^4$). For nonblack bodies the radiated power is less and is written as

$$q = \epsilon A\sigma T^4 \quad (W) \quad (4)$$

where the emissivity ϵ is always between zero and unity.

The radiated power of a black body per unit area is shown in Figure 33 for various temperatures. Thus, for a temperature of 294 K (70°F) the radiation is 424 W/m². If 42 W is to be radiated, a radiator of at least 0.1 m² is required (black, with no incident radiation). If there is 310 W/m² incident (from an environment of 272 K), the net heat transfer is 114 W/m², and a 0.1-m² radiator would radiate a net power of only 11.4 W.

The figure also shows the usual temperature limits of a battery and of a solar array. Frequently the battery has the tightest temperature limits on a communications satellite, while the solar array often exhibits the widest temperature variations. The solar constant (1370 W/m²) is important in determining the temperature of a satellite (and of the earth), and is shown in the figure.

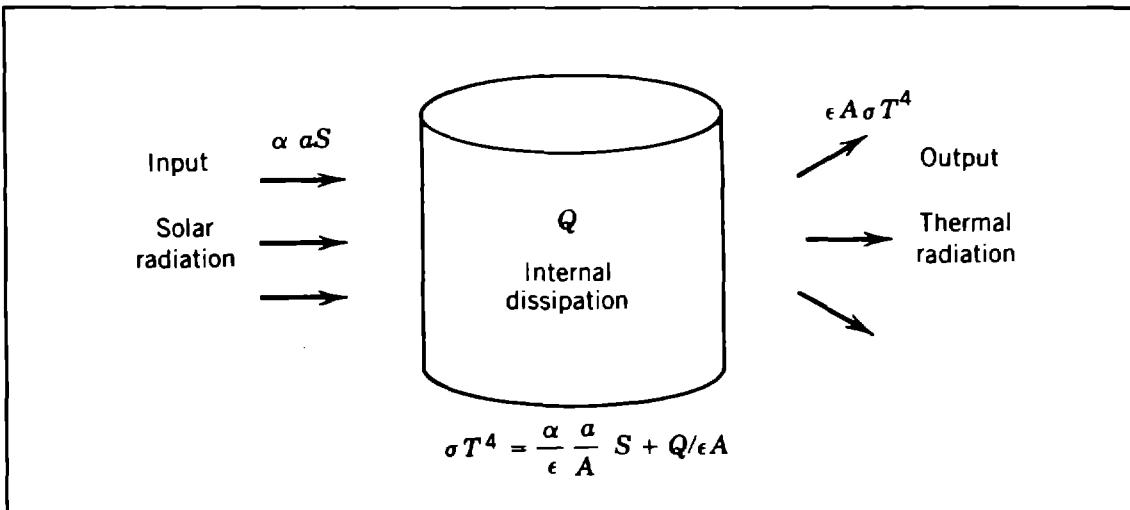


Figure 34 Average equilibrium temperature of a satellite depends on shape (a/A), surface properties (α/ϵ), and internal dissipation (Q) (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

The balance between power absorbed and power radiated is shown in Figure 34. The input solar power is equal to the product of: the surface absorptivity α , the projected area a , and the solar constant S (average = 1370 W/m^2). The output power radiated is equal to the product of: the emissivity ϵ , the total surface area A , the Stefan-Boltzmann constant σ ($= 5.6703 \times 10^{-8} \text{ W/m}^2 \text{ K}^4$), and the fourth power of the absolute temperature T .

On the average the input solar power plus the electrical dissipation Q , must equal the output power radiated. The average equilibrium temperature T_E of a satellite is

$$\sigma T_E^4 = \frac{\alpha}{\epsilon} \frac{a}{A} S + \frac{Q}{\epsilon A} \quad (\text{W/m}^2) \quad (5)$$

The temperature depends on the ratio of absorptivity to emissivity α/ϵ and the ratio of projected to total surface area a/A . Both of these can be changed. The first is of the order of unity, and the second around 1/4.

To control the average temperature the thermal properties of different surfaces are carefully specified. These are both the thermal emissivity and the absorptivity to sunlight. Temperatures of different parts inside a satellite can be controlled by changing the thermal coupling and with electric heaters. Some of the more important thermal tasks in a communications satellite are to: maintain the batteries within narrow temperature limits, keep the thruster fuel from freezing, and dissipate the large amount of heat generated by the power amplifiers.

6.1 Temperature Transients

When there is more heat going into a body than coming out, its temperature rises. The net thermal power input is equal to power from radiant, thermal, or electrical sources minus the dissipated thermal power. The rate at which the temperature rises (dT/dt) is equal to the net thermal power input divided by the thermal mass of the body. The thermal mass is equal to the product of the mass m and the specific heat c .

A thermal equation for an isothermal satellite in sunlight can be written as

$$mc(dT/dt) = \alpha aS + Q - \epsilon A\sigma T^4 \quad (W) \quad (6)$$

The rate of change of the temperature (dT/dt) depends on the solar input αaS , the internal heat generated Q , and the heat radiated $\epsilon A\sigma T^4$. If the conditions are constant, and the time is long enough, the temperature will eventually reach equilibrium.

If the initial temperature is below the equilibrium temperature, the thermal input heats up the body. The solution to the differential equation can be written as

$$\frac{t}{\tau} + C = \ln \frac{1 + T/T_E}{1 - T/T_E} + 2 \tan^{-1} \frac{T}{T_E} \quad (7)$$

where C is the constant of integration, and the functions are the natural logarithm and arc tangent in radians. The time constant τ is defined as

$$\tau \equiv mc/(4\epsilon A\sigma T_E^3) \quad (s) \quad (8)$$

While time can be written explicitly as a function of temperature, it is impossible to write this temperature as a function of time.

It is useful to plot the equation as a graph so that the temperature can be obtained if the time is known. This graph is shown in Figure 35. In practice the initial temperature is usually known, and the graph is first used to determine the constant of integration. Then the time increment (t/τ) is added, and the graph is used again to determine the final temperature.

For radiative cooling the solution is

$$\frac{t}{\tau} + C = \ln \frac{T/T_E + 1}{T/T_E - 1} - 2 \tan^{-1} \frac{1}{T/T_E} \quad (9)$$

This is shown in Figure 36. It is similar in form and use to the equation for heating, except that the initial temperature is lower than the equilibrium temperature.

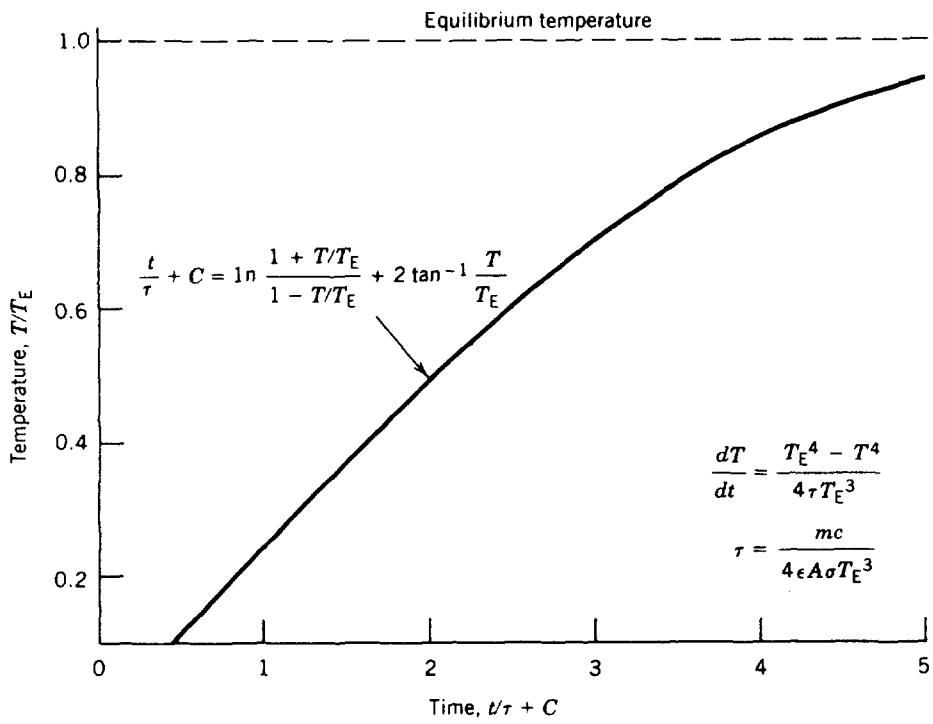


Figure 35 Temperature transient for radiative heating (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

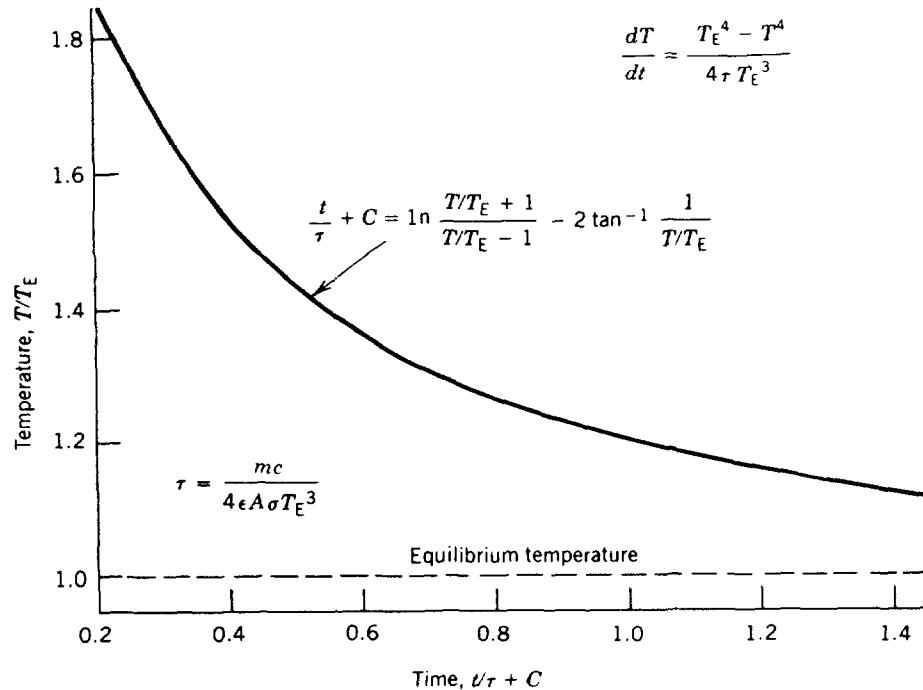


Figure 36 Temperature transient for radiative cooling (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

When close to equilibrium, the temperature difference decreases exponentially. The difference between the body and its equilibrium value decreases by a factor of $1/e$ for each elapsed time constant τ . This temperature difference usually becomes negligible when it is smaller than the accuracy of the temperature measurement, or smaller than any variation in T_E . The body can then be considered to have "reached" equilibrium. This may take from one to perhaps five time constants, depending on whether a factor of $1/e$ or a factor of $1/100$ is required. Thus, determining the temperature for an elapsed time longer than five time constants requires only the steady-state solution.

The time constant for a large compact satellite is many hours. When a geostationary satellite goes into a 1-hr eclipse the equilibrium temperature is around 70 K, but the satellite never gets close to equilibrium. Interior points will lag the average temperature, and may only drop ten or twenty degrees. Solar panels on a geostationary satellite are light weight, with large areas, so the time constant is much shorter (1-5 hr). In a 1-hr eclipse the solar cell temperature may drop to 150 K. Thin aluminum foil has a much shorter time constant (seconds). If insulated from the satellite, the foil may reach very cold temperatures in an eclipse at geostationary orbit.

6.2 Radiatively Coupled Satellite

To determine the temperature at individual points, a satellite is divided into a number of bodies. This division is pictured in Figure 37. Each body is assumed to be isothermal. The thermal coupling between each pair of bodies is determined, which may be by conduction, by radiation, or both. A set of thermal balance equations is written, one for each body. Included are any internal dissipation terms and external radiation inputs and outputs. The set of equations can then be solved (usually with a computer) for the temperature of each body. Both steady-state and transient solutions are possible.

To illustrate the concept of a thermal model, a numerical example will be presented. The complete problem is shown in Figure 38. The spacecraft is a hollow metallic cylinder with a height and radius of

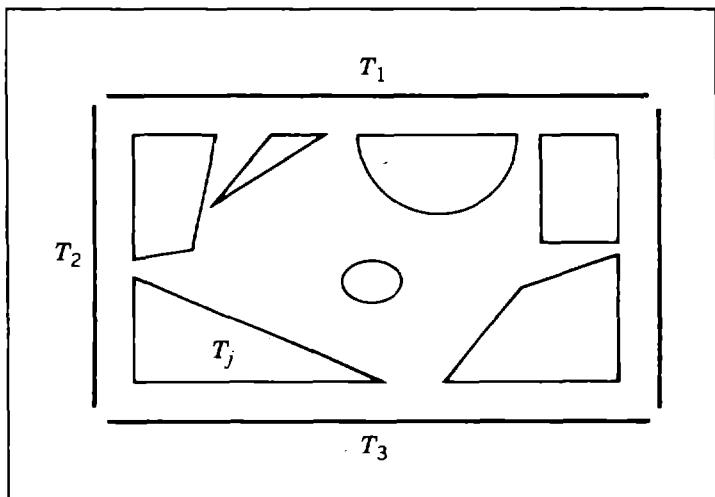


Figure 37 Satellite modeled as n isothermal bodies, surrounded by three walls at fixed temperatures (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

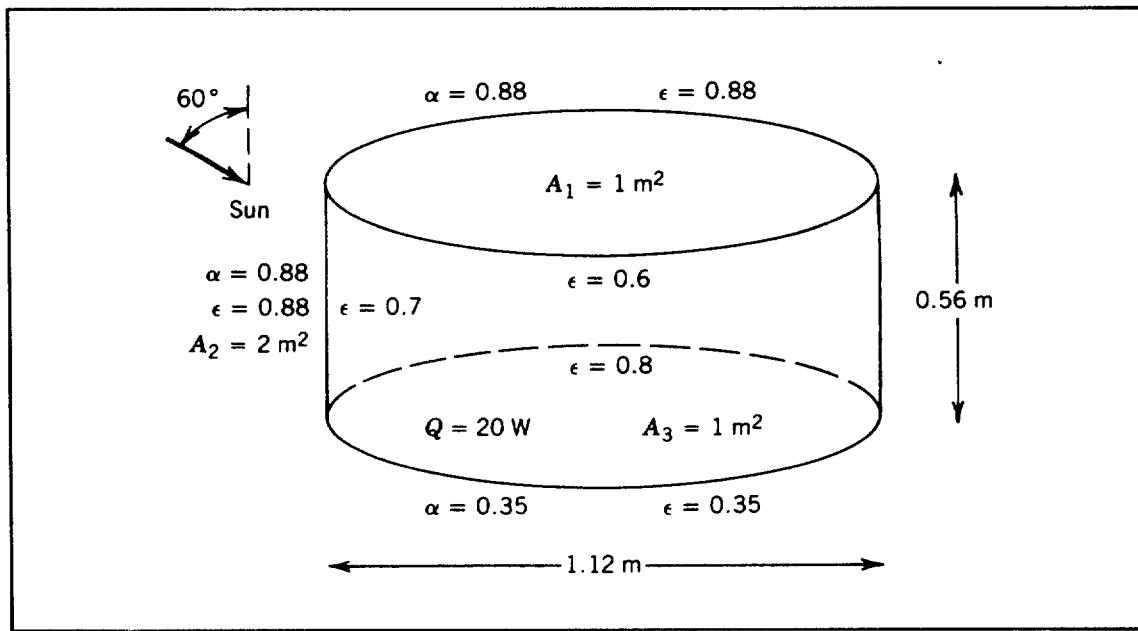


Figure 38 Example of radiatively coupled satellite as three isothermal bodies in the form of a hollow cylinder (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

approximately 0.56 m (actually $1/\sqrt{\pi}$). The thermal emissivity ϵ is given for external and internal surfaces, and the absorptivity to sunlight α is given for the external top and sides. There is 20 W dissipated in the bottom surface. The sun is shining on the satellite, with a 60° angle with the cylinder axis. Each of the three surfaces (cylinder walls, top, and bottom) is assumed to be isothermal. The problem is to determine the temperatures of each surface.

A thermal balance equation is written for each of the three isothermal bodies. This requires calculating the radiation coupling factors which is the fraction of radiation emitted by one body that is absorbed by another body. The three equations are solved for the three unknown temperatures. The solution is shown in Figure 39. The temperature of the top, sides, and bottom is 314 , 287 , and 273 K respectively. The heat flows can be calculated from the temperatures and the radiant coupling factors. Once the heat flows are known, it is possible to predict the effect of changing the couplings between two bodies. For example, the net heat flow from the top to the sides is 55 W ; if the coupling is decreased, the heat flow is reduced, the temperature of the top will increase, and the temperature of the sides will decrease.

6.3 Electrostatic Discharge

The external surfaces of a communications satellite include thermal blankets for insulation, back surface mirrors for maximum radiation, and the cover slides of solar cells. For thermal

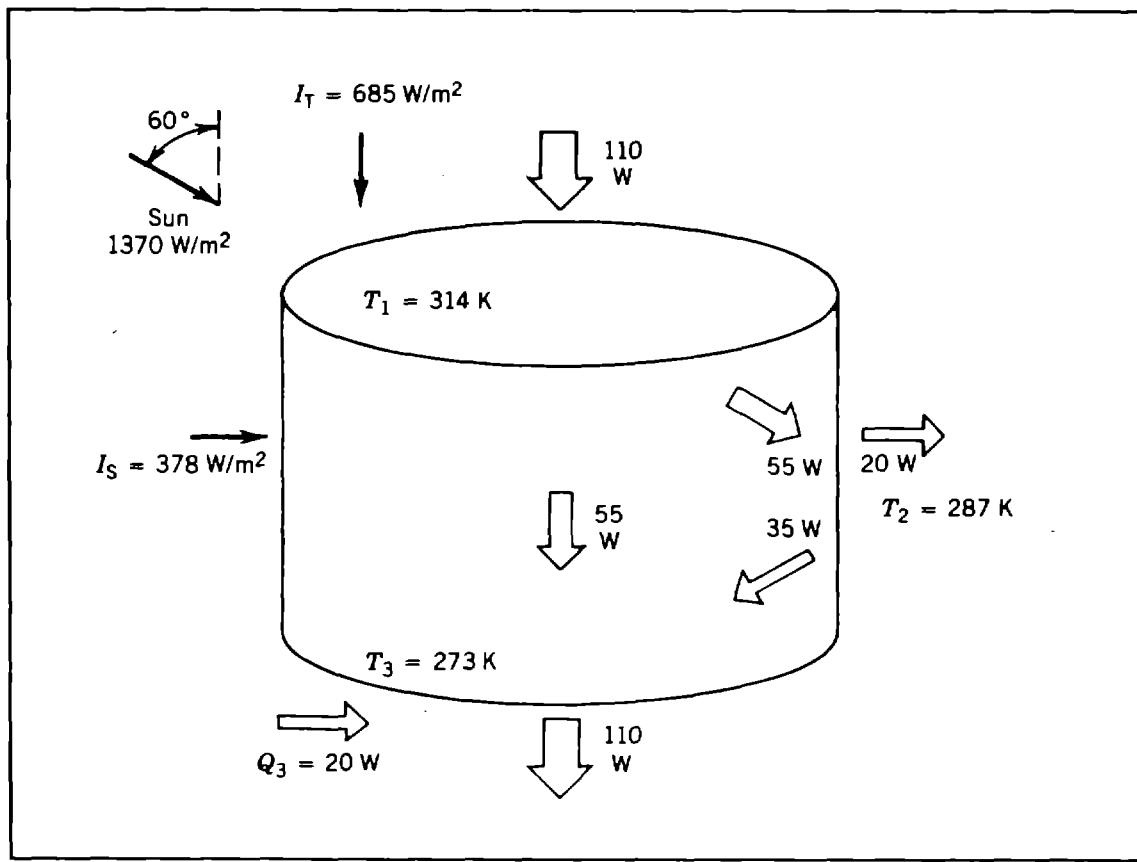


Figure 39 Calculated temperatures and heat flows for satellite example (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

reasons most of the external surfaces are electrical insulators. When bombarded with ionized particle radiation, an electrical charge can result. Since the particles have high energies, the charge increases until there is a voltage breakdown, and an electrostatic discharge.

These discharges produce electrical currents that can induce currents in nearby electrical wires. These in turn can trigger various events in a satellite, and modify stored data. These events have been known for many years, and have been studied under laboratory conditions. In a complex satellite in space it is often difficult to pinpoint the exact cause of a disturbance in normal operations.

7. Spacecraft Testing

Any complex and expensive system requires testing to ensure adequate performance. Communications satellites require extensive testing for two additional reasons: they operate in an environment considerably different from that in which they were built and, after launch, they are inaccessible to routine maintenance and

Space Condition	Critical Components	Simulation
Zero gravity	Array deployment, antenna deployment, liquid motion	Counterweights, balloons, horizontal position
Vacuum (10^{-10} - 10^{-20} torr)	Sliding friction, thermal control, mechanical relay contacts	Use solid lubricant, thermal vacuum tests, reduce current
Solar radiation	Solar array, component temperature	Solar simulator, thermal vacuum tests
Particle radiation	Solar array, semiconductor components	1-MeV electrons
Lack of nearby obstructions	Antenna patterns	Anechoic chamber, outdoor range
Design life of many years	Bearings, batteries, power amplifiers	Accelerated testing

Table 3 Difficulties in Simulating Space (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

repair. The objective of the testing is not necessarily to duplicate the space environment, but approach it sufficiently so that any spacecraft that passes the tests will operate successfully.

The major features of the space environment that are difficult to simulate exactly are zero gravity, high vacuum, solar radiation, particle radiation, lack of nearby obstructions, and a full design lifetime (see Table 3). Except for zero gravity, approximate simulation of these parameters is possible for most spacecraft components. Gravity affects the deployment of solar arrays and antennas, and liquid motion in components such as fuel tanks, heat pipes, and batteries.

Major testing of a spacecraft occurs between integration and shipment to the launch site. This testing requires a considerable length of time (as long as several months). A typical set of tests is shown in Table 4. The basic testing consists of systems performance tests, mechanical vibration tests, and thermal vacuum tests. Various components must be aligned, including sensors, antennas, thrusters, and wheels. The RF test measures the electrical performance of the antennas, not only the intensity distribution in an antenna beam, but also the strength of the sidelobes and possible interference between beams. The mechanical vibration tests and thermal vacuum tests are the most complete tests of the overall spacecraft. Mass properties include the total mass, the position of the center of mass, and various moments of inertia.

Most subsystems are tested separately, to detect problems before spacecraft integration. Each subsystem has performance measurements, vibration tests, and thermal vacuum tests. Specifications for these tests reflect the requirements for each subsystem, the results of computer models of structure and thermal design, and experience with previous spacecraft component testing.

7.1 Solar Cell Testing

The performance of an illuminated solar cell is determined by illuminating it, and measuring its electrical characteristics. A

close spectral match to the solar spectrum can be obtained with a xenon arc lamp. Electrical characteristics are determined by manually or electronically varying load resistances from zero to infinity, and measuring the cell voltage and current. Since solar cell response is a strong function of temperature, the temperature must be determined. The current-voltage data are usually plotted on graph paper with printed power curves, so the maximum power point can be read directly. The short-circuit current I_{sc} and open-circuit voltage V_{oc} can be read directly with digital meters.

In addition to the I-V test, the solar cell is subject to a number of environments, as shown in Table 5. Humidity, peel, and

Test	Parameters or Units Measured
System performance	All units
Alignment 1	Antennas, sensors, thrusters, wheels
Rf range 1	Antennas
Vibration tests	Static, sine, acoustic, shock
Functional	Performance of all units
Thermal vacuum	Temperatures, performance
Alignment 2	Antennas, sensors, thrusters, wheels
Rf range 2	Antennas
Mass properties	Mass, center of mass, moments of inertia

Table 4 Tests of integrated spacecraft before launch (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

Test	Typical Parameter	Observation or Failure ^a
I-V test	Filtered xenon lamp, 1 sun	I_{sc} , V_{oc} , I_{mp} , V_{mp} , temperature coefficient
Humidity	95%, many hours	Electrode separation
Peel	Sticky tape, two directions	Electrode separation
Thermal cycling	70 to 300 K, 1000 cycles	Interconnect failure
Radiation	1-MeV electron beam	Power degradation

^aAbbreviations: I_{mp} , current at maximum power; V_{mp} , voltage at maximum power; I_{sc} , short circuit current; V_{oc} , open circuit voltage.

Table 5 Solar cell testing (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

thermal cycling tests are intended primarily to test the adherence of electrodes and interconnects, while the radiation testing is used to measure the expected performance degradation. The humidity test is used because moisture penetration under the electrode (before launch) can weaken the bond. The peel test, which consists of applying some sticky tape to a cell and peeling it off, does not simulate an actual environment, but it is an easy method of evaluating the adhesion of a cell electrode. The thermal cycling test is often conducted by dipping the cell in liquid nitrogen and then warming it up to room temperature (or above). While this is more severe than the environment that cells experience going into eclipse, it is also a simple test that many cells can survive.

The expected degradation of solar cells in the space radiation environment is determined by a radiation test and a comparison of this test to the space environment. The cells are exposed to a 1-Mev electron beam at normal incidence, and performance degradation is measured. The equivalent 1-Mev electron fluence is calculated based on the known radiation environment and the mission parameters.

A 1-Mev electron accelerator is often the primary source for space radiation simulation. Accelerators produce irradiation rates many orders of magnitude greater than those experienced in space environments; therefore, test times of a few minutes simulate many years in orbit. A successful experiment includes accurate knowledge of the particle energy and the beam intensity. Control samples of solar cells with known properties are used to ensure that beam intensity is correct. Mechanical motion is used to change the positions of the cells, and thereby cancel any variation in cross-sectional beam intensity.

The equivalent 1-Mev electron fluence depends on the shielding on the solar cell, the activity of the sun (in the 11-yr sunspot cycle), the design lifetime of the satellite, and whether the prediction is for the short-circuit current I_{sc} , the open-circuit voltage V_{oc} , or the maximum power P_{max} . In the space environment near earth there are both electrons and protons. With no shielding, the effect of protons trapped in the earth's magnetic field is overwhelming, but with only a little shielding, their effect is negligible. The effects of trapped electrons and solar flare protons are comparable in magnitude. The primary determination of the equivalent 1-Mev electron fluence must be obtained from space flight experiments. This standard can be adjusted for shielding and solar activity, and works well for similar solar cells.

7.2 Vibration Testing

The mechanical launch environment for the payload of a three-stage launch vehicle is shown in Figure 40. The first stage starts at lift-off, later jettisons the solid rocket booster, and ends with the main engine cutoff (MECO). The second stage is ignited, the fairing is jettisoned (no longer needed), and then the engine

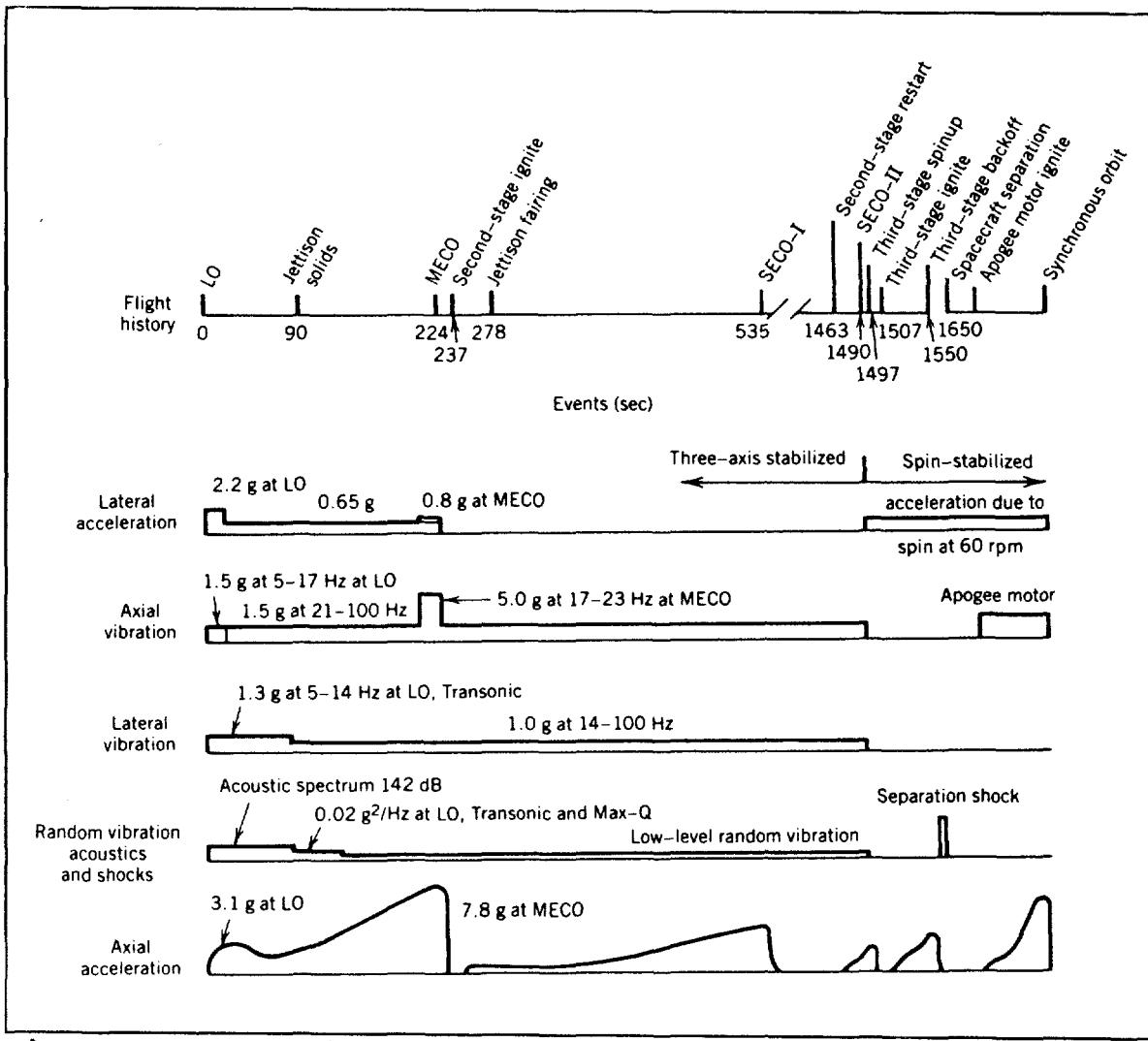


Figure 40 Mechanical launch environment for expendable launch vehicle: LO, lift-off; MECO, main engine cutoff; SECO, secondary engine cutoff (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

is turned off for a coast period (secondary engine cutoff-I, or SECO-1). There is a quarter of an hour coast period in low circular orbit until the equator is reached. The second stage is then restarted for the perigee burn, and ends with the secondary engine cutoff-II (SECO-II). The third stage is spun up (without guidance), ignited, and continues until burnout (BO). The spacecraft then separates from the launch vehicle and is in transfer orbit. After a coast of a few days, the apogee motor is fired, and the satellite is in near-geostationary orbit.

The accelerations and vibrations for the different launch events are shown in the figure. The axial acceleration increases during burn, since the thrust of a rocket motor tends to remain constant and the mass is constantly decreasing. Lateral accelera-

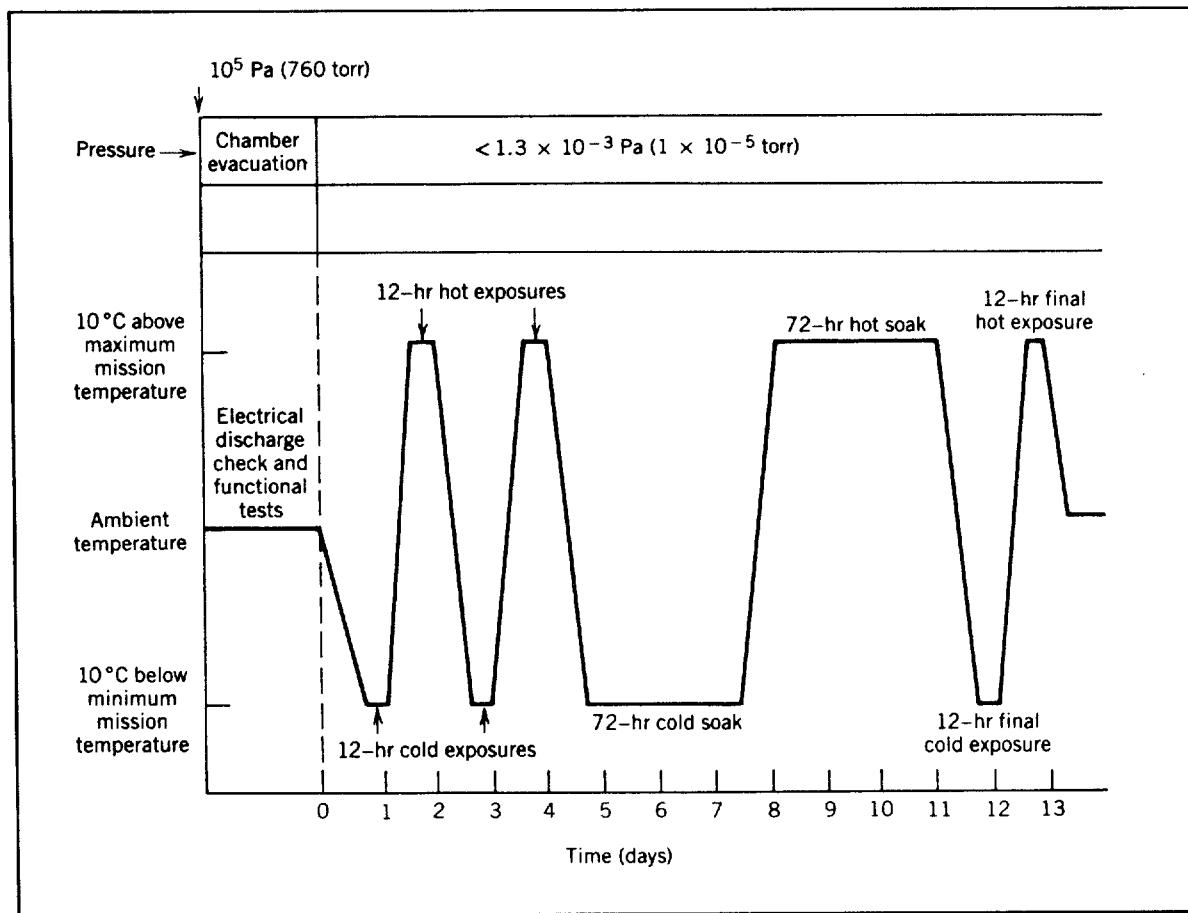


Figure 41 Typical thermal vacuum test cycle (NASA/GSFC, 1977).

tions occur near launch (especially on a windy day) and when the spacecraft is spin stabilized. The vibration levels can be high whenever rockets are firing. The acoustic vibrations are high near lift-off, when there is air to conduct the sound from the motor to the spacecraft. The shocks occur at different points, with one strong shock when the spacecraft separates from the launch vehicle.

To test a spacecraft structure, it is unnecessary to duplicate the entire launch environment. Tests are designed so that, if the spacecraft passes the test, it has a high probability of surviving the launch environment. The main tests are done with an electro-mechanical shaker, that can produce a variety of vibration levels and frequencies. Specific test levels are determined by measuring vibration levels during launch.

7.3 Thermal Vacuum Tests

Environment tests for spacecraft are very important. The closest simulation of the actual space environment is the thermal vacuum test. A typical thermal vacuum test cycle is shown in Figure 41. Cold tests alternate with hot tests for one or more weeks.

A thermal vacuum test has two purposes: to verify the thermal design and to test all the components. It is difficult to accurately predict spacecraft temperatures. Some of these uncertainties may arise from thermal resistance between contacting spacecraft parts, inaccuracies in calculating the heat radiated from one part to another, and the exclusion of many small parts and wires from the mathematical model. Therefore, the first purpose of the thermal vacuum test is to verify the basic thermal design, prove the validity of approximations used in the analysis, and check temperature predictions. This test may be called a thermal balance test and often is done only on the first of a group of satellites. The computer model can often be improved using the test results.

The second purpose of a thermal vacuum test is to provide an environment similar to that encountered in space, and to determine whether all the parts of a spacecraft work in this environment. It is impossible to simulate space exactly, with the extreme vacuum and complex sources of radiation. It is possible to simulate sufficiently to detect most potential problems. Also, it is difficult to simulate the exact time dependence of the environment; it is easier to choose two (or three) extreme conditions. Thus, there would be a "hot" test when maximum temperatures are encountered and a "cold" test with minimum temperatures. This test must be done on every flight spacecraft.

Tests must be long enough for all internal parts to reach their predicted temperatures. A useful tool for monitoring a test is a plot of temperature as a function of $1 - \exp(-t/\tau)$, where τ is the spacecraft time constant. A sample for a small spacecraft is shown in Figure 42. Initial temperature transients due to thermal equilibrium between different parts of the spacecraft tend to disappear sooner. When the temperature is an exponential curve approaching final equilibrium, the points follow a straight line on the graph. As the test proceeds, additional points can be plotted. The equilibrium values can always be determined by a linear extrapolation to the right-hand side (corresponding to an infinite time). This becomes more accurate as closer points are plotted.

7.4 In-Orbit Tests

After the satellite is launched, there is usually an additional period of in-orbit testing. These tests take several weeks and can often be made while the satellite is drifting to its operational longitude or parked at an unused longitude. The main test objectives are:

- To verify deployment of arrays and antennas
- To verify that the satellite system is operating properly
- To verify operation of redundant units
- To provide a baseline as reference for future tests
- To measure those properties that can be evaluated more accurately space, such as actual antenna patterns.

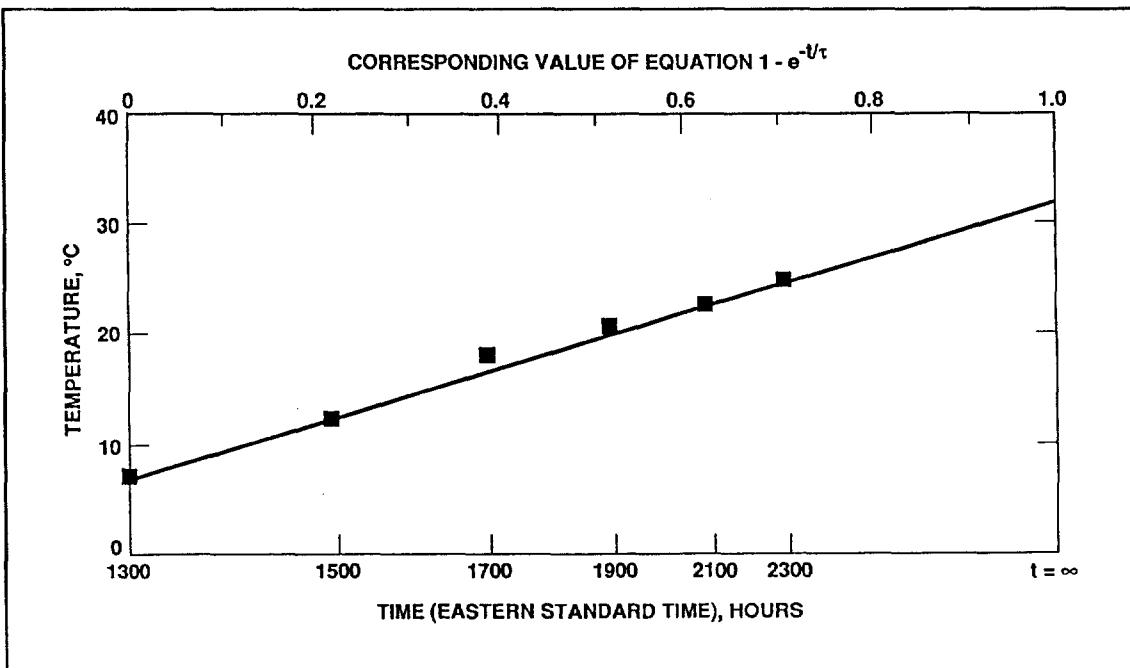


Figure 42 Test temperatures in a thermal vacuum test, extrapolated to equilibrium.

8. Reliability

Once a satellite is launched to geostationary orbit, ordinary repairs cannot be done. So it requires high reliability to perform its mission without the need of repairs. Technically, reliability is the probability that no failure will occur for a given event, or in a given time interval of operation.

There is first a certain probability of a successful launch. From a satellite user's point of view, the launch reliability includes reaching geostationary orbit successfully, and deploying any necessary solar arrays and antenna reflectors. A launch reliability of 0.75 is often used in the initial system design. A few excellent launch vehicles, such as the Atlas Centaur, may have achieved reliabilities of 0.85, but others are less successful. The satellite reliability is the probability that, after a successful launch and deployment, the satellite will complete its mission. A satellite reliability figure of 0.75 is often used as a design goal in building the satellite.

To illustrate some general principles, a reliability curve for human beings is shown in Figure 43. The numbers are actually taken from the 1958 Standard Mortality Table. This graph shows the probability of a new-born infant reaching a certain age. The probability of reaching 5 years of age is almost unity, while the probability of reaching 100 years is quite low.

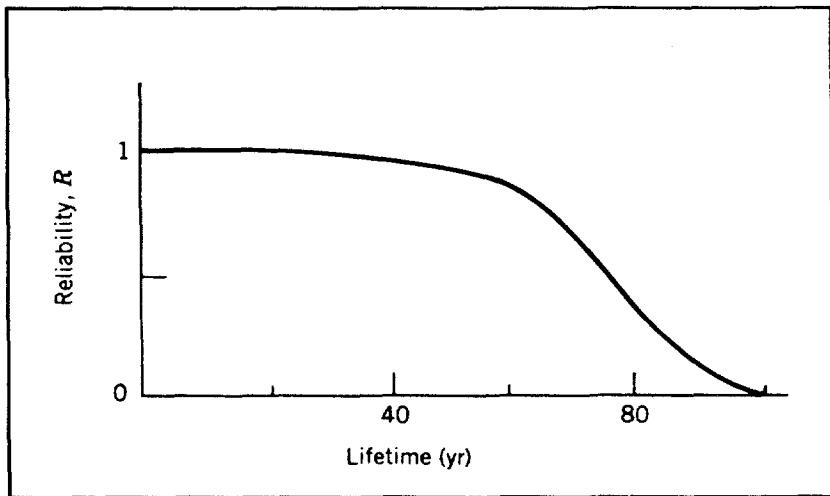


Figure 43 Probability of survival to a given time (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

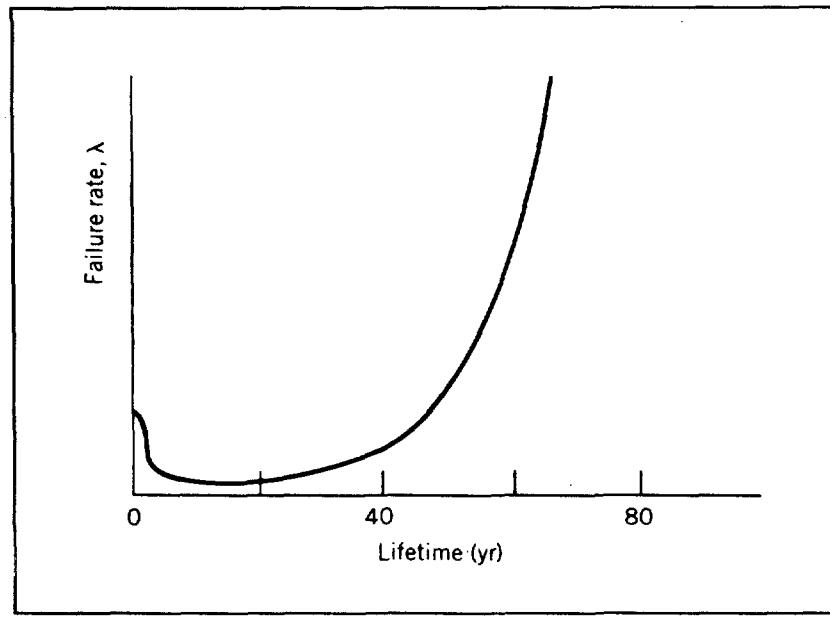


Figure 44 Failure rate of human beings as a function of time (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

The failure rate is defined as the number of failures per unit time divided by the number of units still operating. This is equivalent to the negative of the slope of the reliability curve, divided by the reliability. The corresponding failure rate curve is shown in Figure 44. This shows a low failure rate between the ages of 1 and 40, with higher failure rates at other ages. This is often called a "bathtub" curve, and shows the **early failures** (also called infant mortality), a low constant failure rate, and then an increasing failure rate called **wear-out**. Reliability models are based primarily on low constant failure rates for individual components (also called **random failures**), but other failures cannot be neglected. A fourth type of failure, not shown, is **design failures**. These occur when a mistake is made in the design or manufacture of the item, when all items have a higher failure rate than expected.

8.1 Redundant Components

For many satellite components, it is common to assume a constant failure rate. The failure rate does not decrease with time (as would be true for early failures), nor does it increase

with time (as happens with wear-out failures). A constant failure rate λ leads to an exponential curve for the reliability function

$$R = e^{-\lambda t} \quad (10)$$

The reliability curve R for a constant failure rate is shown in Figure 45, which shows the number of surviving components decreases exponentially.

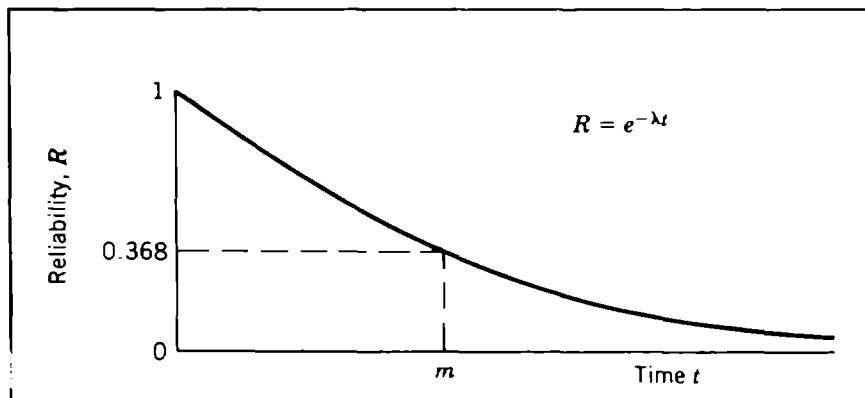


Figure 45 Reliability for a component with a constant failure rate (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

the integration, and a variety of tests to detect weaknesses before launch.

When high reliabilities are required, duplicate or redundant units may be needed. These are connected so that if either unit works, the system will function. In terms of reliability models, these units are considered to be in parallel. (This may or may not correspond to the electrical connections.) Assume there are n spares, and therefore $n + 1$ components in parallel, each with the same constant failure rate. The chance of one failing is one minus the chance of it working. The probability that they all have failed is equal to the product of the probability of each failing. Therefore the reliability of the combination is

$$R_{n+1} = 1 - (1 - e^{-\lambda t})^{n+1} \quad (11)$$

The reliability of a component for a 10-yr spacecraft is shown in Figure 46 for $n = 0, 1$, or 2 . This shows that for a component with 100 FITS (failures in 10^9 hr), the reliability of one component by itself is about 0.99, and for one component with a spare is about 0.9999, a big improvement. This so called 2-for-1 redundancy is provided for many components on a spacecraft. The above equation and figure assumes that the spare has a failure rate equal to the operating unit. In actual practice the failure rate on standby

Most satellite components have backup units, so if one part fails the backup can be used. This is called redundancy. The number of components without redundancy is minimized, since a failure would end the mission. These are called single point of failures. The achievement of high reliabilities requires a good design, careful selection of parts, quality control in

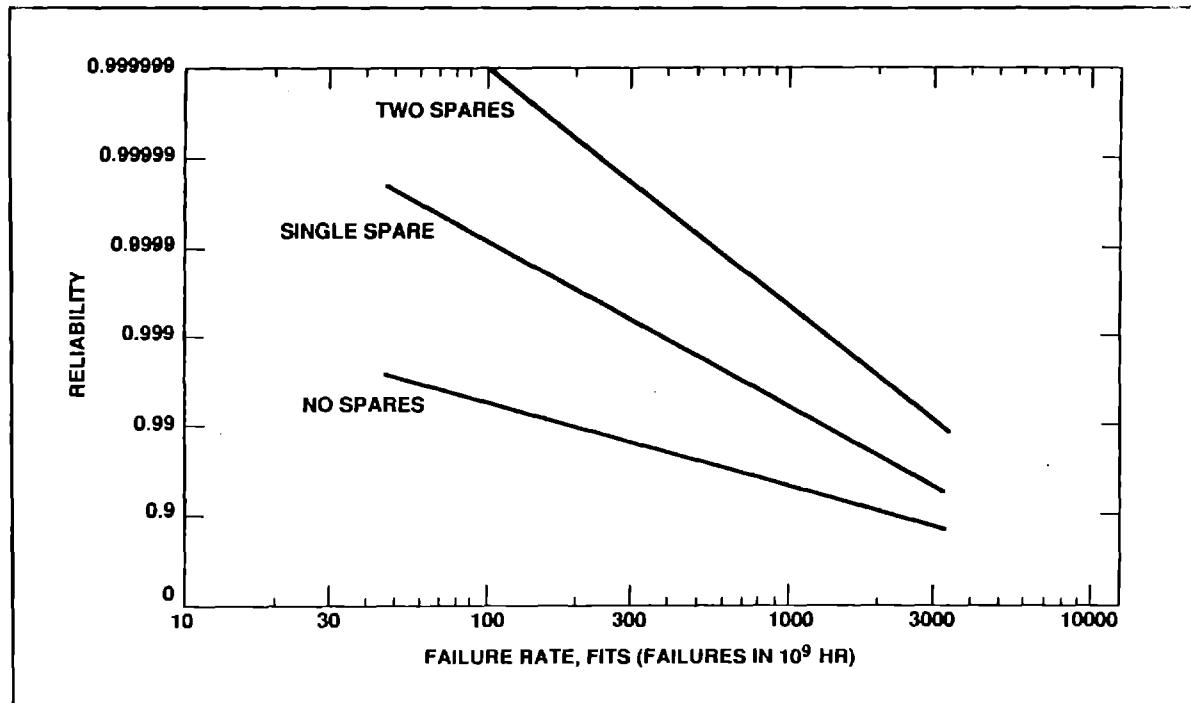


Figure 46 Reliability for components with no spare, one spare, or with two spares, assuming a 10-yr satellite.

units is often less than on the operating unit, and other equations are used to include this.

The preceding formulas assume that failures are independent; that is, a failure in one unit does not affect the failure rate of other units. Frequently this is not true for satellite components. Either a design defect or lack of quality control during manufacture causes a number of similar units to fail. This is portrayed in Figure 47. If there is no redundancy, the reliability of the system is equal to the component reliability ($R_{sys} = R$). If there is twofold redundancy (2 for 1), the system reliability has increased to $R_{sys} = 1 - (1 - R)^2$. If there is fourfold redundancy, the system reliability of $R_{sys} = 1 - (1 - R)^4$ appears to be quite high. The assumption is that any failure is independent of any other failure.

But suppose that when a unit fails there is a 10% chance that all other similar units will fail. The system reliability is then a combination of "no redundancy" 10% of the time and "fourfold redundancy" 90% of the time. This combination is shown in Figure 47 as a dashed line. Redundancy does improve reliability, but not nearly as much as predicted by the simple equations. This effect, called correlated failures, is frequently observed in actual spacecraft. For more than twofold redundancy, it is better to ensure independence by procuring units from different sources.

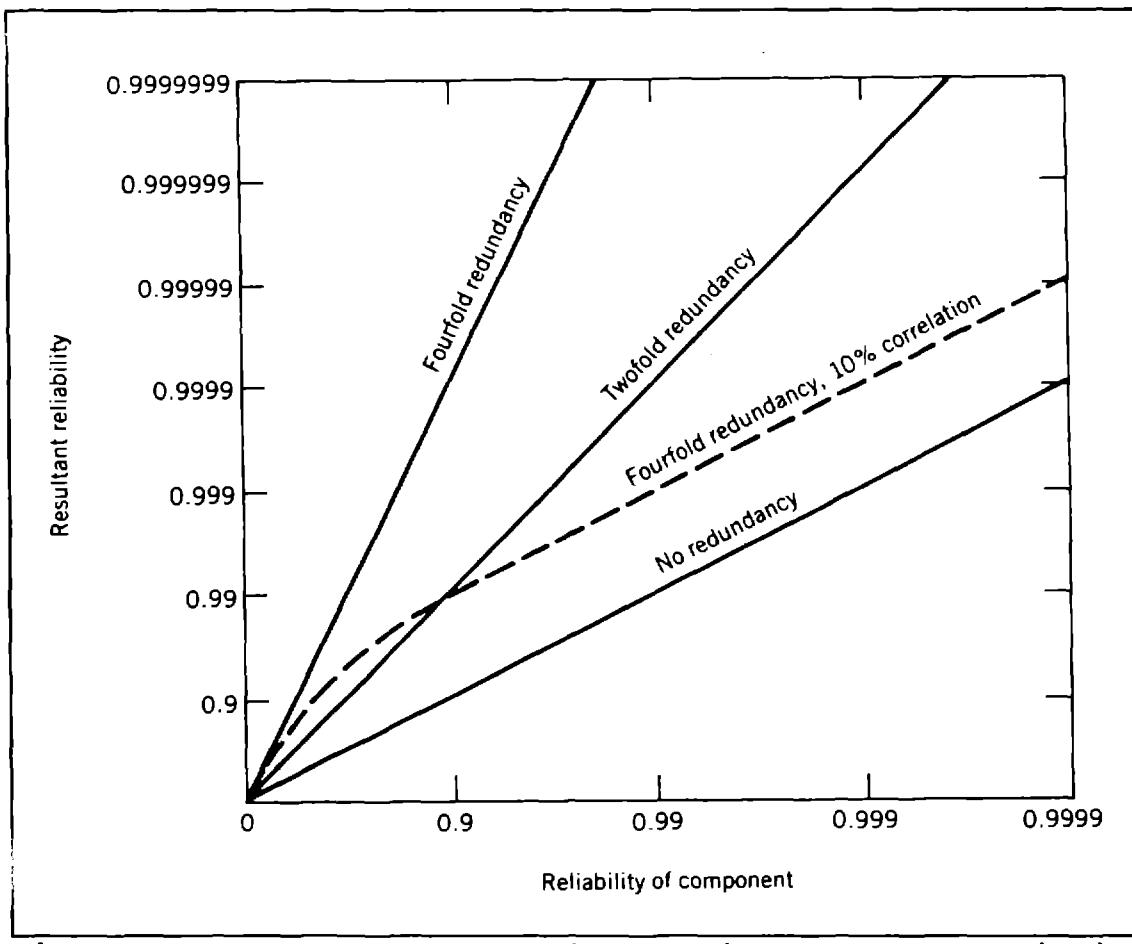


Figure 47 Effect of 10% correlated failures on the reliability of a redundant system (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

8.2 Reliability Model

A mathematical reliability model of a satellite is a useful tool. A model of the spin-stabilized INTELSAT IV satellite series was formulated by J. B. Binckes (personal communication, 1970). The basic criterion for success was that nine of the twelve transponders were operating.

The satellite was divided into six major subsystems, as shown in Figure 48. The subsystems are: communications, telemetry, command, propulsion, attitude control, and electric power. Each subsystem is necessary for the functioning of the satellite, so the satellite reliability is

$$R_{sys} = R_{comm} R_{tm} R_{cmd} R_{prop} R_{att} R_{power} \quad (12)$$

A separate reliability diagram was developed for each subsystem.

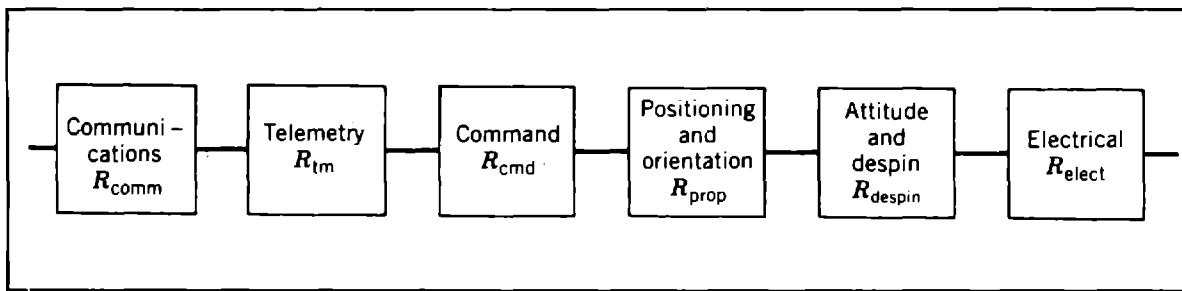


Figure 48 Reliability model of a satellite with six subsystems (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

The telemetry subsystem, shown in Figure 49, is fully redundant. It consists of two separate chains, and every component in one of the two chains must operate. To find the reliability of the telemetry subsystem, the reliability of each component is first calculated. This is done using the failure rates given in the figure. In each chain the components are in series, so these reliabilities are multiplied. The two chains are in parallel, so the unreliabilities must be multiplied. The result is

$$R_{tm} = 1 - (1 - R_1 R_2 R_3 R_4)^2 \quad (13)$$

and a reliability of 0.96 is found for the telemetry subsystem.

The command subsystem is shown in Figure 50. It is fully redundant, except for the antenna which has a very low failure rate (1 FIT). Each operating unit has a spare, so there is 2 for 1 redundancy. There is also cross-strapping, so that spare units can be used in any combination.

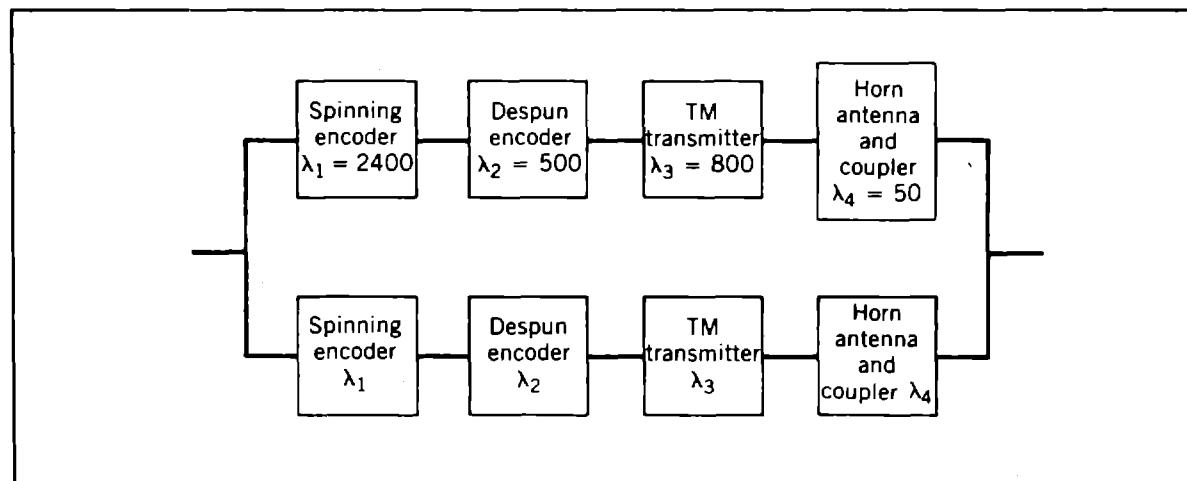


Figure 49 Reliability model of a telemetry subsystem (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

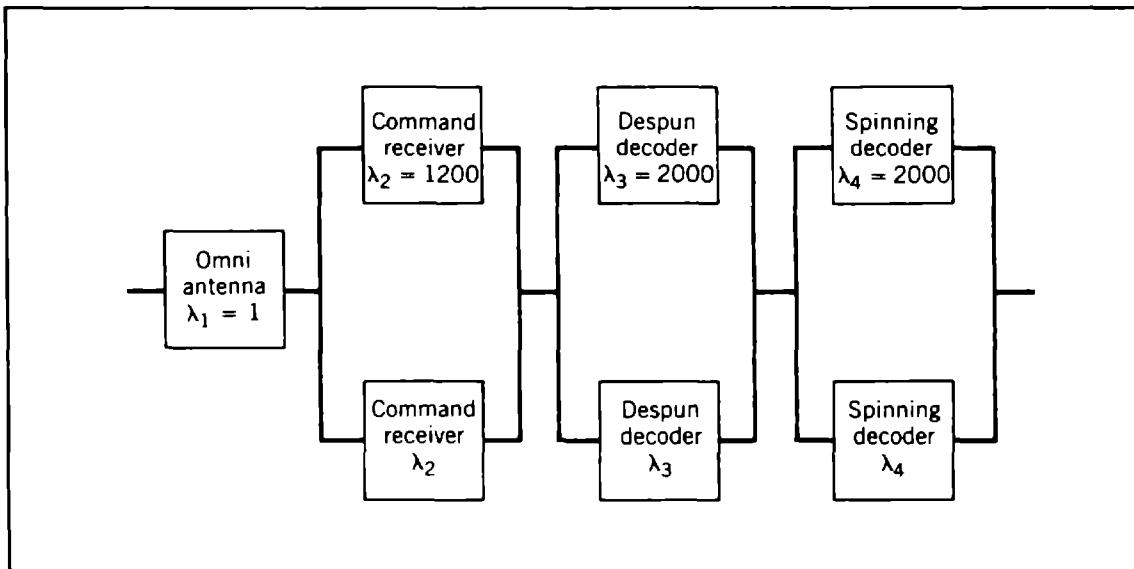


Figure 50 Reliability model of a command subsystem (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

To find the reliability of the command subsystem, the reliability of each component is first calculated by using $R = \exp(-\lambda t)$. The two command receivers are in parallel, so unreliabilities (probability that each fails) must be multiplied. The reliability of the combination is $1 - (1 - R_2)^2$. The reliabilities of the despun decoders and of the spinning decoders are found in a similar manner. The four sections are in series, so the reliabilities must be multiplied. The final result is

$$R_{cmd} = R_1[1 - (1 - R_2)^2][1 - (1 - R_3)^2][1 - (1 - R_4)^2] \quad (14)$$

For a 7-yr satellite, with the failure rates shown in the figure, the reliability is 0.97 for the command subsystem.

The reliability of each of the other subsystems was found in a similar fashion. Some of them had more components and more complicated logic diagrams. But the components were either in parallel or in series, and could be combined with similar equations.

The results were combined to obtain the reliability of the entire satellite. The result is shown as the dashed line in Figure 51. This shows the probability of satellite survival in orbit for different times. The predicted reliability is about 0.6 for the design lifetime of 7 years.

This reliability model was formulated in 1970, almost a year before the first satellite launch. The performance of the INTELSAT IV series of satellites is well documented. Eight satellites were launched, and six were successful, for a launch reliability of

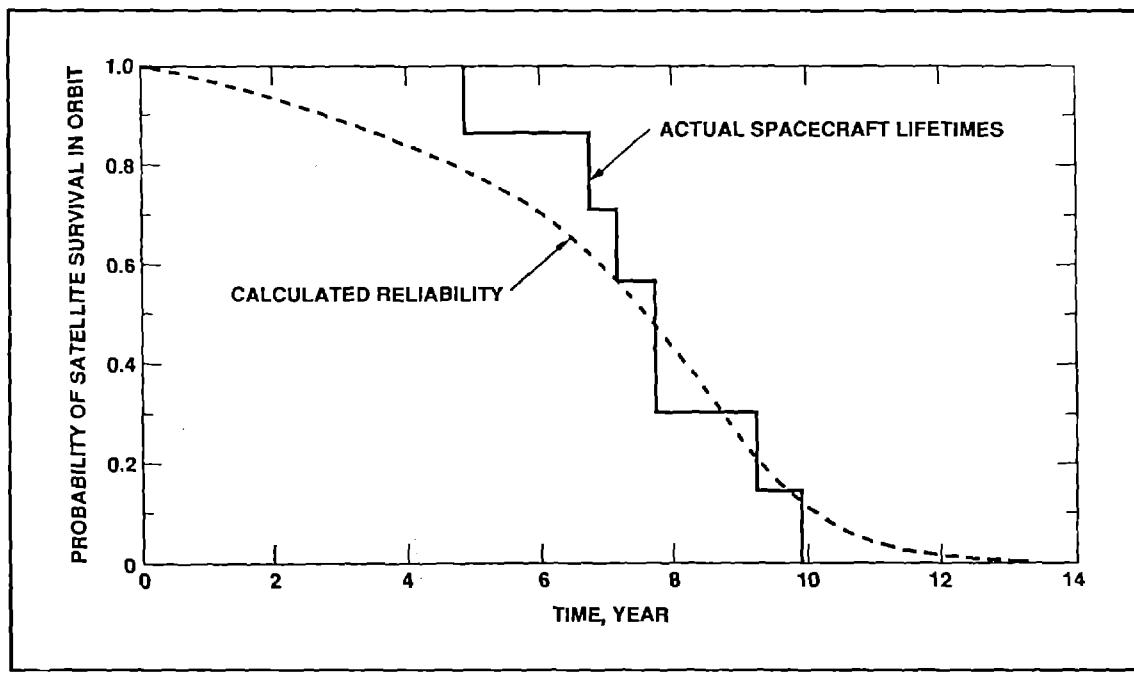


Figure 51 Survival probability of INTELSAT IV satellite in orbit, after initial injection.

0.75. The actual lifetimes of the six operating satellites is shown as the solid line in Figure 51. The agreement is excellent, and deviations tended to be on the good side.

A comparison can also be made on individual components. Table 6 summarizes the predicted number of failures of various components, and the actual number observed. While these comparisons are difficult to make, there is a correlation between the reliability model and the results obtained.

Many spacecraft are designed with two-fold redundancy for most major components. Single points of failure (no redundancy) are minimized. Transponders may only have one spare for every two transponders (3 for 2 redundancy). It is possible to vary the model to maximize the reliability for a given mass (or a given cost). Starting with a spacecraft with no redundancy, spares are added one by one, choosing the unit that can increase the reliability the most for a given mass increase. Figure 52 shows the result of one study (H. Meyerhoff, personal communication, 1982), which did not include the communications subsystem. Each dot represents a possible reliability model. When a heavy spare is added there is a larger gap in the curve, while a light spare just adds a little mass. The curve tends to be smooth, without any sharp bends. Such a study provides some insight into the satellite reliability, but it should be used with caution, since the reliability of each component may have some uncertainty.

Number per Satellite	Component	Failure Rate, FITS	Failures	
			Predicted	Actual
24	Travelling-wave tube amplifier	4000	58.3	36
4	Receiver, pre-amplifier	6000	13.8	34
2	Attitude control electronics	5200	6.1	3
2	Command decoder	4000	5.4	0
12	Transponder switches and filters	500	4.1	0
2	Telemetry encoder	2900	3.9	2
2	Thruster (set of three)	2100	2.8	4
2	Power amplifier (attitude control)	1500	2.0	0
2	Earth sensor	1500	2.0	3
2	Command receiver	1200	1.6	0
2	Sun sensor	1000	1.3	0
2	Hydrazine tank (pair)	800	1.1	0
2	Telemetry transmitter	800	1.1	1
2	Battery and control	600	0.8	6
2	Spot beam	500	0.7	0

^aBased on 78.5 operating years

Table 6 Failures of spacecraft components, predicted and actual flight failures (Copyright, Morgan and Gordon, 1989. Reprinted by permission of John Wiley & Sons, Inc.).

REFERENCES

This paper is based on the author's recent textbook (Gordon and Morgan, 1993), and many of the figures are from it. Most of these figures are also in the author's handbook (Morgan and Gordon, 1989). Information on the geostationary orbit is in Pocha, 1987. For reliability, Bazovsky (1961) is easy to read, and emphasizes practical concepts.

Allison, J. F., R. Arndt, and A. Meulenberg (1975), "A Comparison of the COMSAT Violet and Nonreflective Solar Cells." *COMSAT Tech. Rev.* 5 (2), 211-223.

Bazovsky, I. (1961), *Reliability Theory and Practice*, Prentice-Hall, Englewood Cliffs, NJ.

Gordon, G. D., and W. L. Morgan (1993), *Principles of Communications Satellites*, Wiley, NY.

Gupta, R. K., J. N. Narayanan, A. M. Nakamura, F. T. Assal, and B. Gibson (1991), "INTELSAT VI On-board SS-TDMA Subsystem Design and Performance." *COMSAT Tech. Rev.* 21 (1), 191-225.

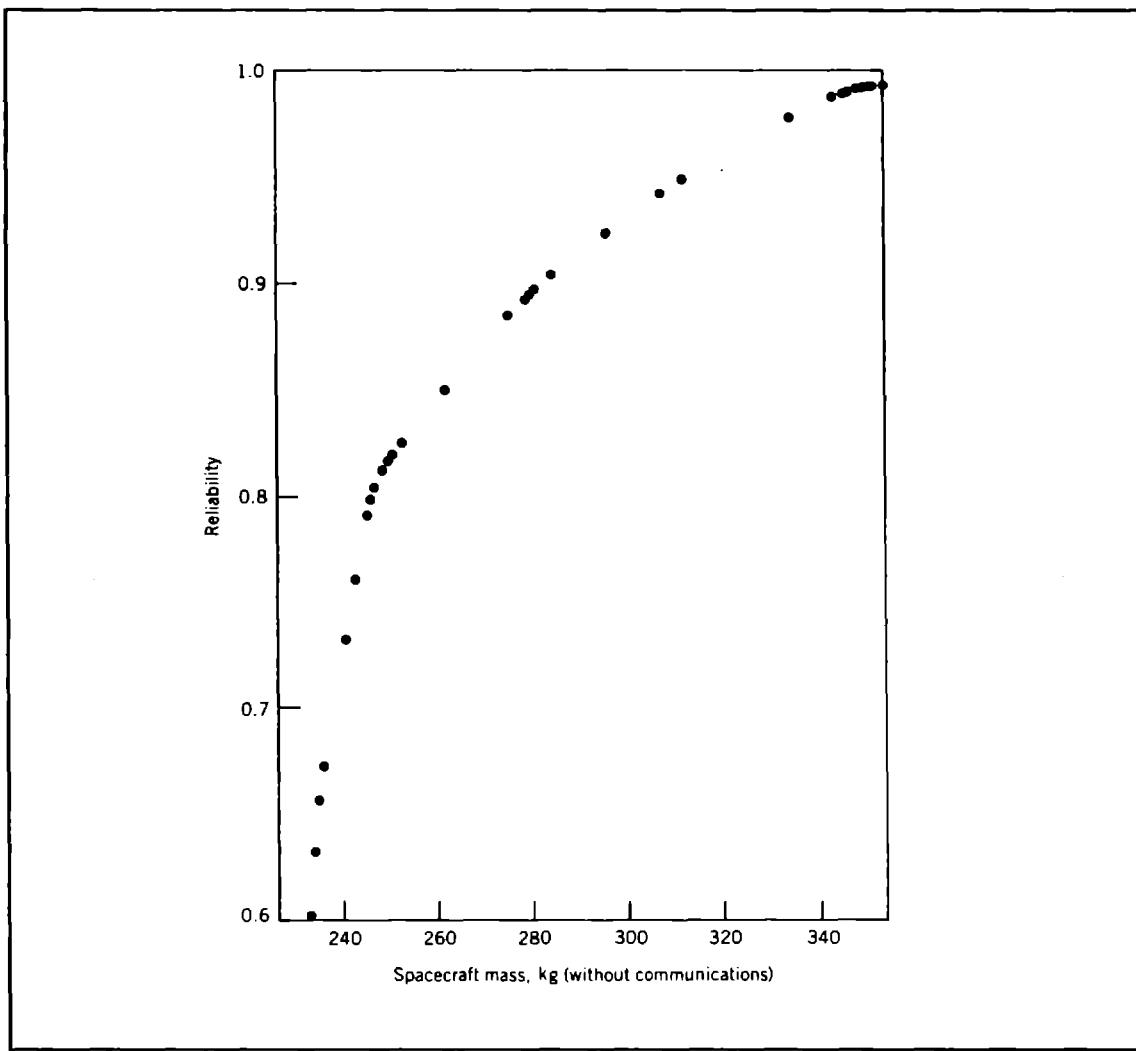


Figure 52 Reliability versus mass for a proposed spacecraft, excluding the communications subsystem.

Morgan, W. L., and G. D. Gordon (1989), **Communications Satellite Handbook**, Wiley, NY.

Pocha, J. J. (1987), **An Introduction to Mission Design for Geostationary Satellites**, D. Reidel Publishing Co., Boston, MA.

Chapter 2

Radiation Environments within Satellites

Henry Garrett
Ballistic Missile Defense Organization
Washington, D.C. 20301

Radiation Environments within Satellites

by

Henry B. Garrett

P. Rustan

S. P. Worden

S. Nozette

Ballistic Missile Defense Organization

Department of Defense

Washington, D.C.

2.1. INTRODUCTION

2.1.1. OBJECTIVES

As electronic components have grown smaller in size and power and have increased in complexity, their enhanced sensitivity to the space radiation environment and its effects has become a major source of concern for the spacecraft engineer. As a result, the description of the sources of space radiation, the determination of how that radiation propagates through material, and, ultimately, how radiation affects specific circuit components are primary considerations in the design of modern spacecraft. The objective of this paper will be to address these 3 aspects of the radiation problem. This will be accomplished by first reviewing the natural and man-made space radiation environments. These environments include both the particulate and, where applicable, the electromagnetic (i.e., photon) environment. As the "ambient" environment is typically only relevant to the outer surface of a space vehicle, it will be necessary to treat the propagation of the external environment through the complex surrounding structures to the point inside the spacecraft where knowledge of the internal radiation environment is required. While it will not be possible to treat in detail all aspects of the problem of the radiation environment within a spacecraft, by dividing the problem into these 3 parts-external environment, propagation, and internal environment-a basis for understanding the practical process of protecting a spacecraft from radiation will be established. The consequences of this environment will be discussed by the other presenters at this seminar.

2.1.2. BASIC CONCEPTS

This section will provide an overview of the basic physical concepts and definitions that will be used throughout the presentation. In particular, the concepts of energy, flux, fluence, and dosage will be briefly described. The reader is referred to the many excellent texts on space physics or astronomy for more detailed explanations^[1, 2].

Consider first the concept of energy. In the case of particles that have a rest mass, the fundamental equation relating particle mass and velocity to energy is:

$$E = 1/2 m v^2 \quad (1)$$

where: m = particle mass
 V = particle velocity
 E = particle kinetic energy

For photons (which have no rest mass), the equivalent equation is:

$$E = h v \quad (2)$$

where:

$h =$	Planck's constant
$v =$	frequency of the light

The fundamental units of energy E to be used in this study will be the erg, the joule, and electron volt. The erg is the basic unit of energy in the CGS system. It has units of $\text{g}\cdot\text{cm}^2\cdot\text{s}^{-2}$. The joule is the fundamental unit of energy in the MKS system (or "System Internationale") and is in units of $1 \text{ kg}\cdot\text{m}^2\cdot\text{s}^{-2}$. Thus, $1 \text{ erg} = 10^{-7} \text{ joules}$ (in MKS units). The electron volt or eV is also a measure of energy such that $1 \text{ eV} = 1.602 \times 10^{-12} \text{ ergs} = 1.602 \times 10^{-19} \text{ joules}$. It is the kinetic energy that an electron will acquire by being accelerated through an electric potential of one volt.

Closely coupled to the concept of energy is that of dose. Simply put, dose is the total energy accumulated in a given volume element of a specific material due to the incident radiation. It is typically given in units of rads or "radiation absorbed dose" for a particular material (the material must be specified because energy absorption is dependent on the material). As an example, for silicon, $1 \text{ rad (Si)} = 100 \text{ ergs/g (Si)}$. The corresponding unit for dose in the MKS system is called the gray. $1 \text{ Gy (gray)} = 1 \text{ joule/kg} = 100 \text{ rad} = 10^4 \text{ ergs/g}$. Here we will, without exception, use the older unit "rad" as this is still the "preferred" unit in space physics and radiation effects analysis. It needs to be emphasized that, for the same incident flux, different materials will be affected differently dependent on the composition of the incoming radiation and the composition of the absorbing material. Other units such as the roentgen (quantity of γ or X-rays that deposit, by ionization and energy absorption, 83 ergs/g in dry air) or the rem (roentgen equivalent per man) are also often used but will not be discussed here.

In addition to the energy and composition of a particle or photon, it is also necessary to describe how many of them there are. This is usually done in terms of intensity or flux and, when speaking in terms of a time interval, fluence. Confusion arises over the concepts of intensity/flux and fluence because there are many different ways to define these quantities. Here we will define the quantity "unidirectional differential intensity" $j(E, \theta, \phi, t)$ as:

"The flux (number of particles or photons per unit time) of a given energy E per unit energy interval dE in a unit solid angle ($d\Omega = 2\pi \cos \theta d\theta d\phi$) about the direction of observation (in the θ, ϕ direction), incident on unit of surface area (dA) perpendicular to the direction of observation."

This is illustrated in Fig. 1^[2]. Typical units are particles-cm⁻²-s⁻¹-sr⁻¹-KeV⁻¹ for protons or electrons and particles-m⁻²-s⁻¹-sr⁻¹-MeV⁻¹- μ ⁻¹ for heavy ions (where μ is "nucleon")—a typical spectrum for iron cosmic rays is presented in Fig. 2^[3]. The "unidirectional integral intensity" (or flux) is defined as the intensity of all particles with energy greater than or equal to a threshold energy E:

$$j_{\geq E} = \int_E^{\infty} j dE \quad (3)$$

with units of particles $\text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}$

We define the "omnidirectional flux" J as:

$$J = \int_{4\pi} jd\Omega \quad (4)$$

Fluence I is the integral of the flux over a given time interval (i.e., one hour, one year, etc.):

$$I = \int_{\delta t} jdt \quad (5)$$

Here, when we refer to omnidirectional fluence $I(\geq E)$, we will normally mean the "omnidirectional integral (in energy) fluence" such that:

$$I_{>E} = \int_E^{\infty} dE \int_{4\pi} d\Omega \int_{\delta t} jdt \quad (6)$$

The units of this quantity are particles- cm^{-2} for some specified threshold energy E (typically 1 MeV or higher for radiation effects) and for a specified time interval (often 1 year).

To allow comparisons between different energies, particle types, and dosages, it is common practice to talk in terms of "1-MeV equivalent" (typically 1-MeV electrons or 1-MeV neutrons in silicon). First, the energy dependence of the damage and energy content of the spectra for the environment to be considered are used to determine what fluence of 1-MeV particles (electrons or neutrons) would produce the same amount of damage or dose in the material (typically silicon or aluminum). A curve for neutrons, in units of MeV-mb (where b is a barn or 10^{-24} cm^2 and the "relative displacement damage" is defined in terms of the cross section times the energy of the incident particle), is given in Figure 3[4]. As an illustration, for 14 MeV neutrons, the 1-MeV neutron dose damage equivalent is given by multiplying the 14 MeV dose by 2.5 (obtained from Fig. 3). (Note: because of variations in the damage parameter with material and property, it should always be kept in mind that the use of a damage equivalent is not exact but an approximation for comparison purposes.)

A final quantity related to energy absorption and flux to be used here is the Linear Energy Transfer or LET. The LET is the energy transferred by radiation per unit length of absorbing material. That is $\text{LET} = dE/dx$. For ionization and excitation effects, it is often expressed in MeV/ μm of the primary particle track length or, if the density of the material is known, MeV- $\text{cm}^2\text{-mg}^{-1}$ (this is typically the "unit" when the reference is to an LET between 1 and 30 and is given by: $\frac{1}{\rho} \frac{dE}{dx}$).

The concept of LET is particularly important when discussing single event upsets (SEU) or "soft errors". These occur when a particle, typically an ionized, high energy atomic nucleus, deposits enough energy in the sensitive region of an electronic device to cause a change in the logic state of the device. Upsets occur only when the energy

deposited exceeds a critical level in the sensitive region of the device. This is often computed in terms of LET. When viewed as a function of LET, the probability of upset is, in its simplest form, a threshold phenomenon: any particle with a minimum LET (L_O) or greater will cause an upset. This behavior is illustrated in Fig. 4 where the energy deposited per unit length (LET) is plotted versus incident particle energy—note how the curve has a peak rate. A useful way of presenting the environment in terms of LET is the Heinrich curve. The Heinrich curve gives the integral flux as a function of LET rather than particle energy. The Heinrich flux F_H is the flux of particles for a single species with a (threshold) LET of LET_0 or greater:

$$F_H(LET_0) = \int_{E_1}^{E_2} f_i(E)dE \quad (7)$$

where: f is the particle flux for the species as a function of energy, and E_1 and E_2 are the energies between which the LET is greater than or equal to the threshold LET_0 (a representative integral Heinrich curve for iron is plotted in Fig. 5^[5]). The LET depends not only on particle energy but the target material as well as the LET versus energy curve will be different for all particle species. Experiments have shown, however, that to first order it is the LET which is important for determining upsets and not the particle energy or its species. The Heinrich flux versus LET plot is the principal means of presenting radiation data for use in SEU calculations just as the particle flux versus energy is the main means of presenting radiation data for dosage calculations.

To summarize, this section has defined the basic terminology used to describe the radiation environment. "Dose", "Flux/Intensity", "Fluence", "LET", and "1-MeV Equivalent" have been defined. The reader is referred to books and articles by Roederer^[2], Smart and Shea^[6], and others for more complete descriptions of these concepts.

2.2. SPACE ENVIRONMENTS

2.2.1. AMBIENT PARTICULATE ENVIRONMENTS

2.2.1.1. OVERVIEW

The high energy particle radiation environment consists of electrons with energies greater than 100 KeV, protons or neutrons with energies greater than 1 MeV, and heavy ions with energies above 1 MeV/nucleon. The populations are characterized in terms of their kinetic energy, charge state (or lack thereof), and composition. Unlike photons which travel uniformly at the speed of light, particles can vary in velocity from a few m/s up to a sizable fraction of the speed of light in the case of cosmic rays. The discussion will be roughly divided into four families of radiation based on these characteristics:

- 1) Trapped radiation (for the Earth, the Van Allen Belts);
- 2) Galactic Cosmic Rays which consist of interplanetary protons, electrons, and ionized heavy nuclei;
- 3) Protons and other heavy nuclei associated with solar proton events.
- 4) Neutrons (primarily Cosmic Ray Albedo Neutrons or CRAN particles).

The first two sources are relatively constant or change on long time scales whereas the third is highly time dependent. The fourth is a secondary population because the

relatively short lifetime of neutrons severely limits any solar produced fluxes at 1 AU or beyond. Each type of radiation has a characteristic spectrum and preferred interaction mode with matter that supports this simple division.

2.2.1.2. TRAPPED RADIATION

First discovered by J. Van Allen and his collaborators on Explorer I, trapped radiation at the Earth consists principally of energetic protons and electrons, with lesser percentages of heavy ions such as O⁺, contained in toroidal belts by the Earth's magnetic field. This toroid is commonly known as the "Van Allen belt(s)"^[7] and consists of (at least) two zones: a low altitude zone, or "inner belt"; and a high altitude zone, or "outer belt". The inner belt extends from ~100s of km to ~ 6,000 km in altitude and is populated by high-energy (~10s of MeV) protons and high energy (1-10 MeV) electrons, while the outer belt, up to 60,000 km in altitude, is predominately high energy electrons. Schematics of the radiation flux contours for the Van Allen belts are illustrated in Fig. 6^[8] (note the difference in scale between the electrons and protons). The detailed mechanism by which particles are entrapped in the belt regions is not well understood nor is the primary source clearly identified (as will be discussed, albedo neutrons are considered an important source of the intense proton and electron fluxes in the inner belt while the outer belt may be primarily due to entrapment of low-energy solar wind plasma by the geomagnetic field followed by local acceleration)-observations of abundance ratios imply both terrestrial and interplanetary sources. Once captured, the motion of charged particles in the Earth's magnetic field is governed by Lorentz force (see later). The trapped radiation environment also exhibits large temporal variations. The inner belt zone, because of the dominance of the Earth's main field, is relatively stable. Most temporal variations in this population occur as the solar cycle proceeds and the Earth's neutral atmospheric density at a given altitude changes causing variations in the altitude at which radiation particles can mirror without being scattered. In contrast, the outer belt, which is more influenced by the Earth's highly variable geomagnetic tail, experiences much greater temporal fluctuations. The electron concentration in the outer zone may experience temporal fluctuations as large as a factor of 1000. Fortunately, most of the physical damage caused by the trapped radiation is largely attributable to the long-term cumulative (or integral) dose received by the spacecraft rather than the instantaneous fluctuations of the radiation.

In this section, to better understand the preceding description, we first discuss the primary characteristics of the Earth's magnetic field as this is the dominate force controlling the formation and changes associated with these trapped radiation belts. Following this discussion, a brief review of the basic concepts of particle entrapment such as gyro radius, mirror point, and pitch angle will be presented. Although the details of these concepts are not critical to an understanding of the basic effects of radiation on spacecraft, they are critical to understanding how the radiation belts are modeled and how one goes about determining the external dose on a particular vehicle. In particular, the concept of adiabatic invariants will be summarized-this concept is critical to an understanding of why we use B-L coordinates, the basis of almost all modern radiation models, in describing the radiation belts. Following these descriptions, several examples of the trapped radiation environment will be presented in terms of the AE8/AP8 radiation models and the predictions compared with actual data. As will become evident, given the rapid temporal variations observed in the space environment, models are only reliable in a long term predictive sense-this is a crucial issue in establishing design margins for spacecraft.

2.2.1.2.1. GEOMAGNETIC FIELD

Above 1000 km, the dominant geophysical environment of interest is the magnetic field of the Earth—the source of the Earth's magnetosphere. Below 1000 km, the Earth's magnetic field, primarily through the control of the ionospheric plasma, plays an important though reduced role in the dynamics of the natural environment. The Earth's geomagnetic field, B , is composed of three distinct components (or current systems): the "core" field, the "crustal" field, and the "external" field^[9]. The core field is a steady field due primarily to the convective motion of the conducting fluid in the Earth's internal core. The distribution of the crustal field, which is the remnant or induced magnetization of ferromagnetic materials near the Earth's surface, results in surface magnetic anomalies associated with the geologic and tectonic features of the crust. The core and crustal fields are usually combined and referred to as the "internal" or "main" field, B_i . B_i varies slowly on the order of a hundred years—currently ~.05% per year—as illustrated in Fig. 7^[9] (be aware that while it is critical to use a magnetic field model appropriate for the year that the magnetic field estimates are required, it is typical in radiation modeling to use the magnetic field model for the year that the radiation belt model was derived—this has caused serious problems in the past). B_i dominates at LEO and accounts for more than ~99% of B even during extremely large geomagnetic storms (massive variations in the Earth's magnetosphere brought about by changes in the solar environment that encompasses the Earth's magnetic field). In contrast, the external field, B_e , which makes up ~1% of the field at LEO, is due primarily to extra-terrestrial sources—primarily the Earth's ring current and the solar wind. It varies rapidly in time (from on the order of milliseconds to periods as long as the 11 year solar cycle) and is closely correlated with geomagnetic activity and solar interactions. Models of the external component of the geomagnetic field are available but are of limited importance to the trapped radiation belts.^[10-12]

Seven basic quantities, called "magnetic elements", are used to specify the geomagnetic field^[9]. Their definitions and relations are given in Fig. 8. Three independent quantities (e.g., (H, D, Z) or (X, Y, Z)) are required to uniquely define B . For spacecraft operating in the radiation belts, the most convenient system to use is either the geographic (also known as geocentric) or the geomagnetic coordinate system. These systems, based on spherical coordinates, are schematically defined in Fig. 9. Geographic coordinates correspond to a geocentric longitude/latitude system based on the Greenwich Prime Meridian. Geomagnetic coordinates are similarly Earth centered, but have the north pole of the system passing near the geomagnetic pole, which is offset from the geographic pole, at ~78.5° latitude and ~291.1°E. Geomagnetic longitude is measured from the great circle passing through the geographic and geomagnetic poles. Details about other coordinate systems can be found in Knecht and Shuman^[9].

Aside from the gravitational field of the Earth, the magnetic field due to the internal geomagnetic field is the most accurately known of the natural environments. It can be crudely modeled in terms of a tilted (-11° from geographic north) magnetic dipole of magnitude $M = 8 \times 10^{25}$ G-cm³ (G is the magnetic unit Gauss). Ignoring the tilt for the moment, in the geomagnetic coordinate system, the magnetic field intensity induced by M at the point (r, θ, ϕ) is given by the expression:

$$B_i = -(M/r^3) \cdot (3 \cos(\theta) + 1)^{0.5} \quad (8)$$

In the Gaussian unit system, r is in cm, and B_i is in G. Given the above value for M , B_i is then found to have a maximum value of ~0.6 G near the polar cap and a minimum value of ~0.3 G near the equator at the Earth's surface.

Eq. 8 is valid only for an idealized configuration of a centered dipole. In reality, large scale discrepancies (as high as $\pm 25\%$) exist between the measured data and the ideal, dipole expression. Modifying the configuration from a centered dipole to an eccentric dipole reduces the discrepancies to the ~10% level, but this is still unacceptably large. For most purposes, the IGRF series of models is the official standard. The latest version of the International Geomagnetic Reference Field, IGRF-87, is a computer model based on numerical fitting of measured data with a magnetic scalar potential expanded in terms of ten spherical harmonics^[13]. The model calculates the seven magnetic elements of B_i for any given geographical location. Specifically, a scalar "potential" is found such that:

$$V(r, \theta, \phi) = a \sum_{n=1}^N \sum_{m=0}^n \left\{ [A_{n,m} \cos m\phi + B_{n,m} \sin m\phi] \left(\frac{a}{r} \right)^{n+1} + [C_{n,m} \cos m\phi + D_{n,m} \sin m\phi] \left(\frac{r}{a} \right)^n \right\} P_{n,m}(\cos \theta) \quad (9)$$

where:

- a = radius of Earth
- r = radial distance in units of a
- N = order of expansion
- θ = colatitude
- ϕ = east longitude
- $A_{n,m}, B_{n,m}$ = constants for internal terms
- $C_{n,m}, D_{n,m}$ = constants for external terms

The magnetic field components are then given by:

$$\begin{aligned} B_r &= \frac{\partial V}{\partial r} \\ B_\theta &= -\frac{1}{r} \frac{\partial V}{\partial \theta} \\ B_\phi &= -\frac{1}{r \sin \theta} \frac{\partial V}{\partial \phi} \end{aligned} \quad (10)$$

Values of A, B, C, and D are presented in Table 1 for the IGRF 1980.0 model. As discussed earlier, the epoch of the magnetic model is important and should correspond either to the same date as the date of the radiation model used or to a model epoch as close to the present as possible if only the current magnetic field value is desired (i.e., IGRF 1987 for 1993).

Figs. 10a and 10b show the 3 dimensional character of the Earth's magnetic field. Fig. 10a is a cross section of the Earth's magnetic field in the noon-midnight meridian showing the structure of the field lines and the plasma regions they contain. Fig. 10b

illustrates typical results from calculations of the magnetic field amplitude over the northern hemisphere at a constant altitude of 400 km. The field amplitude varies from a minimum of 0.25 G near the equator to 0.5 G over the polar caps. Two peaks exist in the magnitude of the magnetic field over the north pole (if vector components are considered, the maximum at 270° east longitude is the true "dip" magnetic pole). Likewise, there are two minima near the equator—the largest of these is responsible for the so-called South Atlantic Anomaly, a region critical in determining radiation exposure in LEO. Finally, it should be noted that geomagnetic storm variations are superimposed on this main field. These are typically less than 0.01 G so that even during a severe geomagnetic storm, magnetic fluctuations are small at low altitudes compared to the average field (even though this is a very small change in the Earth's field, the effect of geomagnetic storms on particle fluxes in the polar ionosphere can be tremendous). They are critical, however, at geosynchronous orbit where the main magnetic field is of comparable strength and dominate the magnetic field at higher altitudes.

2.2.1.2.2. MAGNETIC AND ELECTRIC FIELD EFFECTS

To understand the trapped radiation environment, one must first understand how charged particles move in the Earth's magnetic field. This section provides a very brief overview of the fundamentals of charged particle motion so that the concepts used in modeling the radiation belts can be comprehended. The reader is referred to many excellent texts in this area such as the book by Roederer^[2].

2.2.1.2.2.1. BASIC PARTICLE MOTION

To understand how charged particles become "trapped", it is necessary to review the forces that act on those particles. Except near the upper fringes of the Earth's atmosphere (<100-150 km), collisional and frictional forces on charged particles can, in general, be ignored. The two main forces (F_E and F_M) are then (units of cm-g-s, CGS, are used unless otherwise noted) the electrostatic force:

$$F_E = qE \quad (11)$$

and the magnetic (Lorentz) force:

$$F_M = \frac{q}{c} V \times B \quad (12)$$

where q is the particle charge (including sign), c is the speed of light, V the velocity vector of the particle, B the magnetic field vector in space, and E the electric field vector in space.

Consider the actual motion of a particle subject to equations (11) and (12). Setting the electric field to zero and using the definition of the cross product, equation (12) implies that the force on a charged particle is always perpendicular to both the particle's instantaneous velocity vector and the magnetic field vector. This means that a particle must, in the absence of another force and in the presence of a uniform magnetic field, move in a circle in the plane perpendicular to the magnetic field vector. It may additionally move freely (without any acceleration) along the magnetic field, mapping out a helix around its "center of motion" (Figs. 11, 12, and 13). The radius R_C (called the cyclotron or gyro radius) of this circle is found by equating the centripetal force, mV^2/R ,

to the Lorentz force. In this expression m is the particle mass and V_{\perp} is the component of the velocity perpendicular to B . The expression is:

$$R_c = \frac{mV_{\perp}c}{qB} \quad (13)$$

The frequency with which the charged particle gyrates, the cyclotron frequency, ω_c , is given by:

$$\omega_c = \frac{qB}{mc} \quad (14)$$

where ω_c is in radians per second.

According to equation (12), any particle motion parallel to B is unaffected by B . The particle's motion can be described in terms of a velocity parallel to the field, V_{\parallel} , or perpendicular to the field, V_{\perp} , and a quantity called the particle pitch angle, " α ", the angle the particle motion makes relative to the B direction:

$$\alpha = \sin^{-1}(V_{\perp}/V) \quad (15)$$

$$\alpha = \cos^{-1}(V_{\parallel}/V)$$

The motion of the particle can be pictured as spiraling along the magnetic field direction, executing cyclotron motion around the field while moving along the field (Fig. 13). A charged particle will deviate from these simple motions if there is an electric field or if the magnetic field has temporal changes. As an example, consider the case where the magnetic field increases with distance in a direction perpendicular to the direction of B . In this case as the particle moves from the region of low field strength to high field strength and back again, R_c decreases and increases correspondingly, and the particle traces out a cycloid configuration (Fig. 11). Under the combined influence of both the Earth's electric field (this field is radially directed close to the Earth and points from dawn to dusk at greater distances) and the radial gradient of its magnetic field, charged particles will slowly trace a similar cycloid around the Earth (electrons drifting towards the east and high energy ions towards the west). Although such motion is quite complex, if the magnetic field gradient or E are sufficiently weak, the motion of the particle can be described in terms of its cyclotron motion and a constant "drift velocity".

The final type of motion of interest here results from gradients along the magnetic field and is responsible for the "trapping" of radiation particles in the magnetic field. If the magnetic field converges, then the particle will feel a small force along the direction of the field line. This will cause the particle to decelerate (accelerate) as it moves into the converging (diverging) region (Fig. 12). Eventually (unless the particle has collisions with atmospheric particles-i.e., the mirror point is below some critical altitude which we will define by the magnetic field strength at that position, B_C), the particle will have its motion parallel to the field stopped. However, due to the particle's circular motion (perpendicular to the field), it still experiences the decelerating force which reflects the particle back along the field line. This occurs at the "mirror" point, as determined by the strength of the magnetic field. This point is designated by " B_m " (Fig. 13). As will become evident, B_m and B_C are critical in determining charged particle motions.

2.2.1.2.2.2. INVARIANTS OF THE PARTICLE MOTION

The many different types of motion of a particle in combined E and B fields initially appear to be very difficult to handle. In fact, if the E and B fields are changing rapidly in time and space in comparison to the characteristic motions of the particle, then careful, single particle trajectories must be calculated for each and every particle. Fortunately, the Earth's E and B field generally change very slowly in time and space compared to the characteristic motions of radiation particles. We can then describe the particles' group motions in terms of so-called characteristic invariants of the motion (realizing that they do indeed change slightly) rather than having to calculate each particle trajectory. The adiabatic invariants help determine the scale over which the approximations hold. This section will describe those characteristic constants. First, consider the cyclotron motion of a particle. The motion of a particle in a slowly varying B and E field can be described as a cyclotron motion superimposed on a slow drift of the particle's guiding center, the center of the cyclotron gyration. The energy and angular momentum of a charged particle will remain constant during this motion if the adiabatic invariants are conserved. A particle which mirrors at a field strength of B_m has a pitch angle at an arbitrary B ($B_m \geq B$) given by:

$$\sin^2(\alpha) = \frac{B}{B_m} \quad (16)$$

This may also be expressed in the form of the first adiabatic invariant:

$$\mu = \frac{m v_{\perp}^2}{2 B} = \text{constant} \quad (17)$$

μ is also called the magnetic moment invariant because it is the magnetic moment of the charged particle ($\mu = IA$, where I is the current and A is the area of the "loop"). If the magnetic field changes slowly, the magnetic moment is conserved and the ratio of the particle energy perpendicular to the field will stay constant relative to the magnetic field strength.

The second invariant is called the longitudinal invariant. Before discussing it, consider the dipole magnetic field of the Earth. This field can be crudely represented by a tilted dipole with components:

$$B(r) = \frac{-2M \sin(\lambda)}{r^3} \quad (18)$$

$$B(\lambda) = \frac{M \cos(\lambda)}{r^3} \quad (19)$$

$$B(\phi) = 0 \quad (20)$$

where $M = 0.311 \text{ G-}R_E^3$, R_E is the Earth's radius (1 $R_E = 6371 \text{ km}$), and r is the radial distance from the center of the Earth; $B(r)$, $B(\lambda)$, and $B(\phi)$ are the field components in polar coordinates. For reference, a field line for this dipole field is defined by:

$$r = r_0 \cos^2 \lambda \quad (21)$$

$$\phi = \phi_0 = \text{constant}$$

where r_0 is the radial distance at which the field line crosses the magnetic equator and a field line is the line that would be traced by always moving in the direction of the B vector.

The dimensionless quantity L can be defined where:

$$L = \frac{r_0}{R_e} \quad (22)$$

The value of L is, by the definition of r_0 , the equatorial crossing point of a magnetic field line in terms of R_e . Referring back to the discussion of the effects of a converging magnetic field, if the cyclotron radius is sufficiently smaller than the curvature of the Earth's magnetic field lines, then, to a good approximation, the case of a particle moving back and forth between two magnetic mirrors is as illustrated in Fig. 13. The particle does not move along a surface of constant B , but moves on a surface of constant L between B_m (B_m is constant by virtue of the first invariant) in the northern hemisphere and in the southern hemisphere. In fact, the second invariant Φ defined by:

$$\Phi = \int V_{\parallel} dl \quad (23)$$

is a constant of the particle motion (dl is a unit of length along the particle trajectory). Constant Φ means that as a radiation particle drifts, it traces out a cycloid around the Earth's equator and follows a well defined surface. This surface is defined by the "L shell" of the particle and the two mirror points. This fact has given rise to the use of the McIlwain B-L coordinate system in which a particle population is completely described in terms of the particle flux as a function of B and L values—this is the fundamental underpinning of most existing trapped radiation models in use today.

Particles which have B_m values that fall below about 100 km are generally lost to the atmosphere by collisions with neutral particles so that frequently when plotting particle distributions as a function of α , a gap in the flux is found near 0° and 180° (this is not always true, however; see later). This is called the particle loss-cone angle for obvious reasons.

Combining equations 18-20, it can be shown that:

$$B(\lambda)/B_o = \frac{(4 - 3\cos^2(\lambda))^{1/2}}{\cos^6(\lambda)} \quad (24)$$

The magnetic latitude that a particle with pitch angle α_0 at the magnetic equator will mirror can be determined by combining Eq. 24 with Eq. 16 to get:

$$\sin(\alpha_c) = \frac{\cos^3(\lambda)}{(4 - 3\cos^2(\lambda))^{1/4}} \quad (25)$$

This expression can be used to determine the maximum pitch angle that a spacecraft can see at a given latitude in space. That is, if the spacecraft passes through the magnetic field line at a magnetic latitude of λ , it can only detect particles that have equatorial pitch angles of α_0 or less (for 0° to 90°). All particles with initial pitch angles greater than α_0 at the equator will have mirror points which lie between the spacecraft latitude and the magnetic equator and, therefore, will not be detected by the spacecraft detectors.

The third invariant of the motion, Q , is the flux invariant. This is simply the net magnetic flux inside the longitudinal-invariant surface defined by:

$$Q = \int \bar{B} \cdot d\bar{s} \quad (26)$$

where $d\bar{s}$ is an element of area. This invariant implies that the guiding center of the particle follows slow changes in the Earth's magnetic field.

The three invariants together allow the development of simple time-averaged models of the radiation particle fluxes. It should be remembered, however, that perturbing electric and magnetic fields do occasionally modify the particle motion so that the three invariants are violated. There are perturbations in the magnetosphere which occur on time scales of the cyclotron, the bounce, and the drift periods which violate the three invariants and alter the particle motion. Such variations lead to the diffusion of the particle populations (primarily the electrons) in pitch angle so that the loss cone can sometimes be filled with particles.

2.2.1.2.3. ELECTRON AND PROTON BELTS

There are potentially many different ways to model the Earth's radiation environment. Fortunately, the use of adiabatic invariants and the introduction of the McIlwain B-L coordinates have lead to a standardized means of representing the time-averaged features of the trapped radiation environment. Until the recent CRRES mission, the AE/AP series of radiation models developed by Vette and his colleagues at NASA Goddard have been the principle source of a uniform set of practical models of the Earth's trapped radiation environment. Here, although the CRRES-derived data may ultimately change the details (see Gussenhoven et al.^[14] for an initial comparison of the models with CRRES data), the major characteristics of the Earth's trapped radiation environment will be summarized in terms of these AE/AP models with emphasis on the critical environmental variables. The AP8/AE8 models are based on compiled data from many different satellites.^[15-17] The P and E in the model names AP8 and AE8 refer to "Proton" and "Electron" and 8 is the version number of the models. For a given set of McIlwain B-L coordinates in the range from low Earth orbit to somewhat beyond geosynchronous, AP8 and AE8 provide estimates of the omnidirectional fluxes of protons in the energy range of ~50 keV to 500 MeV and electrons in the energy range of ~50 keV to ~7 MeV. Time-dependent variations of the radiation fluxes such as those due to geomagnetic storms or short term solar modulations are not included in AP8/AE8. However, the models do differentiate between solar cycle maximum and minimum conditions. For protons, a larger flux is predicted at solar cycle minimum than solar cycle maximum. The situation is reversed for electrons with higher fluxes at solar

maximum. Although the AP8/AE8 models are very useful in many applications, it should be remembered that there are other means of representing the environment that may be more appropriate for specific uses. (Note: the CRRES dose results that have recently become available will primarily provide updates to AE8/AP8 for solar maximum.)

A common model of the radiation environment is one where the omnidirectional flux of particles, integrated over an energy interval, is given as a function of B and L coordinates. By definition, omnidirectional means that the particle flux has been averaged over all pitch angles. As discussed earlier, units are typically "particles per square centimeter per second" with the integrated energy channel being from a lower energy to the highest energies capable of being measured by space detectors. Although there are specific cases such as in shielding design for sensitive particle or light detector systems where pitch angle is important, in practice the design community that uses these codes only requires the omnidirectional flux. In principle, it is possible to construct this type of simple model by one measurement of the fluxes at all pitch angles as a satellite moves away from the Earth in the magnetic equatorial plane. The major difficulty in developing such a model of the Earth's radiation belts is that both space and time must be factored into the model. Although the use of the adiabatic invariants and B-L coordinates are very useful in simplifying this task, in reality asymmetries in the Earth's magnetic and electric fields and their time variations introduce significant complications into the modeling process. In particular, "shell-splitting" (particles of the same energy but different pitch angles tend to follow slightly different drift paths around the Earth so that particles observed together at the equator on one side of the Earth are separated in radial distance on the other side), distortions in the Earth's magnetosphere, and similar effects force the inclusion of temporal and local-time (or, less precisely, longitudinal) variables. The AE/AP model fluxes, for example, are parametrically represented by:

$$I(>E, B, L, \tau, T) = N(>E, L) \Phi(>E, L, \tau) G(B, L) \quad (27)$$

where I is the integral omnidirectional flux, $>E$ means for all energies above E, τ is the local time, and T is the epoch (or date). No one spacecraft can collect sufficient data over a long enough time interval or in enough locations to adequately define Eq. 27. Rather data from many different satellites are averaged in discrete B and L bins to determine the B-L variation G; in energy, L, and local time to determine the local time variation Φ (note: B variations were ignored because there was often too little data to allow simultaneous binning in terms of B also); and in energy and L bins to determine the energy variations N. As many different satellites were used, it was important to know the efficiency and geometric factor for the different detectors. As discussed by Vette and his collaborators, there are many regions of spotty spatial coverage^[15]. Likewise, short term temporal variations are not well represented.

Fig. 6 for 1 MeV electrons and 10 MeV protons illustrates the basic structure of the radiation belts as predicted by the AE8 and AP8 models. In particular, the electron contours show a dual peak (the protons have a similar structure but the division between the peaks is less obvious). Typically, therefore, the radiation belts, according to these models, are divided into "inner-zone" and "outer-zone" populations. This division also roughly corresponds, for the electrons, to an inner belt which is weakly affected by geomagnetic storms and an outer belt which is greatly affected by storms. The L-shell region up to L~2.5 is termed the inner-zone while the region beyond L~3 is considered to be the outer-zone with a "slot" region of reduced density in between. The inner-zone electrons peak around L = 1.45 to 1.5. Little variation with geomagnetic activity is seen below L~1.6. The inner-zone protons are very stable, varying inversely with atmospheric density (the fluxes are lower at solar maximum when the atmospheric density is highest).

The proton flux peaks near $L = 1.45$. In the outer zone, the peak L shell varies with energy for the electrons and flux increases can be as great as 10^5 in less than a day during a geomagnetic storm. The outer-zone protons, in contrast, do not show as strong a division into an outer belt as the electrons nor as much variation with geomagnetic activity. Protons with $E \geq 1$ MeV peak at about $L = 3$ while protons with $E \geq 10$ MeV peak at about $L = 2.5$. (Note: all numbers are adapted from Vampola^[18].) To summarize, the AP8/AE8 models describe the following populations:

- a) The slowly varying, stably trapped high energy protons and the inner zone electrons ($\leq 3 R_E$). (Note: the AP8MAX and AP8MIN differ only for altitudes less than about 1000 km.)
- b) The trapped but highly variable outer zone electrons ($\geq 3 R_E$)

Basic uncertainty factors have been defined for these models. These are (for 5 to 10 year averages):

AP8:	(min. and max.):	a factor of 2
AE8:	(min. and max.):	a factor of 2

Below 1000 km, the main trapped radiation environment consists of 2 components: the low altitude extension of the radiation belts (or "horns") at high latitudes and the low latitude South Atlantic (Magnetic) Anomaly. Figures 14 and 15 illustrate these regions for protons and electrons^[19]. The units are particles/cm²-s for $E > 30$ MeV and $E > 0.5$ MeV respectively. The lines represent isoflux contours at 296 km (Fig. 14) and 400 km (Fig. 15). It is interesting to note that the South Atlantic Anomaly, which is the result of low magnetic field intensities in the South Atlantic near Brazil (Fig. 10), is slowly drifting westward with a period of about 1200 years. High energy, trapped particles mirror at a characteristic constant magnetic field strength—thus particles that typically would mirror at higher altitudes above the atmosphere find themselves mirroring at much lower altitudes in this region thereby enhancing the background fluxes.

2.2.1.2.4. LONG-TERM AND SHORT-TERM VARIATIONS

Here the concepts developed in the previous sections will be used to interpret the difficulties that arise in estimating the trapped radiation environment. Of principle interest will be the complications introduced by time variations in the plasma environment over the solar cycle, effects of the decay of the Starfish and Argus nuclear tests, and short term geomagnetic storm effects. These effects currently all complicate attempts to accurately characterize the radiation particle population in an "average sense".

2.2.1.2.5. EXAMPLES: SOLAR CYCLE EFFECTS, STORMS, SUBSTORMS

Figure 16 illustrates approximately one year's worth of hourly averages of the 1.9 MeV omnidirectional electron flux measured at midnight by the geosynchronous satellite ATS 1^[20]. The daily sum of the geomagnetic index K_p (at the bottom of the figure) and the value predicted by the AE model (the horizontal line) are also plotted for comparison. This figure demonstrates two important points. First, the electron radiation flux at $L=6.6$ is highly variable on a daily time scale—some variations being on the order

of 10 to 100. Secondly, the AE model is biased toward the few major geomagnetic storms. This biasing is to be expected as the model is derived by averaging the fluxes. Figure 17 demonstrates electron variations over a slightly longer time period for different L values. The data are 10 day averages of the electrons with energy greater than 0.28 MeV, taken on the 1963 38C spacecraft[20]. Inside of L=1.8 (generally referred to as the inner electron zone) the time variations are quite small, demonstrating the usefulness of an average model in this region. The steady decay of flux levels in the figure is due to the decay of the residue from the artificial Starfish injection event of July 1962. In contrast, outside L=1.8, the fluxes at L=2.2 vary greatly with time due to geomagnetic activity. As a final example of long term variations, consider Figure 18.[21] This plot presents running 27-day averages of omnidirectional electron fluxes at geosynchronous orbit over a solar cycle. Energy channels (in MeV) are indicated along with the corresponding satellite (all are geosynchronous). The fluxes are remarkably stable over this period if sufficiently long averages are taken[21]. The most pronounced variations in the figure are those associated with the 27-day solar rotation period and an apparent semi-annual variation. Paulikas and Blake[21] also found a strong correlation between solar wind velocity and the energetic electron fluxes. The principle point to be made by this and the previous figures is, however, that large time variations are observed in the outer electron zones but that long term averages do indeed smooth out these variations so that average models are useful in many applications.

2.2.1.2.6. PITCH ANGLE MEASUREMENTS

2.2.1.2.6.1. EXAMPLES OF LOSS CONE AND FIELD-ALIGNED DISTRIBUTIONS

As the pitch angle distribution at one point on a field line in principle gives information on the plasma distribution at other locations along the field line, it is important that this function be properly defined. This is not straightforward as it is difficult in practice to accurately determine the pitch angle distribution. Further, particles are observed with pitch angles along the magnetic field direction. These particles are believed to be produced by non-adiabatic processes and thus violate the basic principles on which the time-averaged radiation models are based. Examples of these difficulties will be discussed in this section.

In principle, determination of the pitch angle distribution at the magnetic equator determines the plasma distribution along the entire field line. In practice, spacecraft are seldom at the magnetic equator nor are they able to sample all pitch angles. The issue which arises is what are acceptable limitations on the pitch angle distribution? Figures 19, 20, and 21 present typical examples of proton and electron pitch angle distributions at high energies. These measurements are all from Explorer 45[22, 23] and cover L-shells from 1.7 to 5.2. The electron distributions at L=3 are at 2 day intervals following the large geomagnetic storm of June 17, 1972, and illustrate loss-cone distributions along the magnetic field direction (the June 16 segment illustrates a prestorm interval). Note the gradual decay of the distribution back to the prestorm levels. The proton observations, on the other hand, illustrate variations in pitch angle as a function of L-shell following the storm of December 16, 1971 (Figure 21). Observe the differences in isotropy and how they change with distance from the Earth. Clearly, the variations in particle pitch angle distributions are quite complex at all energies and vary greatly with L-shell. Given the sharply field aligned fluxes on some occasions and the pronounced loss cone distributions on others, it is clear that interpolation of the pitch angle distribution from a limited data set can be a difficult undertaking. The fact that the pitch angle distribution can change over the course of a geomagnetic storm and that it

can vary drastically with energy, greatly increases the number of measurements that must be made and the time intervals over which data must be taken.

2.2.1.2.6.2. COVERAGE LIMITATIONS

In addition to the limitations associated with temporal and pitch angle variations, the AP/AE models are limited by coverage. Specifically, existing models are affected by lack of adequate coverage at some pitch angles and by lack of coverage during the current epoch. These limitations are illustrated in Figure 22^[15]. Here the B/Bo-L coverage for the experiments used in constructing the AP-8 model is illustrated. As is clearly demonstrated, that coverage can be quite limited and over many ranges may only be measured by one satellite.

Although a given AE/AP model may be accurate for the period of time for which the measurements were taken, it does not necessarily follow that it is accurate for another epoch. That is, there are long term variations such as the decay of the Starfish nuclear explosion radiation (see section on man-made environments) that change over much longer time scales than data exists. The solar cycle, for example, follows an irregular 11 year pattern and the level of geomagnetic activity at one solar minimum may differ substantially from that of the next. Thus, for a model to be appropriate for a given time interval, it should ideally include observations from that interval. As this is impossible in practical terms, the data base used should be as close in time as possible if accurate predictions are to be expected. The most recent observations for the AE models (and these are principally from one satellite) are over 15 years old whereas for the AP-8 model they are almost 20 years old! To resolve this issue, the CRRES spacecraft was launched in 1989 to remap the radiation belts and to study radiation effects on microelectronics. The results of the 1 year mission of the CRRES clearly revealed problems with the current models, in particular, a highly variable third belt was discovered between the main 2^[14]. As the CRRES data are published, it is hoped that more accurate models will become a reality.

2.2.1.2.6.3. MAGNETOSPHERIC HEAVY IONS

There is currently only limited information available on heavy ions in the Earth's magnetosphere (CRRES data will hopefully change this). O⁺ and He⁺⁺ are typically observed. The sources of these particles are not clear, however. The helium nuclei (mostly alpha particles) are likely from the solar wind,^[24] while the O⁺ may be primarily of ionospheric origin. In the case of the solar wind, particles probably enter the magnetosphere and are accelerated by radial diffusion. This process, described by Cornwall^[25], adequately describes the magnetospheric helium ion population.^[26, 27] The bulk of the helium nuclei are, however, at energies too low to penetrate the walls of a spacecraft. C, N, and O ions have all been observed though it was not clear whether the particles were trapped in the magnetosphere (recent evidence supports the claim that the bulk of the O⁺ is of ionospheric origin). Models by Adams and his colleagues^[28] assume that there is a small flux of helium nuclei and a smaller flux of heavier nuclei in the magnetosphere above 10 MeV/nucleon. There are also reports of long-lasting enhancements of the low energy heavy ion fluxes after large solar flares, perhaps associated with the Galactic Cosmic Ray (GCR) anomalous component (see later). Blake and Friesen^[24] suggested that the anomalous component particles, which may be only singly ionized initially, become more highly ionized near their geomagnetic cutoff and are subsequently trapped. This leads to a special trapped population of oxygen, nitrogen, neon, and other elements which may comprise the heavy ion component (see

Table 2). As yet, these ions are included in radiation dosage calculations in only very specialized applications although they are of concern for SEU effects.

2.2.1.2.7. CONCLUSIONS

Modeling of the Earth's radiation environment is a complex process at best. For the purpose of studying radiation effects on long duration space missions, the AE/AP models produced by the NSSDC have proven to be very useful. There are, however, limitations to these models such as effects due to large, but short term temporal variations. Important, but less serious, are the effects that inadequate data coverage and the lack of recent data have on their accuracy. It must be realized, however, that any quantitative estimates of the inaccuracies in the current models are very crude as such estimates are intimately tied to the issue of what the models are to be used for. At best, estimates of these inaccuracies are made based on comparisons between the levels of dosage the models would predict for a given orbit and the actual levels measured in that orbit. Vampola^[29] has attempted such estimates and finds the models to be within a factor of 2 for long time averages (~years). It is hoped that the CRRES data for solar maximum can be used to both verify the existing models and to update their predictions.^[14] The improvements in sizing satellite shielding to be expected with better models can not be underestimated--this can be graphically demonstrated (in both senses of the word) by a simple parametric comparison. Figs. 23 (electrons) and 25 (protons) show the range in flux as a function of energy spectrum shape expected based on the current uncertainties in the radiation environment^[30] and the resultant effects on the shielding for both electrons (Fig. 24) and protons (Fig. 26). As a practical example, improvements in the knowledge of just the slope of the electron spectrum for energies greater than 2 MeV might mean perhaps as much as an order of magnitude reduction in the required shielding on a typical spacecraft (though factors of 2 are more likely) if it could be demonstrated that the softer spectra were more appropriate for long mission.

2.2.1.3. NEUTRAL PARTICLE RADIATION

Neutrons (at rest) typically have a half-life of 11.7 minutes before they decay into an electron, proton, and anti-neutrino. As a result, all but the most energetic solar flare neutrons (and therefore extremely small fluences) will have decayed before they can reach the Earth's orbit. Likewise, there is no primary neutron component to the intergalactic cosmic rays. The observed neutron component is, instead, associated with secondary cosmic ray (both solar flare and intergalactic) processes. The neutrons produced by these processes are discussed below.^[6]

A primary Galactic Cosmic Ray (Fig. 27) incident on the Earth's atmosphere will interact with air nuclei to produce a multitude of high energy secondary cosmic rays. These in turn will interact to produce additional particles in a "nuclear cascade". The production of these secondary particles becomes significant around 55 km reaching a maximum value at ~20 km (called the Pfotzer maximum). The intensity falls off below this altitude. At high energies, the so-called "knock-on" process dominates neutron production while at lower energies, "neutron evaporation" dominates. These neutrons interact with the atmosphere to produce various radioactive isotopes such as ¹⁴C. Representative differential energy spectra for neutrons at various depths in the atmosphere are presented in Fig. 28.

About 10% of the secondary cosmic ray neutrons escape into space. Most of these neutrons (which are not controlled by the Earth's magnetic field) will decay (11.7 minute

half-life) into electrons and protons which will be trapped in the Earth's radiation belts (these "cosmic ray albedo neutrons"-CRAN-are a major source of the high energy protons in the belts). A globally averaged albedo neutron spectrum is presented in Fig. 29. Near the Earth, there is an approximate 1:7 ratio of albedo neutrons from the equatorial atmosphere as compared with the polar atmosphere. For reference, a list of major flares and the corresponding ground-level measurements of the secondary neutron flux are presented in Table 3 for 1968 to 1972. Significant ground level neutron events are relatively rare and are normally observed in conjunction with major solar flares.

2.2.1.4. SOLAR FLARES

2.2.1.4.1. SOLAR PROTON EVENTS

A very energetic process, and potentially damaging situation, occurs when very strong magnetic fields in the solar photosphere reach a critical instability. On time scales of seconds, the strong fields are unstable enough to "snap", thereby adjusting and relaxing to remove the instability. A considerable amount of energy, up to 0.1% of the total solar energy output or about 10^{32} ergs, is released during this "flare". A solar flare typically lasts from a few minutes to a few hours and heats the surrounding corona to temperatures in excess of 2×10^7 K. Associated with the heating, large fluxes of atomic particles, primarily electrons and protons, are accelerated and expelled from the Sun. There are also substantial radio bursts and X-ray emissions. Many of the particles escape the Sun and into the interplanetary medium where they follow the solar wind magnetic field out through the ecliptic plane into interstellar space. As a result of the complexity of this field structure, both the intensity and the spectrum of the particles observed at Earth can vary significantly with position. These quantities depend critically on the positions of the Earth and the flare on the Sun relative to each other. Variations as large as 100 in the particle fluxes from the same flare at different points around the Earth's orbit have been observed. A schematic illustrating the time scales for the different solar flare events is presented in Fig. 30 (the 27 day period is the approximate rotation rate of the Sun as seen from the Earth).

The intensity and number of very large solar flares varies dramatically from solar cycle to solar cycle (Fig. 31). The seeming lack of correlation between sunspots and proton events in Fig. 31 illustrates the difficulty in reliably predicting the level or frequency of activity for the future (see, however, Feynman et al.^[31]). Adding to the problem is that most data on large solar flares is from only the last few cycles making statistical studies very difficult. One fact is clear, however, and that is that solar flare related effects occur far more frequently during solar maximum (Fig. 32). Although solar flares are less frequent during solar minima, these flare related effects can be greater in the terrestrial magnetosphere because the interplanetary fields, through which the flare particles travel, are less complicated and the particles can more easily gain access to the Earth's polar caps. As an example, even though the number of geomagnetic storms goes down during minima in solar activity, the level of a given storm, even during the lowest levels of solar activity, can be among the highest ever seen. A particularly good example was the great flare of 1972 which occurred nearly 4 years after the peak in solar activity. It is compared in Fig. 33 with the largest flare of the last solar cycle maximum for $E > 10$ MeV protons.

The intense fluxes of high energy protons that are associated with the solar flares are very damaging to electronics both from the SEU and total dose standpoint (note: solar proton events are often referred to as solar flares; strictly speaking, a solar flare, which involves many processes, may or may not have a classifiable proton event).

Hydrogen and heavy nuclei in the ~1 MeV/nucleon to ~10 GeV/nucleon energy range are ejected during solar proton events. Their intensities are generally a few to several orders of magnitude larger than those of GCR at these lower energies, depending obviously on the size of solar flare. Adams and his collaborators at the Naval Research Laboratory (NRL) have developed several detailed models of the radiation environment associated with a flare [3, 28, 32, 33]. Typically, these and other flare models are based on the large proton event of 1972. These are combined with models of the mean abundances of the heavier ions relative to hydrogen to obtain a worst case model. Figure 34 presents Adam's characterizations of "typical", "worst-case", and "anomalously large" (i.e., very rare) solar flare proton spectra for 1 AU. The worst-case solar flare proton flux is ~5 orders of magnitude larger than the GCR flux, but becomes "softer" above ~10 GeV.

For many years, there has been an ongoing controversy over how to model the probability of the occurrence of solar flare proton events and to estimate their radial variations. Recently, Feynman and colleagues have developed a comprehensive model of solar flares that largely replaces previous models.[31] A major change has been the realization that instead of there being three types of flares (typical, worst-case, and anomalously large) in terms of energy output, there is indeed a continuum when the total output energy of the flares in an event series is integrated. In fact, when plotted in terms of the statistical occurrence frequency, it was found that the likelihood of the occurrence of a solar flare amplitude in a given time interval could be predicted. This new model, as represented in Figures 35 and 36[31], has been used to estimate the likelihood of encountering a given mission-integrated fluence for various interplanetary spacecraft. Fig. 35 plots the cumulative probability of occurrence of a solar proton event fluence of the indicated level or less between 1963 and 1991 (the scale is for a log normal probability such that a log normal distribution will be a straight line). Fig. 36 uses these data to predict the probability of seeing a total proton dose ($E > 10$ MeV) equal to or less than the graphed value for various mission lengths during solar maximum (the 7 years centered around solar maximum—the other 4 years are assumed to have 0 flux). Similar plots are available for other energy intervals and are currently being extended to other ions. In a later subsection, these data are used to estimate the total dose due to protons on a representative space mission. Based on the recommendations of a workshop on the "Interplanetary Particle Environment"[34], the new model assumes a (conservative) radial dependence of r^{-3} for the flux inside 1 AU and a radial dependence of r^{-2} outside 1 AU. This implies that the fluence and dose will be trajectory dependent for interplanetary vehicles such as Galileo and Cassini.

2.2.1.4.2. HEAVY IONS

In addition to intense proton fluxes, solar flare events also typically are accompanied by small but variable amounts of heavy ions. Tables 4 and 5[28] summarize the abundances of the heavy ions found in solar flare events relative to hydrogen. The abundance ratios from the tables are multiplied by the proton flare spectrum to get the individual ion spectra. For the Adams' model, the worst-case composition for the elements from copper to uranium can be estimated by multiplying the abundance ratios of Table 5 by $1.92 \exp(z \cdot 79 / 6.89)$. Although the solar heavy ions are usually assumed to be fully ionized, they may not be, particularly in the $0.5 \leq E \leq 2.5$ MeV/nucleon energy range .[35] Fischer et al.[36] (see also MeWaldt and Stone[37]) showed that solar energetic heavy ions in the energy range $5 \leq E \leq 20$ MeV/nucleon may also not be fully ionized and that the upper limits on the charge to mass ratio of the heavy ions may be as low as 0.1 (~0.5 for fully ionized heavy ions). Breneman and

Stone^[38] have indirect evidence that heavy ions in the energy range 3.5 to 50 MeV/nucleon may have the same distribution of charge states as for the 0.5 to 2.5 MeV/nucleon ions observed by Luhn et al.^[35]. The importance of the charge state of the ions is that it affects their entry into the Earth's magnetosphere—the higher the charge state of a given atomic species, the less likely it is to get through the Earth's magnetic field for the same energy.

2.2.1.5. GALACTIC COSMIC RAYS (GCR)

2.2.1.5.1. OVERVIEW

The third source of background radiation is the Galactic Cosmic Rays or GCRs. Galactic Cosmic Rays are primarily interplanetary protons, electrons, and ionized heavy nuclei with energies ranging from ~1 MeV/nucleon to over $\sim 10^{10}$ eV/nucleon. Fig. 37 compares GCR proton and electron spectra while Fig. 38 displays the observed cosmic ray abundance distribution of the chemical elements in the energy range from ~100 MeV/nucleon to ~1 GeV/nucleon for hydrogen to the iron group.^[39] For comparison, solar system abundances are also shown in the figure. Note that the two abundance distributions are strikingly similar. The major components are hydrogen (93.6 percent) and helium (6.3 percent). The remaining 0.14 percent includes all the rest of the elements. Observations indicate that outside the Earth's magnetosphere, the cosmic ray fluxes are isotropic to within about 10% over their entire energy range, suggesting that they are of galactic and/or extragalactic origin. Within the magnetosphere, however, they are not isotropic. That is, for the low Earth orbit environment, the geomagnetic field provides shielding against incident GCRs (and solar flare particles) as it can effectively deflect through the Lorentz force the lower energy particles. Because of the approximate dipole nature of the geomagnetic field, vertical particle velocities in the polar regions are essentially parallel to the magnetic field resulting in almost no Lorentz force so that the particles can gain direct access. At low inclinations, only particles with sufficiently high energy, or "rigidity", can penetrate through the magnetic shielding as the charged particles are forced into gyro loops by the magnetic field as they approach the Earth. This effect, illustrated schematically in Fig. 39^[40], is discussed in more detail in the next subsection.

2.2.1.5.2. MAGNETIC SHIELDING

Earth's magnetic field serves as an extremely effective shield for low to medium energy cosmic rays. Very succinctly, as in the case of a trapped charged particle, a cosmic ray penetrating the Earth's magnetic field begins to execute a circular motion around the magnetic field. Dependent on its energy, mass, and charge, the gyro-radius or penetration distance of the particle may or may not reach to the inner magnetosphere of the Earth. This ability to penetrate the magnetic field, proportional to the gyro-radius of the particle, is determined uniquely by the cosmic ray's momentum divided by its charge and is called the particle's rigidity:

$$P = \frac{pc}{q} \quad (28)$$

where:

P = rigidity

p = momentum of particle

c = speed of light

q = charge of particle

If pc is in eV, then q is the number of charge units and P is in volts. Typical units are MV or GV. There is a minimum magnetic rigidity a cosmic ray must possess to arrive from a given direction at a given point in the magnetosphere. Regions in the outer magnetosphere and near the poles can be reached by much lower magnetic rigidities than near the Earth's equator. In general, for each point in the magnetosphere and for each direction at that point, there exists a rigidity below which cosmic rays cannot arrive. This value is called the geomagnetic cutoff. For rigidities above this value, cosmic rays can arrive freely, almost as if no magnetic field were present. Fig. 39 illustrates how the rigidity is convolved with the GCR spectrum outside the Earth to obtain the spectrum for a given orbit.[40]

2.2.1.5.3. GCR ION SPECTRA

The most abundant element in cosmic rays is hydrogen. Fig. 40[39] compares the differential energy spectrum of GCR hydrogen (primarily H^+) for solar maximum and solar minimum conditions with other GCR ion spectra. At very high energies, a power law with a spectral index of 2.75 is a good fit. The deviation below about 5 GeV/amu is thought to be due to solar modulation. As the amount of solar modulation depends on the general level of solar activity, there is a variation from solar minimum to solar maximum. Hydrogen, helium, carbon, oxygen, neon, magnesium, silicon, sulfur, calcium, and iron are assumed to be primarily primordial (that is, the hydrogen flux consists of the originally created ions and has not been created by transmutation during transit across the galaxy). It and these heavier ions are thus believed to have similar energy spectra.

The NRL organization, in addition to its flare models, has created a family of GCR heavy ion model spectra corresponding to the curves shown in Fig. 40. The NRL models provide fits to the hydrogen, helium, and iron spectra along with a formula for scaling other elements to these 3 basic spectra. As lithium, beryllium, and boron are assumed to be entirely composed of secondaries and nitrogen is assumed to be a mixture of primary and secondary particles, they have energy dependencies different from these 3 base species. The NRL models take these variations into account by modifying the ratio to either helium or iron as a function of energy. The differential energy spectra of hydrogen (f_1), helium (f_2), and iron (f_{26}) nuclei are estimated for energies above 10 MeV/amu by:

$$f(E,t) = A(E) \sin\{\omega(t-t_0)\} + B(E) \quad (29)$$

where:

- ω = 0.576 radians per year
- t_0 = 1950.6 AD
- t = date (in years)
- E = particle energy in MeV/nucleon
- $A(E)$ = $0.5 (F_{\max} - F_{\min})$
- $B(E)$ = $0.5 (F_{\min} + F_{\max})$

The units of f are particles/ m^2 -ster-s-MeV/ μ . The functions F_{\min} and F_{\max} refer to solar minimum and solar maximum conditions for the following equation and differ by constants:

$$F = 10^m \left(\frac{E}{E_o} \right)^a \quad (30)$$

where:

$$m = C_1 e^{-X_2 \{ \log_{10} E \}^2} - C_2 \quad (31)$$

The exponent of the energy ratio changes from solar minimum to solar maximum. It is given by:

$$a = a_o \left\{ 1 - e^{-X_1 (\log_{10} E)^b} \right\} \quad (32)$$

The fitted parameters for hydrogen, helium, and iron are given in Table 6.^[3] Using these three spectral shapes, the rest of the elements are then modeled using the ratios in Tables 7 and 8.

2.2.1.5.4. GCR VARIATIONS

The flux of GCR increases radial with distance from the Sun. The magnitude of the radial gradient varies with both ion species and energy. For relativistic GCRs the radial gradient is <4%/AU while for GCR particles below 100 MeV/ μ the gradient is <10%/AU. The latitudinal gradient is small, under 1% per degree and may change sign each half solar cycle. The solar wind further modulates the GCR inversely with the 11 year cycle of solar activity. At the maximum of solar activity, GCR intensity is at a minimum and vice versa. The GCR intensity, at moderate energies, varies by a factor of 4 to 8 depending on the energy and ion.^[28] The reason for the modulation of the GCR is apparently the solar wind magnetic field. As the GCR propagation is controlled by this field, any turbulence in the field (such as that associated with solar maximum) makes propagation of the GCR into the inner solar system more difficult--the particles are scattered more effectively. As the solar wind field relaxes during solar minimum, the GCR can more easily reach the inner solar system.

In addition to the temporal and spatial variations in the GCR, there is an occasional systematic change in the energy spectra for some ions. This so-called anomalous component appears as a flattening in the differential energy spectra of ions like helium, oxygen, nitrogen, and neon. The anomalous component may be singly ionized particles with an energy in the range from 1 to 200 MeV/ μ --a component which is not always present near the Earth.^[28] The anomalous component varies by a factor of 100 to 1000 over the 11 year solar cycle--it appeared between 1971 and 1972 and disappeared again during the solar maximum of 1978. Jokipii et al.^[41] have predicted that the anomalous component appears near the Earth only once every other solar minimum (i.e., it will appear again ~1994). The anomalous component seems to be more intense at greater distances from the Sun. Since the composition of the anomalous component suggests that it is interstellar gas accelerated by the solar wind, Fisk et al.^[42] have predicted that only atoms with a first ionization potential higher than hydrogen will display anomalous spectra and that the ions will be singly ionized. If the anomalous component is indeed singly ionized, it can penetrate much more deeply into the Earth's magnetic field and may account for some of the measurements of heavy ions in the inner magnetosphere as mentioned earlier.

2.2.1.5.5. GCR ELECTRONS

Although Galactic Cosmic Rays are primarily interplanetary protons and ionized heavy nuclei, electrons are also a component of GCR. Their measured intensities at energies above ~ 100 MeV, however, are at least 1 order of magnitude smaller than that of the protons (Fig. 37). The model for GCR electrons presented here is taken from NASA SP-8118^[43]. The integral intensity for galactic cosmic ray electrons, in units of ($\text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}$), is given by:

$$I_{\text{mean}} = \frac{2000}{(\alpha E^6 + \beta E^3 + E^2 + \gamma)^{1/4}} + \frac{400}{E^{3/2}} \quad (33)$$

where: $\alpha = 10^{-12}$

$$\beta = 10^{-3}$$

$$\gamma = 6 \times 10^4$$

$$\text{fluence} = \int \int I \, d\Omega \, dt$$

2.2.1.6. INTERPLANETARY e⁻ AND H⁺

Although the interplanetary proton and electron environments are of little impact on the radiation environment, they are included here for completeness. The models for interplanetary electrons and protons are taken from NASA SP-8118.^[43] These particles fall in the energy interval between the very low energy solar wind and the energetic solar protons (approximately 1 keV and 10 MeV). NASA SP-8118 gives the integral intensities (in units of $\text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}$) for these particles in terms of the following functions:

For intermediate energy protons:

$$I_{\text{ave}} = \frac{1.0 \times 10^6}{r^2 E^{1.7}} \quad 10^{-3} < E < 10 \text{ MeV} \quad (34)$$

For intermediate energy electrons:

$$I_{\text{ave}} = \frac{400}{r^2 E^2} \quad 2 \times 10^{-5} < E < 10 \text{ MeV} \quad (35)$$

where E is in MeV, r in AU, and the fluence = $\int \int I \, d\Omega \, dt$. The dose due to these particles is much smaller than that due to the typical solar flare environment and is thus ignored in most radiation calculations.

2.2.1.7. EXTRATERRESTRIAL TRAPPED RADIATION

As in the case of the Earth, many of the other planets in the solar system have been observed to have trapped radiation belts. The species, abundances, energies, and time variations of particles that are trapped in these radiation belts vary greatly depending upon the planet and its magnetic field. Planetary magnetic fields influence the particle spectrum that is observed near a planet in two ways--first, the magnetic field of the planet shields the planet from solar flare particles and from the GCR and, second, it allows particles to be trapped near the planet in radiation belts. Fig. 41^[44] and Table

9[44] compare the magnetic fields and magnetospheres of the various planets. The Earth, for its size, has proportionally one of the most intense magnetic fields in the solar system. Jupiter and Saturn, like the Earth, should have (based on their magnetic fields) intense radiation belts. Subsequent flybys of Jupiter and Saturn have indeed born this out. Here, the discussion will concentrate on the Jovian radiation belts which, after the Earth's, are the best known and are by far the most intense in the solar system. A brief discussion of the Saturnian observations will also be presented (the Uranian and Neptunian belts are also of interest but are not as yet well modeled; Mercury, Mars, and Venus have very weak magnetic fields and therefore no radiation belts).

2.2.1.7.1. JUPITER

The strongest magnetic field in the solar system is that of Jupiter. Since the ability to trap particles magnetically is a function of the magnetic strength, it is little wonder then that it has the most intense radiation belts yet observed. As will be demonstrated, these belts are so intense in fact that they rival the man-made saturated nuclear environment at the Earth—the most intense environment space systems will likely have to fly in. To date (the model is currently undergoing revision to reflect the recent Ulysses flyby of Jupiter), the principle engineering model of these radiation belts is the Divine formulation.[45] This model has many of the characteristics of the AE8/AP8 radiation models and thus can be reviewed in the same manner as the Earth's trapped radiation belts. The following discussion is abstracted from the original paper.

2.2.1.7.1.1. INTRODUCTION

Jupiter has been known to have a magnetosphere since about 1960 when, in analogy with early spacecraft observations of the Earth's radiation belts, it was realized that the Jovian UHF radio emissions could be interpreted in terms of trapped energetic electrons[46]. Early speculation by Brice[47, 48] and others attempted to draw parallels between this hypothesized Jovian magnetosphere and the then current ideas of the Earth's magnetosphere. In order to assess the potential hazard to the Pioneer 10 and 11 spacecraft, crude numerical models of the energetic electrons and protons were developed based on these speculations and the early radio observations. The successful encounters of the Pioneer spacecraft with the Jovian magnetosphere gave rise to a number of quantitative models describing various aspects of the Jovian magnetosphere.[49, 50] In particular, magnetic field models by Smith et al. [51] and Acuna and Ness[52, 53] began to delineate the substantial differences that exist between the Jovian and terrestrial magnetospheres. Pronounced wave-like variations in the high energy particle fluxes led to the proposal that the Jovian magnetosphere was distorted into a thin disc--the so-called magnetodisc theory (Fig. 42[44, 54])--and that this thin disc was populated by a cold plasma consisting of heavy ions originating from Io. The passage of the Voyager 1 and 2 spacecraft, while failing to distinguish between the magnetic-anomaly and magnetodisc models, further refined the particle and field observations. Subsequently, theoretical models have further helped to interpret the observations and have led to the beginning of the development of Jovian magnetospheric models capable of being used to make practical predictions about the environment around Jupiter (see reviews[49, 50]).

2.2.1.7.1.2. JOVIAN COORDINATE SYSTEM

In the Divine model^[45], the independent variables used to define position for the magnetic field and charged particles are jovicentric distance r (commonly in m or R_J), latitude λ (deg or rad), longitude I (deg in System III (1965)^[55]), distance $z=r \sin I$ from the rotational equatorial plane (m or R_J), and distance $R=r \cos I$ from the rotation axis (m or R_J). The value of the Jovian equatorial radius is assumed to be $1 R_J = 7.14 \times 10^7$ m. The common angular speed of rotation of Jupiter's internal magnetic field and of a meridian of constant longitude I in System III (1965) coordinates is assumed to be $\omega=870.536$ deg/day ≈ 12.6 km/s R_J . In this system, I , the longitude, increases westward (opposite to the azimuthal angle in a system of spherical coordinates). Conversions to inertial and other coordinate systems may be derived from Seidelmann and Divine^[55].

2.2.1.7.1.3. MAGNETIC FIELD MODEL

Of the encounters with Jupiter, the retrograde, highly inclined, small perijove trajectory flown by Pioneer 11 has been the most useful for modeling the moments of Jupiter's internal magnetic field. The detailed, 15-coefficient spherical harmonic model O4 derived from the fluxgate magnetometer on Pioneer 11^[52, 53] was used in deriving the model presented here. In this model, the dipole moment will be assumed to have the value $M=1.535 \times 10^{27} \text{ m}^2\text{A} = 4.218 \text{ G-}R_J^3$, and for each field line the magnetic shell parameter L has the constant value:

$$L = (M/B_e)^{1/3} R_J^{-1} \quad (36)$$

where B_e represents the minimum field strength along the line. As in the case of the Earth, among all field lines having the same value of L , the smallest field strength for which Jupiter's atmosphere (equatorial radius $1 R_J$, flattening 0.065) is encountered represents an upper cutoff field strength B_c for stable charged particle trajectories. The O4 model has been used to calculate B_c as a function of L (the upper dashed curve in Fig. 43) and to analyze the energetic charged particle data in the development of the numerical models. Although the O4 magnetic field model was used in deriving the model parameters, the much simpler offset tilted dipole model D4 derived from the Pioneer helium vector magnetometer data^[51] provides adequate accuracy for evaluating model parameters for many applications. The parameter values for this model are presented in Table 10.^[45] The nearly equatorial offset of about $0.1 R_J$ suggests that $L = 1.1$ represents the smallest accessible value of L for the trapped particles (the O4 model yields minimum $L = 1.089$). A transformation of coordinates based on Table 10 allows derivation of distance r_m and latitude λ_m in terms of r , λ , and I . The vector components of the magnetic field may then be derived from the D4 model using standard dipole relations^[51]. The magnetic field strength is given by

$$B = (M/r_m^3)(1+3(\sin_m \lambda)^2)^{1/2} \quad (37)$$

and the shape of a field line, along which the magnetic shell parameter L is constant, is given by

$$L = (r_m/R_J)(\cos_m \lambda)^{-2} \quad (38)$$

The dipole moment has the value $M = 1.538 \times 10^{27} \text{ m}^2 \text{A} = 4.225 \text{ G-RJ}^3$ and the tilt colatitude of 10.77° equals the inclination between the rotational and magnetic equatorial planes represented by $\lambda=0^\circ$ and $\lambda_m=0^\circ$ respectively.

2.2.1.7.1.4. RADIATION BELT e⁻ AND H⁺

The principle radiation belt populations included in this model are, as in the case of the Earth, electrons ($E > 0.06 \text{ MeV}$) and protons ($E > 0.6 \text{ MeV}$). The range of applicability of the energetic electron model extends to the Jovian magnetopause while that of the protons out to $L=12$. The electron model includes a pitch angle dependency within $L=16$ but is considered isotropic beyond that point. The proton model includes a pitch angle dependency within $L=12$. The following paragraphs describe the model characteristics.

For the inner electron and proton models, the independent variables magnetic L shell, local field strength B, pitch angle α with respect to the field line, and particle kinetic energy E were utilized (B and L are of course functions of r, λ , and l). The model populations are assumed independent of time, longitude, and direction azimuth about the field line, as appropriate for stably trapped populations. The mirror point field strength is given by (see earlier):

$$B_m = B(\sin \alpha)^{-2} \quad (39)$$

and the requirement that the intensity vanish for $B_m > B_c$ insures that the drift loss cone is empty for a pitch angle α close to 0° or 180° .

At each value of L, analytic expressions have been developed (these values are tabulated in Divine and Garrett^[45]), based on fits to the spacecraft and radio data, that allow evaluation of the particle integral and differential intensities I and i in terms of α , B, and E. That is, I has been fit to an analytic expression in α , B, and E at discrete values of L such that:

$$I_L = A_L(\alpha, B, E) \quad (40)$$

Variations at other values of L are derived by interpolating between the relevant values. The corresponding differential intensity (in units of $\text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}\text{MeV}^{-1}$) is then given by:

$$i = -\frac{dI}{dE} \quad (41)$$

Electron and proton omnidirectional fluxes are obtained by integration

$$J = 4\pi \int_0^{\pi/2} I(\sin \alpha) d\alpha \quad (42)$$

The spectra represented by equations (40) through (42) were approximately power laws. Values for the equations (40) through (42) were obtained by minimizing the weighted root-mean-square residual in the logarithm of the ratios of the model-predicted count rates to the count rates observed from several detectors at the relevant L values. Improvements in the fits were derived by matching the model predictions to UHF radio observations from Earth.

Figures 44 and 45 illustrate the dependence of the inner belt electrons and protons as defined by equations (40) through (42). The top panels are radial profiles of the omnidirectional integral flux as given by equation (42) evaluated for three energy thresholds at the magnetic equator. The lower panels represent the dependencies of the perpendicular integral intensity I_{\perp} (I at $\alpha=90^\circ$) on kinetic energy and on magnetic latitude. Figures 46 and 47 illustrate contours of constant omnidirectional integral flux for energies greater than 1 MeV for electrons and protons.

In the middle and outer Jovian magnetosphere (i.e., beyond $L=16$), the energetic particle fluxes are extremely time-dependent and are, as a result, difficult to model. However, to adequately estimate radiation effects in the jovian environment, a simple, isotropic formula for the energetic electron fluxes, based on Pioneer 10 and 11 observations, has been included in the model for completeness. This formulation assumes that the peak equatorial fluxes can be described by a function of the form:

$$\log J_O = f(t) - 2.2 \log r - 0.7 \log (0.03 E + E^3/r) \quad (43)$$

Here the electron kinetic energy E has units MeV, r has units R_J , and the omnidirectional integral flux J has units $\text{cm}^{-2}\text{s}^{-1}$. The term $f(t)$, which specifies the time dependence, is assumed to have an average value of 7.43. The maximum value given by equation 43 is assumed to occur along a disk surface at a height z_0 (in R_J) above the jovian equatorial plane given by[56]

$$z_0 = r \tan \alpha \cos (I_O) \quad (44)$$

for $L > 16$ and $r < 20 R_J$, and by

$$z_0 = r_O \tan \alpha \cos \{(I - I_O) - \omega (r - R_O)/V\} \quad (45)$$

for $r > 20 R_J$.

V_O , the "wave speed", is about $40 R_J/\text{hr}$ [56] which is, as might be expected, indistinguishable from the Alfvén speed, V_A , in this regime. The values for I_O , r_O , ω , and α (here α represents inclination of the magnetic axis, not pitch angle) are:

$$\begin{aligned} I_O &= 21^\circ \\ r_O &= 20 R_J \\ \omega &= V_A \cdot 0.9^\circ/R_J \\ \alpha &= 10.77^\circ \end{aligned}$$

The flux falls off away from this surface exponentially with a scale height of $2 R_J$.[56]

$$J = J_o \exp \left(- \frac{|r\lambda - z_o|}{2R_J} \right) \quad (46)$$

2.2.1.7.1.5. COMPARISONS WITH DATA

In this section, a few of the comparisons of the model with jovian data are presented. These comparisons will be limited to the radiation belt model. For the radiation belt models, Figure 48 shows the predicted profile of integral electron flux during the Voyager 1 flyby of Jupiter (for energy thresholds corresponding to two channels of the Voyager LECP) and the fluxes derived by the LECP investigators from their data[57]. The quantitative differences between the model and observed profiles are within a factor of two, and are representative of the uncertainty of the Pioneer and Voyager data sets and of the models themselves. The differences in the shapes of the peaks in these profiles suggest the magnitude of likely longitudinal or temporal variations, or of defects in the model or its data set. Figure 49 shows the predicted profile of synchrotron emission from the energetic electron model, for a wavelength of 21 cm and as would be observed from Earth at a distance of 4.04 AU and $\lambda = 110^\circ$ for a half-power beam width of 1.2 R_J, compared with observed data for two longitudes[58]. The predicted results reflect the fact that the model parameters, which control the electron pitch angle and latitude distributions inside the Pioneer 10 perijove (for L<2.85, they are not strongly constrained by the Pioneer 11 data), have been adjusted to better represent the data. Even so, the fact that the discrepancies are still about a factor of two, demonstrates that the match to the profiles and to the spectrum and polarization of the Earth-based UHF data is imperfect.

2.2.1.7.2. SATURN

Fig. 50^[59] is a schematic illustration of the postulated Saturn magnetosphere. As in the case of Jupiter, Divine^[60] has developed a first order radiation model for Saturn similar to that for Jupiter (this model, although unpublished, is available through JPL^[60]). Based on high energy data from Pioneer 11, Voyager 1, and Voyager 2, the model covers the distance from 2.3 to 13 R_S. It describes the electron distribution at energies between 0.04 and 10 MeV and the proton distribution between 0.14 and 80 MeV. As in the Jupiter model, the first step in the model is to specify the Saturnian magnetic field. Estimates for this field and other relevant quantities are listed in Table 11. Next, the integral and differential intensities for the electrons and protons, as functions of the magnetic field and L, are specified by algorithms very similar to those of the Jupiter model. The integral omnidirectional flux J is then calculated as before by Eq. 42.

The output of the model is presented in Figs. 51-54. The integral omnidirectional flux for the Saturn electron model at three energy thresholds are plotted in Fig. 51 (the dropouts are typically associated with the orbits of the Saturnian moons or rings). The fluxes at the magnetic equator ($B_e/B=1.0$) and at high magnetic latitude ($B_e/B=0$) are represented respectively by the solid and dashed lines. Integral omnidirectional flux for the electrons are plotted at 3 sample values of L in Fig. 52. On the left, the dependencies on local magnetic field strength B are shown for energy thresholds of 0.1, 1.0, and 10 MeV (solid, dashed, and solid lines) at each L. On the right are spectra at the magnetic equator (solid lines) and for high magnetic latitude (dashed lines) at each L. Similar plots for the protons are presented in Figs. 53 and 54 (for energies of 1, 10, and 100 MeV; solid, dashed, and solid).

2.2.2. AMBIENT PHOTON ENVIRONMENTS

2.2.2.1. GENERAL ENVIRONMENT

This document shall discuss the electromagnetic radiation environment in two frequency ranges (although there is an ambient γ -ray environment, this population does not contribute significantly to radiation effects on microelectronics and will not be considered here):

- 1) Optical (infrared, visible, UV): 10^{12} Hz- 10^{16} Hz
- 2) EUV/X-ray: 10^{16} Hz- 10^{21} Hz

First consider the infrared and visible spectrum. The visible (~ 3500 Å to ~ 7000 Å) and infrared (~ 0.7 mm to ~ 7 mm) portions of the spectrum are dominated by the solar flux. As shown in Figure 55^[61], the solar spectral irradiance peaks between ~ 4500 Å to ~ 7000 Å and accounts for most of the solar constant, the total solar energy flux just outside the Earth's atmosphere. It is currently estimated to be ~ 1370 W/m² ($\sim 8.4 \times 10^{11}$ MeV/cm²-s or $1.96 \text{ cal-cm}^{-2}\text{-min}^{-1}$). Although there are additional sources of visible and infrared light (such as light reflected by the Moon, atmospheric dayglow, infrared radiation from the Earth, and light from auroral displays), these contribute little to the penetrating radiation environment as defined in this report (they are, of course, critical to the detailed thermal balance of a spacecraft) and will not be discussed further.

Next consider the UV, EUV, and X-ray portion of the spectrum. As in the case of the visible and infrared portions of the spectrum, the Sun generally dominates the short wavelength portion of the spectrum in direct sunlight. The shortest wavelengths, 10 Å-100 Å or less, are referred to as X-rays. This spectral range is important because it contributes to the ionization of the E-region. The spectral region from about 100 Å to 1000 Å, called extreme ultraviolet or EUV, is related to the photoionization processes of O₂, N₂, and O in the ionosphere and to thermospheric heating. UV radiation is the continuum and line spectrum between roughly 1000 Å and 3500 Å. This spectral region contributes to photo-dissociation, absorption, and scattering processes in the mesosphere, stratosphere, and troposphere. The spectral range from 10-1750 Å is absorbed in the lower thermosphere and effects the production of oxygen atoms and their vertical distribution above the mesopause. The Lyman-alpha line at 1216 Å plays a major role in the mesosphere through the disassociation of O₂, H₂O, and CO₂ and the ionization of nitric oxide. The spectral region between 1750 to 2400 Å leads to the dissociation of O₂ and to ozone production in the mesosphere and stratosphere. Between 2400 Å and 3300 Å, the solar irradiance is responsible for the disassociation of ozone and other trace gases that play a role in the stratospheric budget.

UV, EUV, and X-ray radiation are not only important to atmospheric and ionospheric dynamics but, through material surface changes and photoelectron emission, are a major environmental factor in space system design at all altitudes. The energy in this spectral range is represented by a solar flux between 10^7 and 10^{10} photons/cm²-s below 1000 Å. The flux rises almost exponentially to 10^{16} photons/cm²-s between 1000 Å and 10000 Å. The flux is not constant but varies in time due to a number of factors, one of which is the solar cycle variability. This radiation spectrum is also a complex variable of the atmospheric attenuation as a spacecraft moves in and out of the Earth's shadow^[62].

Other forms of electromagnetic waves considered of importance to radiation analysis are γ -rays, Cerenkov radiation, and Bremsstrahlung radiation. Whereas X-rays are typically photons produced by inner electron shell transitions, γ -rays are photons normally associated with nuclear transitions. As such, although the higher energy X-rays

overlap with the lower energy γ -rays, the latter usually represent wavelengths shorter than .1 Å and frequencies above 10^{20} Hz. Electromagnetic radiation emitted when a charged particle is accelerated by the coulomb field of a nucleus, usually X- or γ -rays from relativistic electrons decelerating in a solid, is called Bremsstrahlung (or "stopping") radiation. This radiation is important as it is responsible for the radiation level below which it is impractical to shield a spacecraft (usually around 3 g/cm^2). The final form of electromagnetic radiation is Cerenkov radiation. It is the ultraviolet and visible light emitted when a charged particle passes through a transparent dielectric medium with a relative speed exceeding the phase velocity of light in the medium.

2.2.2.2. X-RAY FLARES

A primary source of naturally occurring X-rays are the so-called solar X-ray flares. An X-ray flare is associated with the sudden, short-lived brightening of a localized region in the solar chromosphere. X-ray flares nearly always occur in the plage of an active region, and are usually only visible in monochromatic light. They are classified according to their area (0,1,2,3,4) and their visible brightness (F,N,B), and/or by their X-ray intensity (C,M,X). The X-ray classification levels are, in terms of the peak X-ray emission in the 1-8 Å band:

- 1) C event: The peak emission is between 1.0 and $9.9 \times 10^{-3} \text{ ergs/cm}^2\text{-s}$. The actual level is given by the designator (i.e., C5.3 means the peak emission was $5.3 \times 10^{-3} \text{ ergs/cm}^2\text{-s}$). This is the least energetic solar X-ray flare level.
- 2) M event: The M event is a moderate level solar event and is characterized by an X-ray burst peak emission level between 1.0 and $9.9 \times 10^{-2} \text{ ergs/cm}^2\text{-s}$ (M6.6 means $6.6 \times 10^{-2} \text{ ergs/cm}^2\text{-s}$).
- 3) X event: This is a major solar flare as defined by the X-ray burst level. The peak X-ray emission must be in excess of $0.1 \text{ ergs/cm}^2\text{-s}$ (as before, X3.5 means the peak emission was $3.5 \times 10^{-1} \text{ ergs/cm}^2\text{-s}$).

A "typical" X-ray flare is plotted in Fig. 56^[63]. The data are from the November 26, 1982 X-class flare. Note the nearly exponential decay of the X-ray levels following the initial sharp rise.

2.2.3. MAN-MADE ENVIRONMENTS

2.2.3.1. EXOATMOSPHERIC NUCLEAR DETONATIONS

The exoatmospheric (here considered to be above 100 km) detonation of a nuclear weapon produces a variety of radiation environments and effects on spacecraft systems. Here, only a few of the more severe environments will be covered. The environments to be considered are those associated with prompt or transient radiation effects and those associated with the long term effects due to the debris and trapped electrons created by the bomb blast—EMP effects will be ignored. While the threat of nuclear war has faded in the last few years, the probability that a country could decide to carry out an atmospheric "test" has probably grown making these environments perhaps more of concern than ever before. Although the effects of the so-called "saturated nuclear environment" and the single "test" will both be discussed, it is clearly the latter effect that the spacecraft engineer will most likely have to deal with in the immediate future.

2.2.3.1.1. EVENT MORPHOLOGY

Consider first the overall event morphology of a nuclear detonation. In the general case of a single nuclear detonation above 100 km, the initial effect is the creation of a rapidly expanding plasma. This plasma consists of the nuclear fuel, the ionized remains of the bomb case and delivery vehicle, and the ambient atmosphere. The plasma, because of its charged state is confined to a bubble by the Earth's magnetic field. During the first 100 ns, the burst emits so-called "prompt radiation" consisting of γ -rays, X-rays, and neutrons. This radiation pulse expands radially falling off as $1/r^2$ along the line of sight. The delayed effects are associated with the subsequent evolution of the plasma bubble. The hot plasma bubble expands roughly spherically and rises following the initial detonation. Eventually the plasma cools and contracts leaving debris trapped on the magnetic field lines which beta decays emitting electrons. The debris either is eventually lost to the atmosphere or plates out on spacecraft or other space debris/junk. The high energy electrons drift eastward in the Earth's magnetic and electric field creating an artificial radiation belt. This belt is formed in about 6 hours following the initial detonation, after which it begins to decay. Depending on the intensity and location of the belt, it can last from months to years (the effects of belts created by the Starfish and Argus experiments are believed to have been observed through at least one solar cycle of 11 years).

2.2.3.1.2. THE PROMPT ENVIRONMENT

The primary effect of the nuclear weapon-produced X-rays is energy deposition and ionization. Indeed, approximately 70-80% of the energy is in the form of X-rays with ~1% of the energy in neutrons and γ -rays. In the case of low energy X-rays, if the energy deposition takes place sufficiently rapidly, the surface material can be vaporized or actually blown off. Internal stress due to this rapid heating can generate a shockwave in the material that will cause spallation from the interior or exterior surfaces and delamination of dissimilar materials. The intermediate energy X-rays produce similar effects to the low energy electrons, only in the interior of the spacecraft. In addition to these effects, the high energy X-rays can generate photocurrents at the individual device level causing upset or burnout. X-rays are also the major source of system generated electromagnetic pulse (SGEMP). A representative environment (in Cal-cm⁻² where 1 Cal = 4.18×10^7 ergs) versus distance from a 1 MT burst is plotted in Fig. 57.^[64] The X-ray flux falls off as r^{-2} . A typical energy spectrum is presented in Fig. 58.^[65]

γ -rays, because of their higher energy, can not be effectively shielded against. They are initially produced in the first $\sim 10^{-8}$ s by the interactions of the neutrons released during the fission and fusion processes with the nuclear reaction and bomb fragments. γ -rays are subsequently produced by the radiation decay of the nuclear fragments. Although the effects of the γ -rays are similar to the X-rays and fall off as r^{-2} , they are, because of their energy, far more penetrating. The prompt radiation environment (in rads(Si)) for γ -rays versus distance from a 1 MT burst is plotted in Fig. 57^[64] and a representative spectrum in Fig. 58.^[65]

Like the X-rays and γ -rays, neutrons travel radially outward from the detonation site (i.e., their effects fall off as r^{-2}). Unlike X-rays and γ -rays, they do not travel at the speed of light, but rather at a velocity proportional to the square root of their initial energy ($V=(2E/m)^{1/2}$). As a result, the highest energy neutrons reach the spacecraft

first with the lowest energy neutrons arriving last. The pulse duration at the spacecraft is therefore dependent on the distance between the spacecraft and the detonation—typical values are 10^{-3} s. As neutrons are high energy ($E > 1$ MeV) and have low cross-sections as they are not charged, they are very hard to shield against. The much longer duration of the neutron ionization pulse than that of the X-rays or γ -rays may be more damaging even though the others may be more intense. Despite this, displacement damage (described later) is the major effect of the neutrons. The neutron flux as a function of distance is given in Fig. 57^[64] and a typical neutron spectrum in Fig. 58.^[65] The neutron spectrum includes a peak at 14.7 MeV that is characteristic of fusion and is not found in fission spectra.

2.2.3.1.3. DEBRIS ENVIRONMENT

As discussed earlier, ionized debris are created from the vaporization of the weapon's explosion, the casing, the local atmosphere, and any other material in the vicinity of the explosion. While most of the material is constrained to the Earth's magnetic field lines, some of the debris can jet radially across field lines away from the detonation. The majority of the magnetically contained debris is lost to the upper atmosphere at the conjugate points. The small amount of debris that is trapped magnetically will produce electrons by beta decay. Spacecraft passing through the debris cloud will become coated with the debris which can induce a dose (typically small, though the dose rate may be temporarily high) through beta decay, alpha emission, γ radiation, and neutrons. The dose rate within the debris cloud can be roughly approximated by^[64, 66, 67]:

$$D \approx f \frac{5 \times 10^6 Y}{(1+t)^{1.15} V^{2/3}} \quad (47)$$

where:

- D = Dose rate (rad(Si)/s)
- f = Probability that the debris will stick
- V = Volume of debris cloud (km³)
- t = time after burst (s)
- Y = Yield of weapon (MT)

Far more important than the debris environment, is the trapped electron environment it creates. Beta decay of the radioactive debris produces large amounts of trapped electrons. This effect, termed the Argus Effect after the Argus high altitude bomb tests of 1958, can lead to the creation of artificial radiation belts. The extent and intensity of the belt so formed is highly dependent on the altitude, yield, and latitude of the detonation. Representative examples of the electron fluences produced by such belts are compared with the natural environment in Figure 59 for polar and equatorial orbits and various bomb scenarios.^[65, 68] The maximum case shown corresponds to the so-called saturated nuclear environment in which the total kinetic energy density of the radiation particles just equals that of the Earth's magnetic field (i.e., the radiation can no longer be contained by the Earth's magnetic field if the energy of the particles per unit volume exceeds the magnetic field energy per unit volume). The energy spectrum of the trapped, weapon-produced electrons is more energetic (or "harder") than the natural electron spectrum (Fig. 60^[68]). The electrons are typically energetic enough to produce secondary X-rays (actually Bremsstrahlung—see earlier) inside even a well-shielded spacecraft. Rate dependent effects associated with the trapped electrons can be

particularly troublesome as their dose rates are often orders of magnitude higher than the dose rates from naturally occurring electron fluxes. Unlike transient effects which occur only for an extremely short period immediately after the detonation, the duration of the effects due to the trapped electrons are integrated over many orbits.

2.2.3.2. NUCLEAR POWER SOURCES

2.2.3.2.1. RADIOISOTOPE THERMOELECTRIC GENERATORS (RTGs)

Interplanetary missions beyond the orbit of Mars typically can not efficiently utilize solar arrays—the solar energy density is too low. Rather it has become common practice for missions like Voyager, Galileo, Ulysses, and, soon, Cassini to employ RTGs. Here, as an example of the induced radiation environment typical of these systems, the Galileo RTGs will be characterized. It should be remembered, however, that this environment is very device-dependent and will need to be specified for each particular design.

Galileo has 2 RTGs (Fig. 61). These RTGs, called General Purpose Heat Source (GPHS) RTGs, are basically large cylinders containing golf-ball sized pellets of plutonium 238 (^{238}Pu).^[69-71] The ^{238}Pu produces heat by emitting alpha particles—the heat is converted directly into electricity by electric thermocouples. Each RTG contains ~11 kg of plutonium or about 137,000 curies (1 curie represents 3.7×10^{10} radioactive disintegrations per second)—this gives an intensity of about 7,000 n/s-g at the fuel cylinder level. The fuel is in the form of pure plutonium oxide (PuO_2) with 0.7 ppm ^{236}Pu and less than 0.5% ^{232}U and ^{228}Th (Table 12^[69]). The primary radiation environments of concern are neutrons and γ -rays. Table 13^[71] tabulates the normalized differential flux as a function of energy for both RTGs and RHUs. These normalized values must be multiplied by an absolute value that is a function of both distance and angle relative to the RTG. The absolute values, in terms of isodose-isoflux for neutrons and γ -rays, are plotted in Figs. 62 and 63 for a single GPHS RTG with 5 year old fuel (corresponding roughly to the Galileo RTGs at launch) and give absolute flux at a particular distance and angle. The radiation characteristics of the RTG plutonium fuel, because of its short half-life, change rapidly in time. This is reflected in the γ -ray values which need to be multiplied by the values in Fig. 64 to correct for this time factor. In addition, a factor of 2 is recommended as a "worst case" design margin.

2.2.3.2.2. RADIOISOTOPE HEATER UNIT (RHU)

RHUs are intended to provide localized heating within a spacecraft and minimize the use of electrical power. The Galileo design is based on a 1 W unit with approximately 34 curies (2.6 g) of plutonium fuel per unit. Each unit was ~5 years old at launch. Fig. 65 is a cutaway of a typical Galileo RHU. The PuO_2 fuel pellet is encapsulated within a clad of platinum-rhodium alloy for fuel containment. Like the RTGs, it is assumed that the plutonium is basically pure ^{238}Pu with less than 1.2 ppm of ^{236}Pu and less than 0.5% ^{232}U and ^{228}Th . The normalized fluxes in Table 13^[71] are to be multiplied by the values in Table 14^[71] to give absolute values. These are given in terms of distance from the RHU in cm. As for the RTGs, the γ -ray values must be corrected by a time factor (plotted in Fig. 66^[71]). Again, a factor of 2 radiation design margin is recommended for a "worst case".

2.3. SHIELDING EFFECTS AND INTERACTIONS WITH MATTER

From the standpoint of radiation interactions with matter, 3 "particle" families need to be considered:

- | | | |
|----|-------------------|--|
| 1) | photons | primarily EUV, X-ray, and γ -rays |
| 2) | charged particles | protons, electrons, and heavy ions |
| 3) | neutrons | |

While numerous, more exotic particles such as positrons, muons, mesons, etc. exist, these 3 families account for the vast majority of interactions of concern to the spacecraft engineer. In addition, for the impacting particles, mass, charge, and kinetic energy are the principle physical characteristics of interest whereas mass and density are the key characteristics for the target material. Here the various types of interactions will be discussed in terms of these 3 particle families. The effects of the shielding on these particles will be manifested in terms of energy deposited in a volume (dose) or the energy deposited per unit length in the target material (LET) after traversing a specified thickness of shielding. The radiation shielding calculation necessary to determine the environment inside a spacecraft thus breaks down into a 3 step process for each particle:

- 1) Definition of the ambient environment.
- 2) Propagation of that environment through the shield and calculation of the subsequent changes in the spectrum up to the target.
- 3) Estimation of the total energy and/or the energy deposition rate at the target.

The first part of this report detailed the ambient environment. In this section, the latter 2 issues will be addressed. An important factor that should kept in mind when considering these 2 issues is the importance of the cascade process to the final result. In this process, one incident particle produces many secondary particles that may be very different from the incident particle (for example, electrons may generate photons or vice versa). These secondaries in turn generate their own secondaries leading to a complex mix of many different photons and particles. This process repeats until the point of interest is reached or until all the initial particle energy is absorbed. Rather than address this process in its entirety, it will first be broken down into the individual, distinct single particle interactions. The final part of this section will describe how, given the characteristics of these individual interactions, Monte Carlo techniques can then be used to estimate the gross effects of the cascade process. In the simplest models, analytic expressions are fit to these results or to actual measurements to give estimates of the end products produced by the cascade as a function of depth in the shielding material. The reality is that the cascade process is basically probabilistic and too random to be precisely modeled analytically. However, given that the analytic fits typically give very adequate results in most cases of interest to the engineer, models based on their use will be addressed here for practical radiation shielding calculations.

2.3.1. SINGLE PARTICLE INTERACTIONS

The study of the interactions of a single high energy particle such as a photon, neutron, or charged particle with matter forms a major sub-division of the physical sciences. Rather than present a detailed quantitative review of each of these interactions, a qualitative description will be presented for each of the main interactions. This will be supplemented with a quantitative plot of the actual interaction where appropriate. The reader is referred to detailed quantitative reviews of each of the processes such as may be found in Refs. [1] and [72]. In most practical cases, however, the results presented here should suffice for actual computations as the complex

equations modeling the processes are normally pre-coded in the computers codes available for carrying out shielding calculations.

2.3.1.1. PHOTON INTERACTIONS

Photons, which propagate at the speed of light and have no charge or rest mass, interact primarily through the photoelectric effect, Compton scattering, and pair production. These interactions all generate free electrons. Consider first the photoelectron process, the probability of which decreases with increasing photon energy and increases with Z. In the photoelectron process, the photon is completely absorbed by the emitted (typically) outer shell electron. In one case, however, subsequent interactions are possible—that is, if the photon is energetic enough to emit K-shell electrons (inner shell electrons), then this process will dominate ~80% of the time over the emission of outer shell electrons. When an L-shell (or outer shell) electron subsequently drops down to fill the K-shell vacancy, it can emit either an additional X-ray or a low energy Auger electron from the L-shell (dependent on the Z of the material). In Compton scattering, the incident photon is not completely absorbed as the photon is of much greater energy than the atomic electron binding energy. Part of the photon energy goes to scattering the atomic electron (called a Compton electron) and the rest into a scattered, lower energy photon. Pair production takes place for photons at energies of 1.02 MeV or higher. A photon of this energy will be completely absorbed by a high-Z material. A positron-electron pair will then be formed. Figure 67^[73] compares the ranges over which each of the 3 interactions dominate as functions of Z and energy. For reference, in silicon, the photoelectron effect dominates at energies <50 keV, pair production for energies >20 MeV, and Compton scattering at intermediate energies. The products of these interactions (electrons, photons, and positrons) can of course further interact with the target material producing a complex cascade of electrons and photons.

2.3.1.2. CHARGED PARTICLE INTERACTIONS

Charged particles interact with matter primarily in 2 ways: Rutherford scattering and nuclear interactions. Rutherford (or Coulomb) scattering, in which the charged particle interacts with the electric field of the target atom, typically dominates. It results in both excitation and ionization of atomic electrons and can, for sufficiently energetic impacts, transfer enough energy to displace atoms within the lattice structure. As an example, for electrons, a minimum energy of ~150 KeV is required to cause displacement in silicon while only 100 eV is required for protons. Nuclear interactions, where the impacting particle actually interacts with the atomic nucleus, can result in elastic or inelastic scattering and transmutation (through fusion or fission). As an example, a nucleus can absorb a proton and emit an alpha particle. This process, also called spallation, and the recoil atoms that result from displacement can transform a relatively benign proton environment into a SEU-causing heavy ion environment as the heavy ions have much larger LETs compared with the protons. Also, long term exposure to the space radiation environment can, through transmutation, lead to making the spacecraft material itself radioactive.

One quantitative measure of the interaction of a high energy particle with matter is stopping power or energy loss per unit length in a given material. As an example, low energy electrons (~10 KeV) primarily cause ionization. The amount of energy deposited by the latter and protons in producing ionization can be determined from stopping power tables (electrons: Berger and Seltzer^[74]; protons: Janni^[75, 76]). Stopping power is essential in calculating the Heinrich flux necessary for most SEU calculations (see

later). Stopping power (or LET) in terms of MeV-cm²/g is given in Fig. 68 for electrons, protons, and various heavy ions in silicon.^[3]

A second quantitative measure of high energy particle interactions closely related to stopping power is the penetration depth/range or the maximum distance a particle of a given energy can penetrate. This depth can be used to roughly estimate the minimum cut-off energy for a given thickness of spacecraft shielding and hence its effectiveness. Fig. 69 compares the penetration depth of electrons and protons in aluminum for different energies (the range of heavier ions in Al are presented later in Figure 75). Note in particular that an electron at 1 MeV penetrates over 100 times more shielding (~0.2 cm) than a 1 MeV proton (~0.0015 cm). Similarly, it takes a ~20 MeV proton to penetrate the same depth as a 1 MeV electron. As ~0.1-0.2 cm (40-80 mils) is a typical shielding level, it is common to compare the integral dose for $E \geq 1$ MeV electrons with $E \geq 20$ MeV protons as these are the primary contributors to the radiation environment behind the spacecraft shield.

2.3.1.3. NEUTRON INTERACTIONS

Neutron interactions with matter can result in elastic scattering, inelastic scattering, and nuclear transmutation. In elastic scattering, the neutron is not captured but transfers some of its energy to the target atom which can be displaced from its lattice position. It occurs only if the incident neutron has more energy than required for displacement—typically >25 eV. The target atom, referred to as a primary recoil or knock-on, can in turn cause ionization or further displacement damage. In inelastic scattering, the neutron is captured by the nucleus followed by emission of a lower energy neutron. The kinetic energy lost in the process can result in displacement or can excite the atomic nucleus which returns to ground state by emitting a γ -ray. In transmutation, the capture of the neutron can change the atomic isotope, cause fission, or cause the emission of another particle such as a proton or alpha particle. For silicon, the dominate process is displacement and ionization for neutrons with energy ~1 MeV or higher (see Figure 3 which plots relative neutron displacement damage versus energy). The effects of neutron dose are typically not considered in radiation calculations except for the prompt nuclear environment or for evaluating the results of ground testing.

2.3.2. MODELING THE EFFECTS OF SHIELDING

The preceding section has briefly described the basic interactions between single particles and matter. If the detailed evolution of a particle passing through matter is followed, the interaction of the particle with shielding becomes increasingly complex as each interaction gives rise to a cascade of by-products. Fortunately, as each interaction disperses the energy into more by-products, a point is reached where the by-products and the original incident particle (if it still exists) no longer have sufficient energy to excite further interactions—the process has a finite conclusion. Although it is common practice using Monte Carlo techniques to model the detailed passage of a particle through shielding, the computer codes that accomplish this can require super computers or take many hours to carry out the calculations. As a consequence, it is common to run the detailed codes only for a range of variables and then use this information to derive analytic fits to the end products of the multiple particle interactions that are created following a single particle impact. The effects of these by-products are then approximated roughly in terms of displacement damage, energy deposition, or ionization (or electron-hole creation). It is normally these algorithms, not the detailed computations, that are used to actually model radiation effects.

As a specific example of the Monte Carlo results, consider electrons. Electrons are particularly easily scattered in a material. Rather than passing through the material, they and the secondary electrons they generate are scattered into the material. This behavior is illustrated in Fig. 70a which is a computer simulation (Monte Carlo) of the trajectories of 5 electrons impacting on an "infinitely thick" copper target.^[77, 78] Fig. 70b^[77, 78] shows the results for 100 electrons—note that many of the electrons are actually scattered back out of the surface of the material. This behavior becomes ever more complex as the thickness of the shield decreases as illustrated in Fig. 71.^[77, 78] It is readily apparent in these Monte Carlo simulations that the dose is very dependent on the shape (or thickness) of the shield. This scattering of the electrons and their by-products by the shielding means that the details of geometry of the shielding must be considered in any radiation calculations.

Analytic expressions (as an example of this for electrons, see Ref. [79]) have been fit to the results of Monte Carlo calculations (Figs. 70 and 71) and to actual measurements for electrons, protons, heavy ions, neutrons, and photons. Specific characteristics, such as energy deposition, ionization, flux (both forward and back-scattered), dose, etc., can then be predicted as functions of shielding thickness or material. Fig. 72^[80] illustrates this effect for one characteristic—the electron dose versus distance into the shielding material as the incident electron energy is increased. Here the region over which the electron deposits its energy is smeared out along track. Contrast this with Fig. 73^[81] for a proton—a high energy ion deposits its energy primarily near the end of its track. This difference in energy deposition with shielding thickness is often used in designing solid state particle detectors capable of discriminating between high energy electrons and protons. It also must be kept in mind when designing shielding as too much shielding can actually cause cosmic rays of a particular energy to deposit most of that energy at a specific point in a device rather than passing through it. Families of similar parametric curves have been developed that allow rapid calculations of the effects of shielding (the lengthy Monte Carlo calculations are, however, often retained in detailed shielding studies).

These considerations allow a simple description of how the radiation dose environment is determined inside a spacecraft. If that environment is limited to dosage, then the following, very simple 1-dimensional model describes the basic mathematical steps involved:

For dose, assume a target of density ρ , area δA , and thickness $\delta t = \frac{M}{\rho \delta A}$:

- 1) First determine the attenuation effects of the shielding on the ambient flux spectrum of the particles of interest.
- 2) Compute the attenuated flux (number N of particles per unit area δA normal to the surface) versus energy at the target surface. Call this $f(E)$ at energy E such that:

$$f(E) = \frac{N(E)}{\delta A} \quad (48)$$

- 3) Estimate (i.e., from Fig. 72) the change in energy δE in crossing the target thickness δt at the appropriate distance in the shield for a particle of initial energy E :

$$\delta E \approx \delta \tau \frac{dE}{dx} \Big|_E \quad (49)$$

5) The dose per particle of energy E is given by :

$$D(E) \approx \frac{\delta E}{M} \approx \frac{dE}{dx} \Big|_E \frac{\delta \tau}{M} = \frac{1}{\rho} \frac{dE}{dx} \Big|_E \frac{1}{\delta A} \quad (50)$$

6) The total dose at energy E is then given by:

$$D_T(E) \approx N D(E) \approx \frac{1}{\rho} \frac{dE}{dx} \Big|_E f(E) \quad (51)$$

7) The total dose for $E > E_0$ is then given by integrating Eq. 51 over the range E_0 to ∞ .

This process, repeated for many different angles and particles, gives the total dose inside a 3-dimensional volume. The final answer is basically independent of the shape or size of the test point and is only a function of the density of the material.

In actual dosage calculations, because of the various effects of shielding on the energy deposition, 5 shielding geometries are typically considered (Fig. 74). These 5 geometries (as adapted from the descriptions provided by T. Jordan for his "NOVICE" shielding code^[82]) are:

- 1.) Spherical Shell: As the name implies, this configuration represents a hollow sphere of equal thickness in every direction from the dose site which is at the center of the sphere (note: the radius of the sphere void can be shown to be unimportant for large distances). The dosage tends to be lower than for a solid sphere of the same shield thickness. This case resembles a point inside a typical hollow spacecraft.
- 2.) Sphere: The shield is assumed to be uniformly distributed around the dose site with no gap between the shielding material and the dose site (i.e., a point at the center of a solid metal sphere). As scattering takes place relatively close to the dose site, little scattered flux is lost. This case resembles a "spot shield" configuration.
- 3.) Slab (or 2*Slab): A single slab is assumed to be an infinite 2-dimensional surface (Fig. 73). Ideally, particles enter from one side and irradiate the dose site. That is the basic "Single Slab" configuration which assumes no back scattering of electrons and no flux from behind (i.e., infinite back shield)—this approximates the actual case for high energy protons and heavy ions. To estimate the omnidirectional flux for a part between 2 shield planes without back scattering from a second surface, this value is typically doubled (in the NOVICE code, this is called the 2*Slab case; see Double Slab case also).
- 4.) Back Slab: This configuration is similar to the Slab in that the dose site is again assumed to be backed up by an infinite slab. As before, the flux only comes in from one side but now particles can be reflected or scattered back. This often nearly doubles the incident flux for electrons.
- 5.) Double Slab: Here there are 2 identical thin shields—one on each side. In this configuration, the flux is assumed to come from both sides and back scattering from

each is included. This geometry would resemble the case of a flat solar array panel extending out from the spacecraft in a wing configuration.

Which configuration to use depends greatly on the geometry of the spacecraft component being modeled. The spherical shell is often used as the baseline representation as it more closely resembles the shielding around typical circuit boards in the spacecraft interior.

Consider next the detailed steps involved in determining the other radiation quantity of interest, the Heinrich curve for a heavy ion. As outlined in Adams et al.[32] for the CREME code, the steps are:

- 1) The first step is to define the particle spectrum of interest at the surface of the critical volume. In this example, consider the ambient environment for GCR iron at the surface of a spacecraft. Fig. 2 is a plot of the ambient GCR iron flux as a function of energy for 3 cases: 90% worst case, solar minimum, and solar maximum. Call this differential spectrum $f(E)$.
- 2) The attenuation of a high energy ion by shielding can be approximated by (spallation is ignored[32]):

$$f'(E') = f(E) \left[\frac{S(E)}{S(E')} \right] e^{-\sigma \tau} \quad (52)$$

where:

$$\sigma = \frac{5 \times 10^{-26} \eta (A^{1/3} + 27^{1/3} - 0.4)^2}{27} \quad (53)$$

f' = Differential spectrum inside shielding

τ = thickness of shielding

E' = Energy inside spacecraft ($= R^{-1}\{R(E)\tau\}$)

R = Range through shield of ion of energy E (See Fig. 75 for Fe in Al)

R^{-1} = Inverse function of $R(E)$

S = Stopping power of ion in target material (See Fig. 68 for Fe in Si)

A = Atomic mass of ion

η = Avogadro's number

Fig. 76 shows the results of this calculation for Fe behind 0.025" of Al shielding.[32]

- 3) Next, the dE/dx curve for the incident particle species in the material of interest is determined. The $-dE/dx$ curve for Fe in Si is plotted in Fig. 68. Define this curve as $H(L)$ where L is the LET of the particle.
- 4) The incident (internal) differential particle spectrum $f'(E')$ is converted to the differential Heinrich spectrum $h(L)$ by:

$$h(L) = f'(E') \left[\frac{dE'}{dL} \right] \quad (54)$$

This result for Fe impacting Si is presented in Fig. 77 (the spikes at the ends of the curve are numeric and result from dL/dE going to 0 which implies that dE/dL goes to ∞ - a careful analytic evaluation would give a finite value).

- 5) Eq. 54 is integrated over L to give the integral Heinrich LET curve, F_H (note: this is equivalent to Eq. 7):

$$F_H(L) = \int_L^{\infty} h(L)dL \quad (55)$$

The final results are plotted in Fig. 78. These curves, called Heinrich curves, are the type of information normally required for SEU calculations. In the case of GCR, this curve would be calculated for all the GCR and summed to give a final result (see next section for an example).

To summarize, there are many different techniques for estimating the radiation environment behind a spacecraft shield. To limit the amount of computer time required, the "exact" Monte Carlo formalism is often replaced by analytic approximations (called "kernels") when performing the particle transport and shielding calculations. Specifically, tabulated attenuation data, using Monte Carlo techniques, are prepared for various shield geometries (i.e., the slab, spherical shell, and solid sphere geometries illustrated in Fig. 74). Given a 3-D model of the shielding mass and geometry (or a 1-D configuration, depending on the desired level of accuracy), the equivalent shielding at a point as a function of angle and path length is calculated. The input spectra from the environment (neutrons, γ -rays, photons, electrons, positrons, protons, heavy ions, alphas, GCR, etc.) are convolved with this equivalent shielding to calculate the dose (or Heinrich flux) as a function of energy (or LET) and angle. Secondary and Bremsstrahlung particle effects also normally need to be included, particularly for thick shielding. For a more detailed treatment of these interactions see Refs. [1, 72, 83-85], and others. Computer codes for carrying out detailed transport and shielding calculations are available from the Oak Ridge National Laboratory's Radiation Shielding Information Center (RSIC) and various commercial vendors and government laboratories. (The RSIC address is ORNL, Box X, Oak Ridge, Tenn. 37831-6362, Telephone 615-574-6176 or FTS 624-6176.)

2.4. RADIATION ENVIRONMENT ESTIMATES

In this section, we will complete the analysis of the spacecraft radiation environment by combining our knowledge of the ambient radiation environment with the transport/shielding process to examine several practical case studies. These will be, in order, a detailed analysis of the radiation dose environment to be expected for a lunar transfer mission, a comparison of the intense radiation environment at Jupiter with the saturated nuclear environment at the Earth, and a comparison of the effects of shielding and orbit on solar flare and GCR Heinrich fluxes. Each of these case studies will illustrate the basic steps required in carrying out a thorough analysis of the radiation environment inside a spacecraft—the purpose of this review.

2.4.1. EXAMPLE—THE CLEMENTINE PROGRAM

As an illustration of the process of estimating the radiation dose environment within a satellite, consider the case of the Clementine spacecraft lunar transfer orbit sequence. Clementine is an ambitious DoD/NASA mission designed to map the Moon and an Earth orbit-crossing asteroid. It is the first mission by the US to the Moon in over 20 years and will test the effects of the radiation environment on a number of unique, advanced microelectronic systems. The Clementine will also leave behind its lunar transfer stage in a unique, highly elliptical orbit. This interstage and the Clementine

have both been instrumented with radiation dosage and SEU detectors. In addition, each will carry boxes of advanced microelectronics components for direct exposure to the radiation environment. A detailed radiation environment prediction is required in order to allow the identification of radiation sensitive parts and to determine appropriate replacement parts or provide enhanced protective measures. It is also desired to predict the performance of the systems and test components in the radiation environment. These are all typical requirements for a space mission and illustrate the wide range of potential radiation applications.

2.4.1.1. AE8/AP8 RADIATION DOSAGE RESULTS

As has been discussed, the most often used trapped radiation environment models are the GSFC AE8 (electron) and AP8 (proton) solar maximum (or active) and solar minimum (or quiet) trapped radiation models. These models give dosage results that, when averaged over mission lifetimes of the order of the solar cycle, are typically within a factor of 2 of the actual measured dosages. Unfortunately, for time periods shorter than about 5 years, the statistical variations can be great (approaching factors of $10^{\pm 1}$ to $10^{\pm 2}$ for missions of less than a year, Figure 16). Even so, they have formed the basis of almost all trapped radiation calculations since the late 60's. With a properly defined radiation design margin (RDM), their predictions are useful in evaluating a spacecraft radiation hardness design. (Note: the new total dose model for solar maximum based on the CRRES data^[14] became available too late to be used here.)

As has been described, the process of calculating the dosage at the interior of a spacecraft is straightforward but time consuming. First, the B and L coordinates of the spacecraft are estimated from the orbit. The particle integral flux as a function of energy is then computed in terms of B and L from the AE8/AP8 models. The resulting spectra are summed over mission time to give the total integral fluence spectrum in terms of energy and particle species. These spectra, by species, are then used as input to the shielding code (T. Jordan's "NOVICE" code, a commercially available software package, was used here^[82]) which computes the total energy deposited at a point as a function of shield thickness, shield composition, and geometry. Typically, for dosage calculations to be used in a first order estimate of the internal radiation environment, aluminum for the shield and silicon for the dose site are assumed for composition. Geometrical considerations become particularly important for electrons as the electrons can be easily scattered or reflected within the material. As described in earlier sections, several different geometries are usually assumed. Here, 5 geometries are considered (see Fig. 74):

- 1.) Spherical Shell
- 2.) Sphere
- 3.) 2*Slab
- 4.) Back Slab
- 5.) Dubl Slab (Double Slab)

In the NOVICE calculations, the single slab is assumed to be an infinite 2-dimensional surface (Fig. 74) with an infinite back shield-no radiation comes from behind the shield and none is reflected back. For comparison purposes, the code doubles this value (hence, "2*Slab") so that the results can be used to estimate the omnidirectional flux for a part between 2 slabs without scattering (see Dubl Slab case). Unless stated otherwise, the spherical shell geometry will be assumed as the baseline

representation as it more closely resembles the shielding around typical circuit boards in the Clementine interior.

A set of orbital data for the portions of the Clementine orbit within 11 Earth radii was assembled (note: the Clementine orbit, as is true for most missions, is undergoing continual revision; the calculations here represent only one of several options being considered). The Clementine orbit has been divided up into 6 segments. These are:

1.) 1994/01/24 (01:48) -- 1994/01/26 (01:00)	Initial Parking Orbit
2.) 1994/01/26 (01:00) -- 1994/01/30 (06:44)	Earth-Moon Transfer Injection
3.) 1994/01/30 (07:00) -- 1994/02/09 (12:00)	Earth-Moon Transfer Orbit
4.) 1994/02/09 (12:00) -- 1994/02/21 (01:09)	Earth-Moon Transfer Orbit
5.) 1994/05/03 (13:17) -- 1994/05/15 (12:00)	Earth Gravity Assist Flyby
6.) 1994/05/15 (12:00) -- 1994/05/27 (18:08)	Earth Gravity Assist Flyby

For these orbits, the dosage for the 5 different geometric configurations, solar maximum and solar minimum environmental assumptions, and various shielding thicknesses were calculated so as to provide an estimate of the range of doses to be expected. Radiation doses for electrons, protons, and photons (secondary particles) were calculated for all segments when Clementine would be inside $\sim 11 R_E$ (the AE8 models are only useful inside $\sim 11 R_E$ and the AP8 models within $\sim 6 R_E$). The results in terms of dosage are presented in Figs. 79 and 80 for 2 worst case orbital segments—the trapped electrons at solar maximum and protons at solar minimum and the first transfer orbit—and are tabulated in Table 15 for 58 mils of aluminum shielding. Note the wide variation in results for the electrons—as has been discussed, the precise geometric assumptions can make a big difference in results. The main difference between the proton results is that between the 2-D and 3-D geometric assumptions—plane versus sphere. In many cases, the spherical shell geometry is the most appropriate as it resembles the structure of a spacecraft surrounding circuit boards in the interior of the spacecraft. Spherical shell results for the Clementine interstage mission are summarized in Fig. 81 (the interstage will have a perigee of ~ 500 km, an apogee of $\sim 160,000$ km, and an inclination of 67°). Note that the proton dose due to trapped radiation is very low in comparison to the trapped electrons for the interstage.

2.4.1.2. SOLAR FLARE PROTONS

The solar flare proton environment will potentially be the most severe radiation environment for a lunar transfer orbit like that for Clementine—the spacecraft typically spends little time in the trapped radiation environment as this occurs near perigee—more time is spent at apogee far from the Earth and more directly exposed to the solar flare environment. The maximum dosage expected, as predicted by the JPL model at 1 AU for solar flare protons, is presented in Figs. 82 and 83 in terms of the probability of the plotted values not being exceeded for missions of 1 to 7 year durations during the active part of the solar cycle. The model predicts essentially ~ 0 dosage for the 4 years of solar minimum (roughly May 1995 to May 1999). The results are summarized in Table 16 for a 1 year mission within the current active solar cycle (the main Clementine mission will be from February 1994 to August 1994 while the interstage will last about 450 days; for reference, the next maximum is currently assumed to start on May 26, 1999 and to last 7 years). The JPL model, unlike its predecessors, provides total mission dosage as a function of expectation. Specifically, it predicts the likelihood (expressed as a percentage) that a specified maximum integral fluence will not be exceeded in a given time interval (or, vice versa, for a specified likelihood percentage, the corresponding maximum fluence expected). For the 1 year Clementine mission during solar maximum,

the model would predict a dosage equal to or less than 40,000 rads (Si) 99% of the time behind 20 mils of shielding or less than or equal to 2,000 rads (Si) 50% of the time. This method allows a project to establish mission dose levels based on the desired probability of success. (Please refer to Section 2.2.1.4 for a more detailed explanation of the JPL solar flare model.)

Timing can be everything in the radiation world and the Clementine mission is no exception. Clementine, flying as it does 4 years after solar maximum, may catch the last year of the present solar maximum period (it should be kept in mind that the 1972 solar flare, perhaps the worst we have seen in the space age, occurred almost precisely 4 years after solar maximum!). In that case, the maximum dosage expected as a function of likelihood of not being exceeded is given by Table 16 (assuming ~1 year and 58 mils of shielding). The new proton flare model provides design guidelines as to the maximum dosage that can be expected—it is not a hard and fast standard. As such, it can not be used without a properly specified RDM. The RDM is where the project management makes its input—basically the "risk" margin that the project is willing to assume for a solar flare. A typical RDM is a factor of 2 to 4 although higher values are also common.

2.4.1.3. SUMMARY OF RADIATION DOSAGE ESTIMATES

This section has reviewed the radiation environment for a proposed mission, Clementine and its interstage. This exercise has served to demonstrate the practical steps required in carrying out an estimation of the radiation environment within a spacecraft. For this environment, each of the major contributors to radiation dosage (i.e., trapped electrons, trapped protons, and solar flare protons) have been identified. Estimates of the total dose (assuming a 58 mil aluminum shield and a spherical shell) for the trapped environment range from 900 to 1000 rad(Si) for the electrons and ~200 rad(Si) for the protons (as the mission duration is so short, however, deviations of a factor of ~10 are easily likely!). The solar flare proton dosage is, for a 90% probability that the value won't be exceeded, 4600 rad(Si). The corresponding dosages for the interstage give a total dosage of ~10,000 rad(Si) for the electrons, ~100 rad(Si) for the protons, and 4600 rad(Si) for the flare protons (95%). Whatever levels are adopted, it is always necessary to establish a project-defined radiation design margin (RDM). For an RDM of 2 times the predicted dose, this would give ~12 krad(Si) as the maximum design dose for a 58 mil aluminum shell for the Clementine and 30 krad (Si) for the interstage. As a final issue, shielding geometry is clearly a major driver in determining dose, particularly for the electrons. For particularly sensitive components, it is recommended that detailed dosage calculations always be carried out if the initial levels estimated by this process are a cause for concern. Spot shielding and clever placement of the instruments can significantly lower the dosage that these components see without necessitating their replacement—all steps being followed on the Clementine mission.

2.4.2. JOVIAN MODEL APPLICATIONS

The jovian radiation model^[45] was developed for several practical applications. These applications, as would be expected, ranged from establishing radiation dosage guidelines to single event upset modeling. 2 unique applications will be discussed here. First consider the effects of total dose. In Fig. 84, the model predictions for the total electron dose expected for Galileo as a function of different levels of aluminum shielding and 2 time intervals—closest encounter plus 5 orbits and after the 2 year extended orbit (closest encounter plus 13 orbits)—are graphed. These predictions are also compared with the range of dosages expected for a spacecraft in a circular orbit at various altitudes

above the Earth in the saturated nuclear environment after 1 year. The typical shielding level on Galileo and a hardened DoD spacecraft is at least 3 g-cm^{-2} . The hardness levels of the Galileo electronic components is $\sim 150 \text{ krads(Si)}$, commensurate with this level of shielding. The use and impact of the model in determining the dose levels for the radiation hardness design of Galileo is obvious. It is also very interesting that these results are comparable with the 1-year nuclear saturated environment for the Earth—at 3 g-cm^{-2} , the levels of dosage are remarkably similar and imply that some radiation environments at least as severe as the saturated nuclear environment exist in space.

Another application of the Jupiter radiation model was in studying anomalies on Voyager. During the Voyager I flyby of Jupiter, 42 anomalies were observed. It has been postulated^[86] that these anomalies were caused by arc discharges during the flyby. Although surface charging was ruled out as a possible cause, subsequent estimates of the total electron fluence indicated that there was a sufficient radiation environment of high energy electrons present to cause internal electron charging. That is, it has been proposed that the $E > \sim 100 \text{ KeV}$ electron environment deposited sufficient charge in an external cable on Voyager to generate arc discharges that could cause POR anomalies. To test this assumption, the total fluence of electrons at $E > 1 \text{ MeV}$ and $E > 10 \text{ MeV}$ and protons between $15 < E < 26 \text{ MeV}$ were computed as a function of time. The resulting normalized curves are plotted versus the cumulative sum of Voyager POR anomalies in Fig. 85 (cumulative was used as the charge buildup is a cumulative process). As this figure implies, internal electron-caused charging is a possible source of the Voyager anomalies in a temporal sense as the energetic electron fluence roughly follows the pattern of POR events. The major evidence for buried charge as the cause, however, came from an estimate of the charge deposited in each arc that would be necessary to cause the observed POR upsets. This indicated that the total $E > 100 \text{ KeV}$ (i.e., the minimum energy necessary to penetrate several mils of shielding and deposit charge) electron fluence was very close to the actual level required to account for all 42 arcs. This is believed to be one of the earliest examples of internal arc discharging occurring in space ever published and demonstrates another application of radiation modeling. Subsequent ground testing of the Galileo vehicle supported the buried charge hypothesis and revealed several potential sources of such arcing that led to its redesign.

2.4.3. HEINRICH FLUX ESTIMATES

The Heinrich flux curve is the starting point for most SEU calculations. Here, several practical examples of the effects of source, location, and shielding will be presented. These calculations are based primarily on the results of Adams^[87]. Fig. 86 is the first example and illustrates the effects of shielding on the solar flare and GCR environments in interplanetary space. Fig. 86a is a Heinrich curve showing the effects aluminum shielding (0.173 g-cm^{-2} (0.064 cm) to 10.8 g-cm^{-2} (4.0 cm)) has on the Adams worst case composite flare. This flare has the same spectrum as the August 4, 1972 flare and a 90% worst case enrichment of the heavy ions.^[28] Fig. 86b is a similar plot for the effects of shielding on the GCR (elements 1 to 26). The aluminum shielding in this example ranges from 2 g-cm^{-2} ($.74 \text{ cm}$) to 100 g-cm^{-2} (37 cm)-a very thick shield! It is clear from these plots that the solar flare environment, dominated by protons, can be fairly effectively attenuated by shielding whereas the GCR environment is little effected even for the lighter components. Even so, the GCR environment does not begin to dominate until LETs well above $20\text{-}30 \text{ MeV-cm}^2\text{-mg}^{-1}$.

Next consider the effects of magnetospheric shielding. In Fig. 87^[87], the Heinrich integral flux diagram for a spacecraft with 0.064 cm of aluminum shielding is graphed as

a function of orbital inclination at an altitude of 400 km. A 90% (Adams) worst case flare and an GSFC AP trapped proton environment are assumed. As noted by Adams^[87], below an LET of .5 MeV-cm²-mg⁻¹, the population is dominated by the trapped protons. As the LET increases and the inclination increases, more and more flare particles can penetrate to the spacecraft. The steps correspond to the large drops in abundance beyond iron, nickel, bismuth, and uranium in the Adams flare model heavy ions. Fig. 88, for the same environment, illustrates the variations produced by increasing the altitude for a 60° inclination orbit and for 0.064 cm of aluminum shielding. What variations exist are primarily below 0.5 MeV-cm²-mg⁻¹ and are due almost entirely to the trapped proton environment. These variations go through a maximum between 4000 and 6000 km.

Of what value are these results? First, they predict that, even for substantial levels of shielding, the solar flare environment normally will dominate SEU calculations even at high LET levels in interplanetary space. Indeed, these figures provide upper limits on the SEU environment expected at 1 AU (for no magnetospheric shielding). Second, given the increasing importance of solar flares to the overall SEU environment, statistical techniques such as represented by the JPL solar flare models will need to be utilized for predicting SEU effects. Third, in Earth orbit, the picture is very complex because of geomagnetic shielding, but, at low LETs (i.e., LET < 5-10 MeV-cm²-mg⁻¹), the trapped protons will normally dominate the environment. Unfortunately, with the desire to fly "commercial parts" and CCDs, these LET levels, which have previously been too low to be of concern, are beginning to become an issue.

2.5. CONCLUSION

The objectives of this section were to address 3 aspects of the radiation problem. First, a review was provided of the natural and man-made space radiation environments. This description included both particulate and, where applicable, electromagnetic (i.e., photon) environments. Secondly, the methods used to propagate the external environment through the complex spacecraft structures surrounding the point where the internal radiation environment was required were discussed. Finally, examples of the environment inside the spacecraft were presented. While it was not intended to treat in detail all aspects of the problem of the radiation environment within a spacecraft, by dividing the problem into these 3 parts—external environment, propagation, and internal environment—a basis for understanding the process of predicting the internal spacecraft radiation environment has been established. The consequences of this environment are discussed by the other presenters at this seminar.

2.6. ACKNOWLEDGEMENTS

We would like to acknowledge several sources of information not explicitly referenced in this report. First, Dr. M. Ratliff of SYSCON carried out many of the Clementine and solar flare calculations presented. Several of the tables and the information on heavy ions were made available on disk by Dr. P.A. Robinson of JPL. One of us (H. Garrett) would like to thank his wife, Katherine Garrett, for typing many of the references and captions and for her patience during the several months of preparation of the manuscript. Particular thanks go to A. Johnston of JPL for soliciting this paper and for helping with many of the figures and preparations required to prepare the paper for publication.

2.7. TABLES

TABLE 1. Spherical-Harmonic Coefficients for the IGRF 1980.^[13]

n	m	g	h	dg/dt	dh/dt	n	m	g	h	dg/dt	dh/dt
1	0	-29988		22.4		8	0	20		0.8	
1	1	-1957	5606	11.3	-15.9	8	1	7	7	-0.2	-0.1
						8	2	1	-18	-0.3	-0.7
2	0	-1997		-18.3		8	3	-11	4	0.3	0.0
2	1	3028	-2129	3.2	-12.7	8	4	-7	-22	-0.8	-0.8
2	2	1662	-199	7.0	-25.2	8	5	4	9	-0.2	0.2
						8	6	3	16	0.7	0.2
3	0	1279		0.0		8	7	7	-13	-0.3	-1.1
3	1	-2181	-335	-6.5	0.2	8	8	-1	-15	1.2	0.8
3	2	1251	271	-0.7	2.7						
3	3	833	-252	1.0	-7.9	9	0	6			
						9	1	11	-21		
4	0	938		-1.4		9	2	2	16		
4	1	783	212	-1.4	4.6	9	3	-12	9		
4	2	398	-257	-8.2	1.6	9	4	9	-5		
4	3	-419	53	-1.8	2.9	9	5	-3	-7		
4	4	199	-298	-5.0	0.4	9	6	-1	9		
						9	7	7	10		
5	0	-219		1.5		9	8	1	-6		
5	1	357	46	0.4	1.8	9	9	-5	2		
5	2	261	149	-0.8	-0.4						
5	3	-74	-150	-3.3	0.0	10	0	-3			
5	4	-162	-78	0.2	1.3	10	1	-4	1		
5	5	-48	92	1.4	2.1	10	2	2	1		
						10	3	-5	2		
6	0	49		0.4		10	4	-2	5		
6	1	65	-15	0.0	-0.5	10	5	5	-4		
6	2	42	93	3.4	-1.4	10	6	3	-1		
6	3	-192	71	0.8	0.0	10	7	1	-2		
6	4	4	-43	0.8	-1.6	10	8	2	4		
6	5	14	-2	0.3	0.5	10	9	3	-1		
6	6	-108	17	-0.1	0.0	10	10	0	-6		
7	0	70		-1.0							
7	1	-59	-83	-0.8	-0.4						
7	2	2	-28	0.4	0.4						
7	3	20	-5	0.5	0.2						
7	4	-13	16	1.6	1.4						
7	5	1	18	0.1	-0.5						
7	6	11	-23	0.1	-0.1						
7	7	-2	-10	0.0	1.1						

TABLE 2. Comparison of the Composition of Heavy Ions in the Magnetosphere with the Galactic Cosmic Rays and the GCR Anomalous Component.[88]

Element	Relative Composition ($O^{16} = 1.0$)		
	Anomalous Component	Magnetosphere	Galactic CR
C	$0.230 \pm .090$	$0.200 \pm .019$	$1.130 \pm .030$
N	$0.220 \pm .090$	$0.210 \pm .041$	$0.270 \pm .020$
O	1.000	1.000	1.000
Ne	$0.070 \pm .040$	$0.080 \pm .020$	$0.180 \pm .010$
Mg	$0.002 \pm .002$	$0.006 \pm .004$	$0.200 \pm .010$
Si	<.020	$0.004 \pm .002$	$0.140 \pm .006$
S	---	<0.004	$0.035 \pm .003$
Ar	---	<0.003	$0.013 \pm .002$
Fe Group	---	$0.050 \pm .020$	$0.084 \pm .001$

TABLE 3. Proton Event Classification System

Date	PECS	Maximum Neutron Monitor Increase (Percent)	Maximum 30 MHz Riometer Absorption (db)	Peak Flux ($\text{cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$) (directional) $> 30 \text{ MeV}$	Integrated Intensity (cm^{-2}) $> 30 \text{ MeV}$
1968 Oct 4-6	1 2 0	—	1.5	6.3	2.6×10^5
1968 Oct 31	2 3 0	—	5.5	10.0	$[1.2 \times 10^6]^{\dagger}$
1968 Nov 1	2 3 0	—	4	11.7	
1968 Nov 18	2 3 3	14	12.5	404.0	1.7×10^7
1968 Dec 3	1 2 0	—	(slight)	1.7	$[3.2 \times 10^6]^{\dagger}$
1968 Dec 5	2 3 0	—	4.7	31.0	
1969 Feb 25	1 2 3	16	2.1	41.5	2.1×10^6
1969 Mar 30	1 1 2	9	1.4	13.0	1.3×10^6
1969 Apr 11	3 3 0	—	>16	123.0	1.6×10^7
1969 Nov 2	3 3 0	—	13	737.0	2.1×10^7
1970 Jan 31	1 2 0	—	3	6.2	2.7×10^5
1970 Mar 6	1 3 0	—	5	0.9	1.0×10^5
1970 Mar 29	1 2 0	—	1.8	20.2	1.7×10^6
1970 July 24	1 2 0	—	4.5	0.8	5.8×10^4
1970 Aug 14	2 2 0	—	3.0	2.7	3.9×10^5
1970 Nov 5	1 2 0	—	3.5	1.7	2.8×10^5
1971 Jan 24	3 3 3	26	14.5	408.0	2.7×10^7
1971 Apr 6	1 2 0	—	3.8	5.0	2.0×10^5
1971 Sep 1	2 3 3	16	5.2	162.0	1.3×10^7
1972 May 28	1 2 0	—	2.6	2.7	5.3×10^5
1972 Aug 4	4 4 2	20	>22	21 000.0	6.2×10^8
1972 Aug 7	3 4 2	8	3.1	384.0	3.0×10^7

PECS. The notation PECS stands for the three digit proton event classification system as defined by Smart and Shea [1971] as follows:

First Digit	Second Digit	Third Digit
E > 10 MeV Satellite Measured Proton Intensity Digit $\text{cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$	Daylight Polar 30 MHz Riometer Absorption	Sea Level Neutron Monitor Increase
-3 From 10^{-3} to 10^{-2}	—	—
-2 From 10^{-2} to 10^{-1}	—	—
-1 From 10^{-1} to 10^0	—	—
0 From 10^0 to 10^1	No measurable increase	No measurable increase
1 From 10^1 to 10^2	Less than 1.5 dB	Less than 3%
2 From 10^2 to 10^3	From 1.5 dB to 4.6 dB	From 3% to 10%
3 From 10^3 to 10^4	From 4.6 dB to 15 dB	From 10% to 100%
4 Greater than 10^4	Greater than 15 dB	Greater than 100%

The first digit represents the measurement of the $E > 10 \text{ MeV}$ proton flux by a satellite within the earth-moon system, the second digit represents the 30 MHz absorption measured by a sunlit polar riometer, and the third digit represents the response of a high latitude sea level neutron monitor. An "X" in the first digit indicates that not enough satellite data are available to make a determination of the maximum flux of protons greater than 10 MeV. A digit within parenthesis is provisional, based on partial data.

*Proton flux data derived from Polar Cap absorption measurements.

(sc) Peak flux associated with geomagnetic storm.

| |† Composite integrated proton flux of 2 separate events.

Table 4. Mean and Worst-Case Solar Particle Event Composition for Low Z Ions.^[3]

Element	Mean case	Worst case	Element	Mean case	Worst case
H	1	1	P	2.3×10^{-7}	1.1×10^{-6}
He	1.0×10^{-2}	3.3×10^{-2}	S	8.0×10^{-6}	5.0×10^{-5}
Li	0	0	Cl	1.7×10^{-7}	8.0×10^{-7}
Be	0	0	Ar	3.3×10^{-6}	1.8×10^{-5}
B	0	0	K	1.3×10^{-7}	6.0×10^{-7}
C	1.6×10^{-4}	4.0×10^{-4}	Ca	3.2×10^{-6}	2.0×10^{-5}
N	3.8×10^{-5}	1.1×10^{-4}	Sc	0	0
O	3.2×10^{-4}	1.0×10^{-3}	Ti	1.0×10^{-7}	5.0×10^{-7}
F	0	0	V	0	0
Ne	5.1×10^{-5}	1.9×10^{-4}	Cr	5.7×10^{-7}	4.0×10^{-6}
Na	3.2×10^{-5}	1.3×10^{-5}	Mn	4.2×10^{-7}	2.3×10^{-6}
Mg	6.4×10^{-5}	2.5×10^{-4}	Fe	4.1×10^{-5}	4.0×10^{-4}
Al	3.5×10^{-6}	1.4×10^{-5}	Co	1.0×10^{-7}	5.5×10^{-7}
Si	5.8×10^{-5}	1.9×10^{-4}	Ni	2.2×10^{-6}	2.0×10^{-5}

Table 5. Mean Solar Particle Event Compositions for High Z Ions.^[3]

Element	Mean case	Element	Mean case	Element	Mean case
Cu	2.0×10^{-8}	Sn	2.0×10^{-10}	Lu	2.0×10^{-12}
Zn	6.0×10^{-8}	Sb	1.4×10^{-11}	Hf	8.0×10^{-12}
Ga	2.0×10^{-9}	Te	3.0×10^{-10}	Ta	9.0×10^{-13}
Ge	5.0×10^{-9}	I	6.0×10^{-11}	W	1.0×10^{-11}
As	3.0×10^{-10}	Xe	2.7×10^{-10}	Re	2.0×10^{-12}
Se	3.0×10^{-9}	Cs	2.0×10^{-11}	Os	3.0×10^{-11}
Br	4.0×10^{-10}	Ba	2.0×10^{-10}	Ir	3.0×10^{-11}
Kr	2.0×10^{-9}	La	2.0×10^{-11}	Pt	6.0×10^{-11}
Rb	3.0×10^{-10}	Ce	5.0×10^{-11}	Au	1.0×10^{-11}
Sr	1.0×10^{-9}	Pr	8.0×10^{-12}	Hg	1.0×10^{-11}
Y	2.0×10^{-10}	Nd	4.0×10^{-11}	Tl	9.0×10^{-12}
Zr	5.0×10^{-10}	Pm	0	Pb	1.0×10^{-10}
Nb	4.0×10^{-11}	Sm	1.0×10^{-11}	Bi	6.0×10^{-12}
Mo	2.0×10^{-10}	Eu	4.0×10^{-12}	Po	0
Tc	0	Gd	2.0×10^{-11}	At	0
Ru	9.0×10^{-11}	Tb	3.0×10^{-12}	Rn	0
Rh	2.0×10^{-11}	Dy	2.0×10^{-11}	Fr	0
Pd	6.0×10^{-11}	Ho	4.0×10^{-12}	Ac	0
Ag	2.0×10^{-11}	Er	1.0×10^{-11}	Th	2.0×10^{-12}
Cd	7.0×10^{-11}	Tm	2.0×10^{-12}	Pa	0
In	9.0×10^{-12}	Yb	9.0×10^{-12}	U	1.2×10^{-12}

Table 6. Parameters for the Adams' Model for Hydrogen (f_1), Helium (f_2), and Iron (f_{26}) GCR Spectra. [3]

Parameter	Hydrogen	Helium	Iron
α_0	-2.20	-2.35	-2.14
E_0	1.1775E5	8.27E4	1.175E5
b	2.685	2.070	2.64
X_1 for F_{\min}	0.117	0.241	0.140
X_1 for F_{\max}	0.079	0.180	0.102
X_2	0.80	0.83	0.65
C_1	6.52	4.75	6.63
C_2	4.0	5.10	7.69

Table 7. Hydrogen to Nickel differential energy spectra parameters for the NRL models of Galactic Cosmic Rays^[3]

Element	Ratio:	Energy dependence modeled by:
(Z):		
H (1)	-	$\equiv f_1$
He (2)	-	$\equiv f_2$
Li (3)	0.330	$f_2 G(E)$
Be (4)	0.176	$f_2 G(E)$
B (5)	0.480	$f_2 G(E)$
C (6)	3.04 E-2	f_2
N (7)	-	$f_2 W(E)$
O (8)	2.84 E-2	f_2
F (9)	6.06 E-4	f_2
Ne (10)	4.63 E-3	f_2
Na (11)	1.02 E-3	f_2
Mg (12)	6.02 E-3	$f_2 S(E)$
Al (13)	1.07 E-3	f_2
Si (14)	4.63 E-3	$f_2 S(E)$
P (15)	2.34 E-4	f_2
S (16)	9.30 E-4	$f_2 S(E)$
Cl (17)	0.070	$f_{26} Q(E)$
Ar (18)	0.130	$f_{26} Q(E)$
K (19)	0.090	$f_{26} Q(E)$
Ca (20)	2.1 E-1	f_{26}
Sc (21)	0.042	$f_{26} Q(E)$
Ti (22)	0.147	$f_{26} Q(E)$
V (23)	0.070	$f_{26} Q(E)$
Cr (24)	0.140	$f_{26} Q(E)$
Mn (25)	0.100	$f_{26} Q(E)$
Fe (26)	-	$\equiv f_{26}$
Co (27)	3.40 E-3	f_{26}
Ni (28)	5.00 E-2	f_{26}

Where:

f_1, f_2, f_{26} = Spectra for H, He, and Fe (given by Eq. 29)

$$G(E) = \begin{cases} 0.021 f_2 & \text{for } E < 3000 \text{ MeV}/\mu \\ 0.729 E^{-0.443} f_2 & \text{for } E > 3000 \text{ MeV}/\mu \end{cases}$$

$$W(E) = \{8.70E-3 \exp(-0.4(\log_{10}E - 3.15)^2) + 7.6E-3 \exp(-0.9(\log_{10}E - 0.8)^2)\}$$

$$S(E) = \begin{cases} 1.0 & \text{for } E < 2200 \text{ MeV}/\mu \\ 1.0 + 1.56 \times 10^{-5} (E - 2200) & \text{for } E > 2200 \text{ MeV}/\mu \end{cases}$$

$$Q(E) = 16 \{1 - \exp(-0.075 E^{0.4})\} E^{-0.33}$$

Table 8. Ratios for Copper to Uranium abundances relative to Iron for the NRL Galactic Cosmic Ray models^[3]

Z	Element	Ratio to Iron	Z	Element	Ratio to Iron
29	Cu	6.8 E-4	61	Pm	1.9 E-7
30	Zn	8.8 E-4	62	Sm	8.7 E-7
31	Ga	6.5 E-5	63	Eu	1.5 E-7
32	Ge	1.4 E-4	64	Gd	7.0 E-7
33	As	9.9 E-6	65	Tb	1.7 E-7
34	Se	5.8 E-5	66	Dy	7.0 E-7
35	Br	8.3 E-6	67	Ho	2.6 E-7
36	Kr	2.3 E-5	68	Er	4.3 E-7
37	Rb	1.1 E-5	69	Tm	8.9 E-8
38	Sr	3.6 E-5	70	Yb	4.4 E-7
39	Y	6.8 E-6	71	Lu	6.4 E-8
40	Zr	0.7 E-5	72	Hf	4.0 E-7
41	Nb	2.6 E-6	73	Ta	3.6 E-8
42	Mo	7.1 E-6	74	W	3.8 E-7
43	Tc	1.6 E-6	75	Re	1.3 E-7
44	Ru	5.3 E-6	76	Os	5.6 E-7
45	Rh	1.5 E-6	77	Ir	3.7 E-7
46	Pd	4.5 E-6	78	Pt	7.2 E-7
47	Ag	1.3 E-6	79	Au	1.3 E-7
48	Cd	3.6 E-6	80	Hg	2.3 E-7
49	In	1.4 E-6	81	Tl	1.8 E-7
50	Sn	7.5 E-6	82	Pb	1.7 E-6
51	Sb	9.9 E-7	83	Bi	9.0 E-8
52	Te	5.7 E-6	84	Po	0
53	I	1.5 E-6	85	At	0
54	Xe	3.5 E-6	86	Rn	0
55	Cs	5.8 E-7	87	Fr	0
56	Ba	6.0 E-6	88	Ra	0
57	La	5.3 E-7	89	Ac	0
58	Ce	1.6 E-6	90	Th	9.0 E-8
59	Pr	3.0 E-7	91	Pa	0
60	Nd	1.1 E-6	92	U	5.4 E-8

PHYSICAL DATA

OBJECT	EQUATORIAL RADIUS [km]	MASS [kg]	DIPOLE MAG MOMENT [G-cm ³]	$\vec{E} = (10^{-8}) (\vec{V}_S \times \vec{B})$ ↓ INDUCED E-FIELD AT SURFACE [V/cm]
SUN	6.960×10^5	1.991×10^{30}	$\sim 3.4 \times 10^{32}$	$\sim 4.4 \times 10^{-1}$
MERCURY	2.439×10^3	3.303×10^{23}	5×10^{22}	1.0×10^{-5}
VENUS	6.050×10^3	4.870×10^{24}	~ 0	—
EARTH	6.378×10^3	5.976×10^{24}	8.1×10^{25}	2.5×10^{-3}
MOON	1.738×10^3	7.353×10^{22}	0	—
MARS	3.398×10^3	6.421×10^{23}	$< 10^{22}$	$< 10^{-6}$
JUPITER	7.14×10^4	1.899×10^{27}	1.59×10^{30}	1.84×10^{-1}
IO	1.82×10^3	8.916×10^{22}	$\sim 6.5 \times 10^{17}$	$\sim 2 \times 10^{-10}$
SATURN	6×10^4	5.686×10^{26}	4.3×10^{28}	5.0×10^{-3}
TITAN	2.56×10^3	1.359×10^{23}	—	—
URANUS	2.6145×10^4	8.66×10^{25}	$\sim 1.9 \times 10^{28} (?)$	$\sim 1.6 \times 10^{-2} (?)$
COMETS	$\sim 1-10$	$\sim 10^{12}-10^{17}$		

TABLE 9. Physical data on the size and magnetic fields of the planets

TABLE 10. Parameters of D4 Dipole, Based in System III (1965)
Coordinates

Dipole Offset Vector	Magnetic Moment Vector
Latitude $\lambda = 5.12^\circ$	tilt colatitude 10.77°
West longitude $l = 155.6^\circ$	west longitude $l = 200.8^\circ$
Magnitude $r = 0.1010 R_J$	magnitude $M = 4.225 \text{ G } R_J^3$
Components	Components
$x = -0.0916 R_J$	$M_x = -0.738 \text{ G } R_J^3$
$y = -0.0416 R_J$	$M_y = +0.280 \text{ G } R_J^3$
$z = +0.0090 R_J$	$M_z = +4.151 \text{ G } R_J^3$

TABLE 11. Parameter values for Saturn used in the model calculations.

QUANTITY	VALUE	SOURCE
Radius	$R_S = 6.033 \times 10^7 \text{ m}$	CO84, p. 354
Gravitation	$GM_S = 3.794 \times 10^{16} \text{ m}^3/\text{s}^2$	CE89
Rotation	$\omega = 810.76 \text{ deg/day}$	CO84, p. 360
Magnetic field coefficients	$\left. \begin{array}{l} g_1^0 = 0.21535 \text{ G} \\ g_2^0 = 0.01642 \text{ G} \\ g_3^0 = 0.02743 \text{ G} \\ G_1^0 = -0.0001 \text{ G} \\ G_1^1 = -0.00001 \text{ G} \end{array} \right\}$	CO84, p. 363, Table II

TABLE 12. Isotopic Composition of the Production Grade $^{238}_{94}Pu$ used in the Galileo GPHS-RTGs.[69]

Plutonium Isotope	Weight Percent	Half-Life Years	Radioactivity Curies/gram Pu _x O ₂
236	0.00012	2.8	476
238	81.0	87.8	15.4
239	15.2	24,000	0.054
240	2.9	6,600	0.195
241	0.8	13.0	100.4
242	0.1	370,000	0.0034

TABLE 13. Normalized Differential γ and Neutron Fluxes for the Galileo RTGs and RHUs.[71] The data are based on 1.2 ppm ^{236}Pu and a fuel age of 5 yrs. The flux will change with fuel age and should be corrected using Figures 64 and 66.

Gammas			Neutrons		
Energy Width (MeV)	Normalized Number		Energy Width (MeV)	Normalized Number Flux	
	Width	Flux in Group (/cm ² -s)		in Group (n/cm ² -s)	
7.0	- 6.0	1.09 (-5)	10.0	- 8.55	9.64 (-4)
6.0	- 5.0	3.29 (-5)	8.55	- 6.66	4.59 (-3)
5.0	- 4.0	9.72 (-5)	6.66	- 5.18	9.93 (-3)
4.0	- 3.0	2.96 (-4)	5.12	- 4.46	9.66 (-3)
3.0	- 2.616	2.65 (-4)	4.46	- 4.04	1.34 (-2)
2.616	- 2.614	1.56 (-1)	4.04	- 3.14	8.73 (-2)
2.614	- 2.0	1.54 (-2)	3.14	- 2.45	1.76 (-1)
2.0	- 1.75	2.79 (-3)	2.45	- 1.91	1.57 (-1)
1.75	- 1.5	3.89 (-2)	1.91	- 1.49	1.46 (-1)
1.5	- 1.25	1.75 (-2)	1.49	- 1.16	9.76 (-2)
1.25	- 1.0	2.21 (-2)	1.16	- 0.9	7.60 (-2)
1.0	- 0.75	1.18 (-1)	0.9	- 0.702	5.87 (-2)
0.75	- 0.585	9.96 (-2)	0.702	- 0.546	4.35 (-2)
0.585	- 0.584	2.59 (-2)	0.546	- 0.331	5.45 (-2)
0.584	- 0.50	6.05 (-2)	0.331	- 0.201	2.98 (-2)
0.50	- 0.40	9.17 (-2)	0.201	- 0.122	1.66 (-2)
0.40	- 0.35	4.60 (-2)	0.122	- 4.49 (-2)	1.48 (-2)
0.35	- 0.30	4.56 (-2)	4.49 (-2)	- 1.7 (-2)	3.07 (-3)
0.30	- 0.275	1.92 (-2)	1.7 (-2)	- 5.55 (-4)	7.77 (-4)
0.275	- 0.239	2.41 (-2)	5.55 (-4)	- 3.0 (-5)	1.17 (-6)
0.239	- 0.238	3.91 (-3)	3.0 (-5)	- 3.0 (-6)	1.62 (-9)
0.238	- 0.200	3.49 (-2)	3.0 (-6)	- 1.1 (-7)	4.65 (-11)
0.200	- 0.175	2.51 (-2)	1.1 (-7)	- 2.5 (-8)	1.74 (-12)
0.175	- 0.153	1.70 (-2)			
0.153	- 0.152	2.86 (-2)	Integral Total		1.000
0.152	- 0.125	9.19 (-2)			
0.125	- 0.1213	9.82 (-4)			
0.1213	- 0.1211	9.41 (-6)			
0.1211	- 0.1000	1.58 (-2)			
0.1000	- 0.0099	1.18 (-3)			
0.0099	- 0.0080	2.24 (-2)			
Integral Total			1.000		

TABLE 14. Galileo RHU neutron and γ radiation distance variation factors to be used in conjunction with Table 13 to give absolute values of the flux, fluence, and dose as a function of distance from the RHU.^[71] The γ -ray flux and dose are based on a ~5 old fuel at launch and must be corrected by the curve presented in Figure 66.

Distance From RHU (cm)	Neutron		Gamma	
	Peak Flux (cm ⁻² s ⁻¹)	Fluence (cm ⁻²)	Peak Flux*	Dose** rad(Si)
0	6.1×10^2	1.6×10^{11}	4.7×10^4	5.3×10^3
2	9.5×10^1	2.4×10^{10}	7.3×10^3	8.3×10^2
4	4.0×10^1	1.0×10^{10}	2.5×10^3	2.9×10^2
6	2.4×10^1	6.0×10^9	1.5×10^3	1.7×10^2
8	1.2×10^1	3.0×10^9	1.0×10^3	1.1×10^2
10	8.1×10^0	2.0×10^9	6.4×10^2	7.2×10^1
15	3.9×10^0	9.8×10^8	3.0×10^2	3.4×10^1
20	2.3×10^0	5.8×10^8	1.8×10^2	2.0×10^1
50	3.9×10^{-1}	9.8×10^7	3.0×10^1	3.4×10^0

TABLE 15. Summary table for dosage behind 58 mil shielding for the trapped electron (AE8) and trapped proton (AP8) environments. Results correspond to differing levels of geomagnetic activity, geometry, and orbit. Units are rads(Si).

CLEMENTINE TRAPPED RADIATION ENVIRONMENT ESTIMATES

Active Electrons		58 Mils	DOSAGE (Rads)					
Orbit Phase	Date	SPHERE	SPH SHELL	2*SLAB	DUBL	SLAB	BACK SLAB	
1. Low Earth Orbit	94/01/24	.001150	.000437	.000154	.000315	.000315		
2. Earth-Moon Trans	94/01/26	184.000	83.500	32.600	61.300	61.400		
3. Earth-Moon Orbit	94/02/04	778.000	359.000	127.000	260.000	260.000		
4. Earth-Moon Orbit	94/02/15	466.000	201.000	78.400	143.000	143.000		
5. Earth Grav Assist	94/05/06	247.000	88.400	30.100	60.400	60.400		
6. Earth Grav Assist	94/05/24	665.000	267.000	107.000	183.000	183.000		
	TOTAL DOSE	2340.001	998.900	375.100	707.700	707.800		

Active Protons		58 Mils	DOSAGE (Rads)					
Orbit Phase	Date	SPHERE	SPH SHELL	2*SLAB	DUBL	SLAB	BACK SLAB	
1. Low Earth Orbit	94/01/24	.001800	.001800	.001310	.001310	.001310		
2. Earth-Moon Trans	94/01/26	29.000	29.000	9.780	9.780	9.780		
3. Earth-Moon Orbit	94/02/04	104.000	104.000	31.900	31.900	31.900		
4. Earth-Moon Orbit	94/02/15	66.800	66.800	19.400	19.400	19.400		
5. Earth Grav Assist	94/05/06	0.000	0.000	0.000	0.000	0.000		
6. Earth Grav Assist	94/05/24	0.000	0.000	0.000	0.000	0.000		
	TOTAL DOSE	199.802	199.802	61.081	61.081	61.081		

Quiet Electrons		58 Mils	DOSAGE (Rads)					
Orbit Phase	Date	SPHERE	SPH SHELL	2*SLAB	DUBL	SLAB	BACK SLAB	
1. Low Earth Orbit	94/01/24	.001120	.000434	.000153	.000312	.000312		
2. Earth-Moon Trans	94/01/26	128.000	56.300	21.700	41.400	41.400		
3. Earth-Moon Orbit	94/02/04	557.000	250.000	87.500	181.000	181.000		
4. Earth-Moon Orbit	94/02/15	466.000	201.000	78.400	143.000	143.000		
5. Earth Grav Assist	94/05/06	247.000	88.400	30.100	60.400	60.400		
6. Earth Grav Assist	94/05/24	665.000	268.000	107.000	183.000	184.000		
	TOTAL DOSE	2063.001	863.700	324.700	608.800	609.800		

Quiet Protons		58 Mils	DOSAGE (Rads)					
Orbit Phase	Date	SPHERE	SPH SHELL	2*SLAB	DUBL	SLAB	BACK SLAB	
1. Low Earth Orbit	94/01/24	.001800	.001800	.001310	.001310	.001310		
2. Earth-Moon Trans	94/01/26	29.100	29.100	9.830	9.830	9.830		
3. Earth-Moon Orbit	94/02/04	104.000	104.000	32.000	32.000	32.000		
4. Earth-Moon Orbit	94/02/15	66.700	66.700	19.400	19.400	19.400		
5. Earth Grav Assist	94/05/06	0.000	0.000	0.000	0.000	0.000		
6. Earth Grav Assist	94/05/24	0.000	0.000	0.000	0.000	0.000		
	TOTAL DOSE	199.802	199.802	61.231	61.231	61.231		

		DOSAGE (Rads)						58 Mils
		SPHERE	SPH SHELL	2*SLAB	DUBL	SLAB	BACK SLAB	
TOTAL ACTIVE DOSE		2539.80	1198.70	436.18	768.78	768.88		
TOTAL QUIET DOSE		2262.80	1063.50	385.93	670.03	671.03		

TABLE 16. "Worst Case" solar proton dosage predictions for ~1 year Clementine mission scenario. A 58 mil aluminum shell is assumed.

Probability Dose Does Not Exceed:	Maximum Dose Rad (Si)
50%	466
90%	2750
95%	4640
99%	12700

2.8. FIGURES

Fig. 1. The flux (number of particles or photons per unit time) of a given energy E per unit energy interval dE in a unit solid angle ($d\Omega=2\pi \cos \theta d\theta d\phi$) about the direction of observation (in the θ, ϕ direction), incident on unit of surface area (dA) perpendicular to the direction of observation.[2]

Fig. 2. The cosmic ray iron spectrum: The solid curves are for solar maximum (lower) and solar minimum (upper). The dashed curve is the 90% worst-case iron spectrum, which is implied by comparison with the cosmic ray helium spectrum.[3]

Fig. 3. Illustration of a neutron displacement damage equivalence curve. The curves (actual and fitted), in units of MeV-mb (where b is a barn or 10^{-24} cm^2), are for neutrons. The "relative displacement damage" is defined in terms of the cross section times the energy of the incident particle. As an example, for 14 MeV neutrons, the 1-MeV neutron dose damage equivalent is given by multiplying the 14 MeV dose by 2.5.[4]

Fig. 4. Linear Energy Transfer Function (LET) versus Energy. L_0 corresponds to a constant value of LET. As illustrated, there can be multiple values of energy (E_1 and E_2 here) that correspond to the same value of LET.

Fig. 5. Integral LET spectrum for iron.[5]

Fig. 6. The Earth's radiation belts in idealized dipole space, according to the AP8 and AE8 models. Average omnidirectional integral fluxes above energy thresholds are shown.[8]

Fig. 7. Decay of the Equatorial Dipole Field Strength with Time.[9]

Fig. 8. Geomagnetic Field Magnetic Elements[9].

Fig. 9. Spherical Coordinate System for Geographic and Geomagnetic Coordinates.[9]

Fig. 10a. Cross section of the Earth's magnetic field in the noon-midnight meridian showing the structure of the field lines and the plasma regions they contain.

Fig. 10b. Illustration of the magnetic field amplitude over the northern hemisphere at a constant altitude of 400 km. The field amplitude varies from a minimum of 0.25 G near the equator to 0.5 G over the polar caps.

Fig. 11. Motion of a charged particle (positive in this case) in a magnetic field in the absence of an electric field. The magnetic field is constant in the lower half of the figure and equal to B_1 . It is constant and equal to B_2 in the upper part of the figure. The figure illustrates the effects of a gradient in a direction perpendicular to the direction of the magnetic field.

Fig. 12. Motion of a charged particle in a converging magnetic field in the absence of an electric field. F_1 is the force along the magnetic field that results from the field convergence (or divergence) and is responsible for the mirroring effect (see text).

Fig. 13. Motion of a Charged Particle in a Dipole Magnetic Field.

Fig. 14. Proton Flux Densities at 296 km Altitude.[19]

Fig. 15. Electron Constant Flux Contours at 400 km Altitude.(E>0.5 MeV)[19]

Fig. 16. Hourly averages of ATS1 1.9 MeV omnidirectional electron flux at local midnight.[20] Daily sum of K_p is plotted at the bottom of the figure. The horizontal line is the AEI-7 model equatorial flux at $L = 6.6$ and local midnight.

Fig. 17. 10 day averages of the inner zone electron fluxes greater than 0.28 MeV (in units of count rate) measured by the satellite 1963-38C. The effects of the decay of the Starfish nuclear debris and of the 4 major geomagnetic perturbations are evident.[20]

Fig. 18. ATS-1, ATS-5, and ATS-6 energetic electron fluxes (running 27-day averages) as a function of time[21]. ATS-5 data were normalized to ATS-6 data in mid-1974. The energetic thresholds for the ATS-1, ATS-5, and ATS-6 channels are shown on the figure. The flux averages for each year are also indicated (solid horizontal lines). Superimposed on this graph are the Zurich monthly sunspot numbers.

Fig. 19. Pitch angle distributions observed at selected times during the geomagnetic substorm of September 8, 1977. Electron counts per 8 ms sample plotted versus μ , the cosine of the pitch angle (α), for various energies.[89]

Fig. 20. Explorer 45 observations of the electron pitch angle distribution at $L = 3$, approximately once every two days, for the period following the storm of June 17, 1972. Prestorm distributions for June 16, 1972 are also shown.[22]

Fig. 21. Equatorial proton pitch angle distributions from Explorer 45 for every 0.5 L on inbound orbit 98. Elevated fluxes at pitch angles from 90° to 180° for the energies of <30 keV are due to reflected sunlight. The data are for an interval following a storm on December 16, 1971.[23]

Fig. 22. $B/B_0 \cdot L$ coverage for experiments used in constructing the AP-8 model. The model and the data used have a cutoff at $L = 6.6$.[15]

Fig. 23. Normalized integral proton flux as a function proton energy. The range of spectral shapes is what is to be expected given the current knowledge of natural variations in this population.[30]

Fig. 24. Depth-dose per omnidirectional proton as a function of spherical aluminum thickness based on the spectra in Figure 23.[30]

Fig. 25. Normalized integral electron flux as a function electron energy. The range of spectral shapes is what is to be expected given the current knowledge of natural variations in this population.[30]

Fig. 26. Depth-dose per omnidirectional electron as a function of spherical aluminum thickness based on the spectra in Figure 25.[30]

Fig. 27. Schematic drawing of a cosmic ray shower. N and P are high energy neutrons and protons while the lower case n and p are used to denote disintegration product neutrons and protons. Pions, mesons, electrons, and gamma rays are indicated by conventional symbols.[6]

Fig. 28. Neutron spectra at various depths from top of the atmosphere. The data are adjusted to solar minimum conditions at the midlatitude geomagnetic latitude of $\sim 42^\circ$ (equivalent to a vertical cutoff rigidity of 4.5 GV).[6]

Fig. 29. A model "global averaged" albedo neutron spectrum.[90]

Fig. 30. Schematic plot of the relative variations in time of the amplitudes of the X-ray, radio noise, high energy particle, and solar plasma fluxes for a "typical" large solar flare.

Fig. 31. Solar-flare proton fluences (vertical lines) superimposed on the sunspot cycle (continuous curve). Solid lines correspond to $E > 3$ MeV, dashed lines to $E > 1$ MeV proton fluences.

Fig. 32. Yearly Event Fluences for Protons of Energy > 30 MeV vs Year Relative to Sunspot Maximum.[31]

Fig. 33. Comparison of the $E > 10$ MeV proton flux as a function of time for the great flares of August 1972 and the October 1989.

Fig. 34. The peak proton flux differential energy spectra for: F_m , a "typical" event; F_w , a "worst-case" event (90% confidence level); and F_a , an "anomalously large" event.[28]

Fig. 35. Distribution of Solar Event Fluences for Solar Active Years between 1963 and 1991 for Protons of Energy > 10 MeV for which Daily Averaged Flux exceeds $1.0 \text{ cm}^{-2} \cdot \text{s}^{-1} \cdot \text{sr}^{-1}$. The straight line is the selected log normal distribution.[31]

Fig. 36. Fluence Probability Curves for Protons of Energy Greater than 10 MeV for Various Mission Lengths.[31]

Fig. 37. The energy spectra of cosmic-ray protons (line) and electrons (points) as measured near Earth.[39] Below a few GeV, interstellar spectra are strongly influenced by the Sun.

Fig. 38. Relative abundance of the elements from hydrogen to the iron group, normalized to that of carbon ($C=100$).[39]. The barred points represent the cosmic-ray abundances measured near Earth. The dots (without error bars) represent the elemental abundances in the solar system.

Fig. 39. Schematic representations^[40] of the effects of the Earth's magnetic field on the Cosmic Ray environment as functions of particle rigidity and orbital inclination. The top 3 spectra represent the Cosmic Ray spectrum outside the Earth's field. The second set of figures represents a an orbit average geomagnetic transmittance factor which is a function of rigidity and orbital inclination. The third plot is the result of convolving the 2 upper sets of curves and is the orbit average Cosmic Ray exposure as a function of rigidity for 0° , 30° , and 50° inclination orbits.

Fig. 40. Cosmic-ray energy spectra^[39] of the more abundant nuclear species as measured near Earth. Below a few GeV/nucleon, these spectra are strongly influenced by modulation within the solar system. The different curves for the same species represent measurements at various levels of general solar activity, the lowest intensity being observed at the highest activity level.

Fig. 41. Comparison of planetary-magnetospheres.^[44, 91]

Fig. 42. Schematic representation of Jupiter's magnetosphere illustrating the various plasma regions and particle flows.^[54]

Fig. 43. Equatorial and cutoff magnetic field strengths for the Jupiter magnetic field and charged particle models.^[45]

Fig. 44. Sample dependencies^[45] for Jupiter energetic electron model. (a) Omnidirectional integral flux at three energies as a function of L shell. (b) Integral intensity at 90° pitch angle at the equator and for three L shell values as a function of energy. (c) Integral ($E > 3$ MeV) intensity at 90° pitch angle for different L shells as a function magnetic latitude.

Fig. 45. Sample dependencies^[45] for Jupiter energetic proton model. (a) Omnidirectional integral flux at three energies as a function of L shell. (b) Integral intensity at 90° pitch angle at the equator and for three L shell values as a function of energy. (c) Integral ($E > 20$ MeV) intensity at 90° pitch angle for different L shells as a function magnetic latitude. The increase of the intensity at $L = 7.0$ and $l_m > 55^\circ$ illustrates the kind of model defect that can occur when carefully crafted algebraic forms are used to extrapolate beyond the range of available data, and this increase is unphysical.

Fig. 46. Contours for electron fluxes above 1 MeV at Jupiter.^[45]

Fig. 47. Contours for proton fluxes above 1 MeV at Jupiter.^[45]

Fig. 48. Comparison^[45] of Voyager 1 LECP electron flux profiles with predictions from the Jupiter energetic electron model. Arrows indicate left or right flux scales.

Fig. 49. Comparison of UHF strip brightness observed from Earth at 21-cm wavelength with predictions from the Jupiter energetic electron model.^[45]

Fig. 50. A schematic representation of Saturn's magnetosphere in the r-z (trajectory) plane as revealed by the LECP data^[59]. Tick marks on the trajectory are at 6-hour intervals from day 318 to 320. Note the Titan-associated mantle region outside $\sim 17 R_s$, and the presence of closed field lines in the tail lobe region. The phase of the second-order anisotropies relative to the magnetic field (B) is shown for both electron (e) and protons (p). R, Rhea; T, Titan; MP, magnetopause.

Fig. 51. Integral omnidirectional flux for the Saturn electron model^[60] at three energy thresholds. The fluxes at the magnetic equator ($B_e/B=1.0$) and at high magnetic latitude ($B_e/B=0$) are represented respectively by the solid and dashed lines. The horizontal axis changes scale at $L=5$.

Fig. 52. Integral omnidirectional flux for the Saturn electron model^[60] at three sample values of L (numbers adjacent to the curves). On the left the dependencies on local magnetic field strength B are shown for energy thresholds of 0.1, 1.0, and 10 MeV (solid, dashed, and solid lines) at each L. On the right, spectra are shown for the magnetic equator ($B_e/B=1.0$, solid lines) and for high magnetic latitude ($B_e/B=0$, dashed lines) at each L.

Fig. 53. Integral omnidirectional flux for the Saturn proton model^[60] at three energy thresholds. The fluxes at the magnetic equator ($B_e/B=1.0$) and at high magnetic latitude ($B_e/B=0$) are represented respectively by the solid and dashed lines. The horizontal axis changes scale at L=5.

Fig. 54. Integral omnidirectional flux for the Saturn proton model^[60] at three sample values of L (numbers adjacent to the curves). On the left, the dependencies on local magnetic field strength B are shown for energy thresholds of 1, 10, and 100 MeV (solid, dashed and solid lines) at each L. On the right spectra are shown for the magnetic equator ($B_e/B=1.0$, solid lines) and for high magnetic latitude ($B_e/B=0$, dashed lines) at each L.

Fig. 55. The spectral distribution of solar irradiance^[61] from a wavelength of 10^{-3} Å to 10 m.

Fig. 56. X-ray flux (watts/m²) between 1 and 8 Å versus time for the November 26, 1982 solar flare as measured by the NOAA GOES spacecraft at geosynchronous orbit.^[63] Also shown are the various classes of X-ray flares.

Fig. 57. The prompt radiation environment in space for a 1 megaton nuclear weapon as a function of distance from the detonation.^[64] The neutron environment is in terms of particle fluence (n/cm²). The γ -ray environment is in dosage (rads(Si)). The X-ray environment is in terms of energy fluence (cal/cm²).

Fig. 58. Schematic representations of the prompt radiation differential spectra for nuclear weapon generated neutrons, γ -rays, and X-rays.^[65, 92]

Fig. 59. The natural and manmade electron radiation dose environments for a.) polar and b.) equatorial orbits with altitudes between 10^2 and 10^5 nautical miles. The plots are in units of a.) e⁻/cm² 10 days for E>5 MeV and b.) units of e⁻/cm² 10 days for E>1 MeV or rads(Si) for 0 shielding. Curves are presented for nuclear saturated environments (see text), a single continental US detonation, a single low latitude nuclear detonation, and the natural environment (for solar maximum and solar minimum). The shading for the equatorial plot shows the range of expected variations. (Adapted from [68] and [65].)

Fig. 60. Comparison of the natural and fission-produced equatorial electron flux spectra as a function of energy. The variation in the natural environment spectrum is with altitude (in terms of L-shell). The spectra are normalized to 1 at E=0.5 MeV.^[68]

Fig. 61. Illustration of the construction of a Galileo General Purpose Heat Source (GPHS) Radioisotope Thermoelectric Generator (RTG).^[69, 70]

Fig. 62. Neutron isodose isoflux plots for a single 4500 Watts(th) GPHS RTG (point kernel calculation).^[71]

Fig. 63. Gamma isodose-isoflux contours for a single 4500 Watt(th) GPHS RTG (5-year-old fuel).^[71] Values should be increased by a factor of 2.0 for the worst case; to determine refined estimates, use factors taken from Table .

Fig. 64. RTG γ -ray flux correction factor for fuel aging.^[71]

Fig. 65. Schematic of a Galileo Lightweight Radioisotope Heater Unit (LHRHU).^[69]

Fig. 66. LHRHU γ -ray flux correction factor for fuel aging.^[71]

Fig. 67. Illustration of the relative importance of the three photon interactions as a function of Z and photon energy. The solid lines correspond to equal interaction cross sections for the neighboring effects. The dashed line illustrates the situation for photon interactions with silicon.^[73]

Fig. 68. The stopping power (or LET) in MeV-cm²/g versus energy per atomic mass unit for a variety of ions (MeV/ μ) and electrons (MeV) in silicon.^[3]

Fig. 69. Minimum penetration energy for electrons and protons versus shield thickness.

Fig. 70a. Computer plot showing five electron trajectories in a Cu target with a 20 keV incident beam normal to the surface.^[77, 78]

Fig. 70b. Monte Carlo results for 100 electron trajectories in a Cu target with 20 keV incident electron energy.^[77, 78]

Fig. 71. Effect of varying target thickness; electron trajectories in an Al target of thickness 0.25 μ m, 1.3 μ m, and infinity, for an incident electron energy of 20 keV.^[77, 78]

Fig. 72. Electron dose as a function of depth for CaFMn TLDs.^[80] The dosage is normalized to 1 at the front face.

Fig. 73. Proton energy deposition (MeV/mm-proton) as a function of depth for CaF₂ (density of 3.18 g/cm³).^[81]

Fig. 74. 5 shielding configurations considered in the NOVICE code for calculating dosage. Note that examples 3 and 4 are doubled by the code.

Fig. 75. Ion range versus energy in Al^[32] for H, He, C, O, Ar, and Fe. The range is in units of g-cm² and the energy in MeV/ μ .

Fig. 76. Shielding attenuated cosmic ray differential iron spectra for the 3 cases in Figure 2: 90% worst case (upper), solar minimum (middle) and solar maximum

(lower).^[32] These spectra are for 1 AU (no magnetospheric shielding) and behind 0.025 inches of aluminum shielding.^[32]

Fig. 77. The same three cases as in Figures 2 and 76, but now the spectra have been transformed into differential LET spectra.^[32]

Fig. 78. The same three cases as in Figures 2 and 76, but now the spectra of Figure 77 have been integrated to give the integral LET spectra (Heinrich curves) for Fe.^[32]

Fig. 79. Radiation dosage from the trapped proton environment (AP8) for solar-quiet conditions and for the first Earth-Moon Transfer Orbit {1994/01/30(07:00)--1994/02/09(12:00)}.

Fig. 80. Radiation dosage from the trapped electron environment (AE8) for solar-active conditions and for the first Earth-Moon Transfer Orbit {1994/01/30(07:00)--1994/02/09(12:00)}.

Fig. 81. Mission summary plot for the total radiation dose expected for the Clementine Interstage. This assumes a 450 day mission and a 95% confidence flare environment. Dosage is plotted as a function of aluminum shield thickness for a spherical shell geometric configuration.

Fig. 82. Solar flare proton dosage (rad(Si)) for a 50% probable maximum dose after 1,2,3,5, and 7 years into the active portion of the solar cycle beginning on May 26, 1999. Results are for 1 AU (no magnetospheric shielding).

Fig. 83. Solar flare proton dosage (rad(Si)) for a 99% probable maximum dose after 1,2,3,5, and 7 years into the active portion of the solar cycle beginning on May 26, 1999. Results are for 1 AU (no magnetospheric shielding).

Fig. 84. Comparison between the 1 year dose from the saturated nuclear radiation environment for a range of Earth orbits and the total dose expected for the Galileo Jupiter mission after 5 orbits and after a 2 year extended mission.

Fig. 85. Correlation of Voyager Power On Reset (POR's) Anomalies with the High Energy ($E > 1$ MeV and $E > 10$ MeV) Electron Environment.^[86] Also shown for comparison is the high energy proton environment (15 MeV $< E <$ 26 MeV).

Fig. 86 Integral LET spectra as functions of spacecraft wall thickness for a) the Adam's worst case composite solar flare^[87, 93], and b) the GCR environment (elements 1 to 26)^[87, 93].

Fig. 87. Integral LET spectra inside a spacecraft (with 0.064 cm aluminum walls) in a 400 km circular orbit.^[87] The 90% worst case environment is assumed in the interplanetary medium and the GSFC AP trapped proton environment at the Earth. The LET spectra are for the various orbital inclinations indicated.

Fig. 88. Integral LET spectra inside a spacecraft with 0.064 cm aluminum walls that is in a circular orbit at a 60° inclination.^[87] As in Fig. 87, the 90% worst case environment is assumed in the interplanetary medium and the GSFC AP trapped proton environment at the Earth. The LET spectra are for the various altitudes as shown.

2.9. REFERENCES

1. Condon, E.U., and H. Odishaw, *Handbook of Physics*. 2 ed. 1967, New York: McGraw-Hill.
2. Roederer, J.G., *Dynamics of Geomagnetically Trapped Radiation*. Physics and Chemistry in Space, Vol. 2. 1970, New York: Springer-Verlag. 166.
3. Adams, J.H., Jr. *Cosmic Ray Effects on Microelectronics, Part IV*. 1986, Naval Research Laboratory, Washington D.C.: NRL Memorandum Report 5901.
4. van Lint, V.A.J. *Radiation Effects Testing Short Course*. in *IEEE 1986 Nuclear and Space Radiation Effects Conference*. 1986. Providence, Rhode Island: IEEE.
5. Petersen, E.L. *Single Event Upsets in Space: Basic Concepts*. in *Tutorial Short Course, IEEE Nuclear and Space Radiation Effects Conference*. 1983. Gatlinburg, Tenn.: IEEE.
6. Smart, D.F., and M.A. Shea, *Galactic Cosmic Radiation and Solar Energetic Particles*. Handbook of Geophysics and the Space Environment, ed. A.S. Jursa. Accession No. ADA 167000. 1985, Springfield, VA: National Technical Information Service. Chapter 6.
7. Van Allen, J.A., *Geomagnetically Trapped Radiation*. Space Science, 1963, New York-London: J. Wiley & Sons, Inc.
8. Daly, E.J., *The Evaluation of Space Radiation Environments for ESA Projects*. ESA Journal, 1988. 12 : pp. 229-247.
9. Knecht, D.J., and B. Shuman, *The Geomagnetic Field*, in *Handbook of Geophysics and the Space Environment*, A.S. Jursa, Editor. 1985, National Technical Information Service: Springfield, VA. Chapter 4.
10. Olson, W.P., and K.A. Pfitzer, *A quantitative model of the magnetospheric magnetic field*. J. Geophys. Res., 1974. 79: p. 3739.
11. Voigt, G.H., *A three dimensional, analytical magnetospheric model with defined magnetopause*. Z. Geophys., 1972. 38 : p. 319-346.
12. Tsyganenko, *Magnetosphere Model*... Annals. Geophys., 1976. 32: p. 1.
13. Barraclough, D.R., *International Geomagnetic Reference Field: The 4th Generation*. 1987. 48: p. 279.
14. Gussenhoven, M.S., et al., *Preliminary comparison of dose measurements on CRRES to NASA model predictions*. IEEE Trans. Nuc. Sci., 1991. 38(6, December): pp. 1655-1662.
15. Vette, J.I., M.J. Teague, D.M. Sawyer, and K.W. Chan. *Modeling the Earth's Radiation Belts*. in *Solar-Terrestrial Prediction Proceedings*. 1979. Boulder: NOAA.

16. Sawyer, D.M., and J.I. Vette. *AP-8 Trapped Proton Environment for Solar Maximum and Solar Minimum*. 1976, NSSDC/WDC-A-R&S: 76-06.
17. Vette, J.I. *The AE-8 Trapped Electron Model Environment*. November, 1991, NSSDC/WDC-A-R&S: 91-24.
18. Vampola, A.L. *The Space Particle Environment in Materials in Space*. 1988. Langley Res. Cntr.: NASA.
19. Smith, R.E., and G.S. West, ed. *Space and Planetary Environment Criteria Guidelines for Use in Space Vehicle Development, 1982 Revision*. Vol. 1, TM-82478. 1983, NASA.
20. Chan, K., M.J. Teague, N. Schofield, and J.I. Vette, *Modeling of electron time variations in the radiation belts*. in *Quantitative Modeling of Magnetospheric Processes*, W.P. Olson, Editor. 1979, American Geophysical Union: Washington, D.C. p. 121.
21. Paulikas, G.A., and J.B. Blake, *Effects of the Solar Wind on Magnetospheric Dynamics: Energetic Electrons at the Synchronous Orbit*, in *Quantitative Modeling of Magnetospheric Processes*, W.P. Olson, Editor. 1979, American Geophysical Union: Washington, D.C. p. 180.
22. Lyons, L.R., and D.J. Williams, *The storm and poststorm evolution of energetic (35-560 keV radiation belt electron distributions)*. J. Geophys. Res., 1975. **80**: pp. 3985-3994.
23. Taylor, W.W.L., and L.R. Lyons, *Simultaneous equatorial observations of 1 to 30-Hz waves and pitch angle distributions of ring current ions*. J. Geophys. Res., 1976. **81**: pp. 6177-6183.
24. Blake, J.B., and L.N. Friesen. *15th International Cosmic Ray Conference*. 1977.
25. Cornwall, J.M., *Radial Diffusion of Ionized Helium and Proton: A Probe for Magnetospheric Dynamics*. J. Geophy. Res., 1972. **77**: p. 1756.
26. Spjeldvik, W.N., and T.A. Fritz, *Theory of Charge States of Energetic Oxygen Ions in the Earth's Radiation Belts*. J. Geophys. Res., 1978. **83**(A4): p. 1583.
27. Fritz, T.A., and W.N. Spjeldvik, *Simultaneous Quiet Time Observations of Energetic Radiation Belt Protons and Helium Ions: the Equatorial α/p Ratio Near 1 MeV*. J. Geophys. Res., 1979. **84**: pp. 2608-2618.
28. Adams, J.H., Jr. , R. Silverberg, and C.H. Tsao. *Cosmic Ray Effects on Microelectronics, Part I: The Near-Earth Particle Environment*. August 25, 1981, Naval Research Laboratory, Washington D.C.: NRL Memorandum Report 4506.
29. Vampola, A.L., *Radiation effects on space systems and their modeling*, in *Space Systems and Their Interactions with Earth's Space Environment*, H.B. Garrett and C.P. Pike, Editors. 1980, AIAA: New York. p. 339.
30. Janni, J., and G. Radke, *The radiation environment and its effects on spacecraft*, in *Quantitative Modeling of Magnetospheric Processes*, W.P. Olson, Editor. 1979, American Geophysical Union: Washington, D.C. p. 634.

31. Feynman, J., T.P. Armstrong, L. Dao-Gibner, and S. Silverman, *A New Interplanetary Proton Fluence Model*. J. Spacecraft, 1989.
32. Adams, J.H., Jr., J.R. Letaw, and D.F. Smart. *Cosmic Ray Effects on Microelectronics, Part II: The Geomagnetic Cutoff Effects*. May, 1983, Naval Research Laboratory, Washington D.C.: NRL Memorandum Report 4506.
33. Tsao, C.H., R. Silberberg, J.H. Adams Jr., and J.R. Letaw. *Cosmic Ray Effects on Microelectronics, Part III: Propagation of Cosmic Rays in the Atmosphere*. August 9, 1984, Naval Research Laboratory, Washington D.C.: NRL Memorandum Report 4506.
34. Feynman, J., and S. Gabriel, eds. *Interplanetary Particle Environment, Proceedings of a Conference*. JPL 88-28. 1988, Jet Propulsion Laboratory: 167.
35. Luhn, A., et al., *Ionic Charge States of N, Ne, Mg, Si, S in Solar Particle Events*. Advances in Space Research, 1984. 4: p. 161-164.
36. Fischer, S., K. Vandas, B. Kuznetsov, and R. Ramatay, *Determination of the Effective Charge of Solar Cosmic Ray Nuclei Using the Earth's Magnetic Field*, in *Advances in Space Research*. 1984, pp. 169-172.
37. McWaldt, R.A., and E.C. Stone, *A Search for Deuterium Tritium and ^3He in Large Solar Flares*. 1983, Bull. of the Am. Phys. Soc. p. 742.
38. Breneman, H.H., and E.C. Stone, *Solar and Photospheric Abundance from Solar Energetic Particle Measurements*. Astrophys. J., 1985. 229: pp. L57-61.
39. Meyer, P.R., R. Ramaty, and R. Webber, *Cosmic Rays - Astronomy with Energetic Particles*, in *Physics Today*. 1974,
40. Smart, D.F., and M.A. Shea, (*Private Communication*). 1989
41. Jokipii, J.R., E.H. Levy, and W.B. Hubbard, *Effects of Particle Drift on Cosmic Ray Transport. I. General Properties*. Astrophys. J., 1977. 213: p. 861.
42. Fisk, L.A., B. Kpzlovsky, and R. Ramaty, *Astrophys. J. (Ltrs.)*, 1974. 190: p. 135-137.
43. Divine, T.N. *Interplanetary Charged Particle Models*. March, 1975, NASA: SP-8118.
44. Garrett, H.B., and G.S. Spitale, *Effects of the Solar and Extra-Terrestrial Environment on Space Power Systems*, in *Space Power*. 1984, NASA: p. 225-249.
45. Divine, T.N., and H.B. Garrett, *Charged particle distributions in Jupiter's magnetosphere*. J. Geophys. Res., 1983. 88(A9, Sept.): pp. 6889-6903.
46. Drake, F.D., and H. Havatum, *Non-thermal microwave radiation from Jupiter*. Astron. J., 1959. 64: pp. 329-330.
47. Brice, N.M., and G.A. Ioannidis, *The magnetospheres of Jupiter and Earth*. Icarus, 1970. 13: pp. 173-183.

48. Ioannidis, G., and N. Brice, *Plasma densities in the Jovian magnetosphere: Plasma slingshot or Maxwell demon?* Icarus, 1971. 14: pp. 360-363.
49. Gehrels, N., ed. *Jupiter*. 1976, University of Arizona Press: Tucson, AZ.
50. Dessler, A.J., ed. *Physics of the Jovian Magnetosphere*. Cambridge Planetary Science Series, eds. W.I. Axford, G.E. Hunt, and R. Greeley. 1983, Cambridge University Press: New York. 544.
51. Smith, E.J., L Davis, Jr., and D.E Jones, *Jupiter's magnetic field and magnetosphere*, in *Jupiter*, T. Gehrels, Editor. 1976, University of Arizona Press: Tucson, AZ. pp. 788-829.
52. Acuna, M.H., and N.F. Ness, *Results from the GSFC Fluxgate Magnetometer on Pioneer 11*, in *Jupiter*, T. Gehrels, Editor. 1976, University of Arizona Press: Tucson, AZ. p. 1976.
53. Acuna, M.H., and N.F. Ness, *The main magnetic field of Jupiter*. J. Geophys. Res., 1976. 81: pp. 2917-2922.
54. Anon. *Highlights of Conference in Physics of the Jovian and Saturnian Magnetospheres*. 1981. John Hopkins University.
55. Seidelmann, P.K., and T.N. Divine, *Evaluation of Jupiter longitudes in System III (1965)*. Geophys. Res. Lttrs., 1977. 4: pp. 65-68.
56. Carbary, J.F., *Periodicities in the Jovian magnetosphere: Magneto-disc models after Voyager*. Geophys. Res. Lttrs., 1980. 7: pp. 29-32.
57. Krimigis, S.M., et al., *Hot plasma environment at Jupiter: Voyager 2 results*. Science, 1979. 206 : pp. 977-984.
58. dePater, I., and H.A.C. Dames, *Jupiter's radiation belts and atmosphere*. Astron. Astrophys., 1979. 72 : pp. 148-160.
59. Krimigis, S.M., et al., *Low-energy charged particles in Saturn's magnetosphere: Results from Voyager 1*. Science, 1981. 212 (4491): pp. 225-231.
60. Divine, T.N., *Numerical Models for Electron and Proton Distributions in Saturn's Radiation Belts*. 1990, Jet Propulsion Laboratory.
61. White, O.R., ed. *The Solar Output and its Variation*. 1977, Colorado Associated University Press: Boulder, CO.
62. Garrett, H.B., and J.M. Forbes, *A Model of Solar Flux Attenuation During Eclipse Passage and Its Effects on Photoelectron Emission from Satellite Surfaces*. Planet. Space Sci., 1981. 29: pp. 601-607.
63. Allen, J.H., *GOES X-Ray Measurements*. EOS, 1985.
64. Dolan, P. 19. *Nuclear—EMP, X-Ray, Thermal, Nuclear Radiation*. in *Space Systems and Technology Workshop II; Natural Environments Panel Proceedings*. 1982. Los Angeles, CA: USAF Space Division.

65. Kuller, W.G., and D.W. Hanifen, *Fragility of Space Operation to Exoatmospheric Nuclear Bursts*. 1980, Air Force Weapons Laboratory: Kirtland AFB, NM. p. 35.
66. Dolan, P., *Part I: "Phenomenology"; Part II: "Damage Criteria"*, in *Capabilities of Nuclear Weapons (U)*, P.J. Dolan, Editor. 1972, Defense Nuclear Agency: Washington, DC.
67. Gladstone, S., and P.J. Dolan, *The Effects of Nuclear Weapons*. 1977, Washington, DC: US Dept. of Defense and US Dept. of Energy.
68. Kuller, W.G., J.F. Janni, J.A. Frazier, J.L. Mullis, and D.R. Hollars, *White Paper on Nuclear Effects on the Space Based Radar System*. 1979, Air Force Weapons Laboratory: Kirtland AFB, NM. p. 24.
69. Mannheimer, H. *Environmental Impact Statement for Galileo and Ulysses Missions*. January, 1986, NASA Office of Space Transportation:
70. Salisbury, D.F., *Radiation Risk and Planetary Exploration—The RTG Controversy*. The Planetary Report, 1987. VII (3): pp. 3-7.
71. Hoffman, A. *Galileo Orbiter Functional Requirements Book, Environmental Design Requirements*. May 21, 1987, JPL: GAL-3-240, Rev. C.
72. Particle Data Group, *Passage of Particles through Matter; Review of Particle Properties*. Phys. Lttrs., B, 1990. 239 (III): pp. 1-38.
73. Winokur, P.S. *Total-Dose Radiation Effects (From the Perspective of the Experimentalist)*. in *IEEE Nuclear and Space Radiation Effects Conference Short Course*. 1992. New Orleans, LA: IEEE NSREC.
74. Berger, M.J., and S.M. Seltzer. *Penetration of Electrons and Associated Bremsstrahlung through Aluminum Targets*. 1968, NASA, Washington, DC: SP-169, pp. 285-322.
75. Janni, J.F., *Proton Range-Energy Tables, 1 keV-10 GeV, Part 1*. 1982, Atomic Data and Nuclear Data Tables. pp. 150-339.
76. Janni, J.F., *Proton Range-Energy Tables, 1 keV-10 GeV, Part 2*. 1982, Atomic Data and Nuclear Data Tables. pp. 341-529.
77. Curgenven, L, and P. Duncumb. *Report 303*. July 28, 1971, Tube Investments Research Laboratories:
78. Brown, D.B. *Total Dose Effects at Dose Rates Typical of Space*. in *NSREC Short Course*. 1990. IEEE NSREC.
79. Jordon, T.M. *Electron Dose Attenuation Kernels for Slab and Spherical Geometries*. November, 1981, AF Weapons Laboratory: AFWL-TR-81-43.
80. Stassinopoulos, E.G., *Electron Depth-Dose Profiles in CaFMn TLDs (Private Communication)*. 1993.
81. Stassinopoulos, E.G., *Proton Energy Deposition in CaF₂ (Private Communication)*. 1993.

82. Jordon, T.M. *NOVICE: A Radiation Transport/Shielding Code; Users Guide*. January 2, 1987, Experimental and Mathematical Physics Consultants: 87.01.02.01.
83. Fermi, E., *Nuclear Physics*. 1950, Chicago, IL: University of Chicago Press.
84. Evans, R.D., *The Atomic Nucleus*. 1982, McGraw-Hill, 1955 and Kliegler Publishing, 1982.
85. van Lint, V.A.J., T.M. Flanagan, R.E. Leadon, J.A. Naber, and V.C. Rogers, *Mechanisms of Radiation Effects in Electronic Materials*. Vol. 1. 1980, New York, NY: Wiley-Interscience.
86. Leung, P., A.C. Whittlesey, H.B. Garrett, P.A. Robinson Jr., and T.N. Divine, *Environment-Induced Electrostatic Discharges as the Cause of Voyager 1 Power-On Resets*. *J. Spacecraft and Rockets*, 1986. 23(No. 3, May/June 1986): pp. 323-330.
87. Adams, J.H., Jr. *The Ionizing Particle Environment Near Earth*. in *AIAA Aerospace Sciences Meeting*. 1982. AIAA.
88. Robinson, P.A., Jr., ed. *Introduction to Spacecraft Environments and the Anomalies They Cause*. June 15, JPL D-5489. 1988, Jet Propulsion Laboratory.
89. Baker, D.N., R.D. Belian, P.R. Higbie, and E.W. Hones, Jr. *The use of >30 KeV electron anisotropies at 6.6 R_e to predict magnetospheric substorms*. in *Solar-Terrestrial Predictions Proceedings*. 1980. NOAA.
90. Claflin, E.S., and R.S. White, *A study of equatorial inner belt protons from 2 to 200 MeV*. *J. Geophys. Res.*, 1974. 79: p. 959.
91. Slavin, J., and et al. *Planetary Mach Cones: Theory and Observation*. 1984, Jet Propulsion Laboratory.
92. Rudie, N.J., *Principles and Techniques of Radiation Hardening*. Vol. I-III. 1976, Hollywood, CA: Western Periodicals Co. pp. 1-17.
93. Sexton, F.W. *Measurement of Single Event Phenomena in Devices and ICs*. in *IEEE Nuclear and Space Radiation Effects Conference Short Course*. 1992. New Orleans, LA: IEEE NSREC.

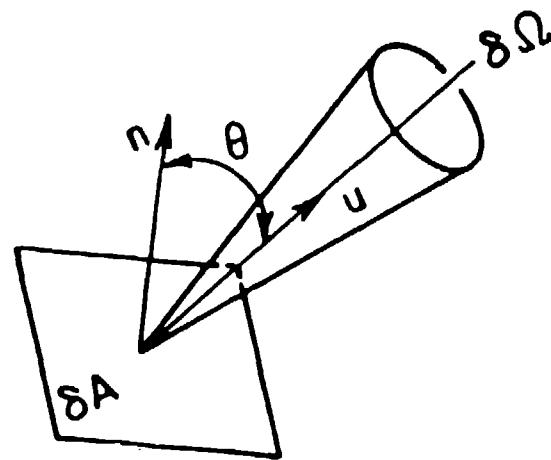


Fig. 1. The flux (number of particles or photons per unit time) of a given energy E per unit energy interval dE in a unit solid angle ($d\Omega=2\pi \cos \theta d\theta d\phi$) about the direction of observation (in the θ, ϕ direction), incident on unit of surface area (dA) perpendicular to the direction of observation. [2]

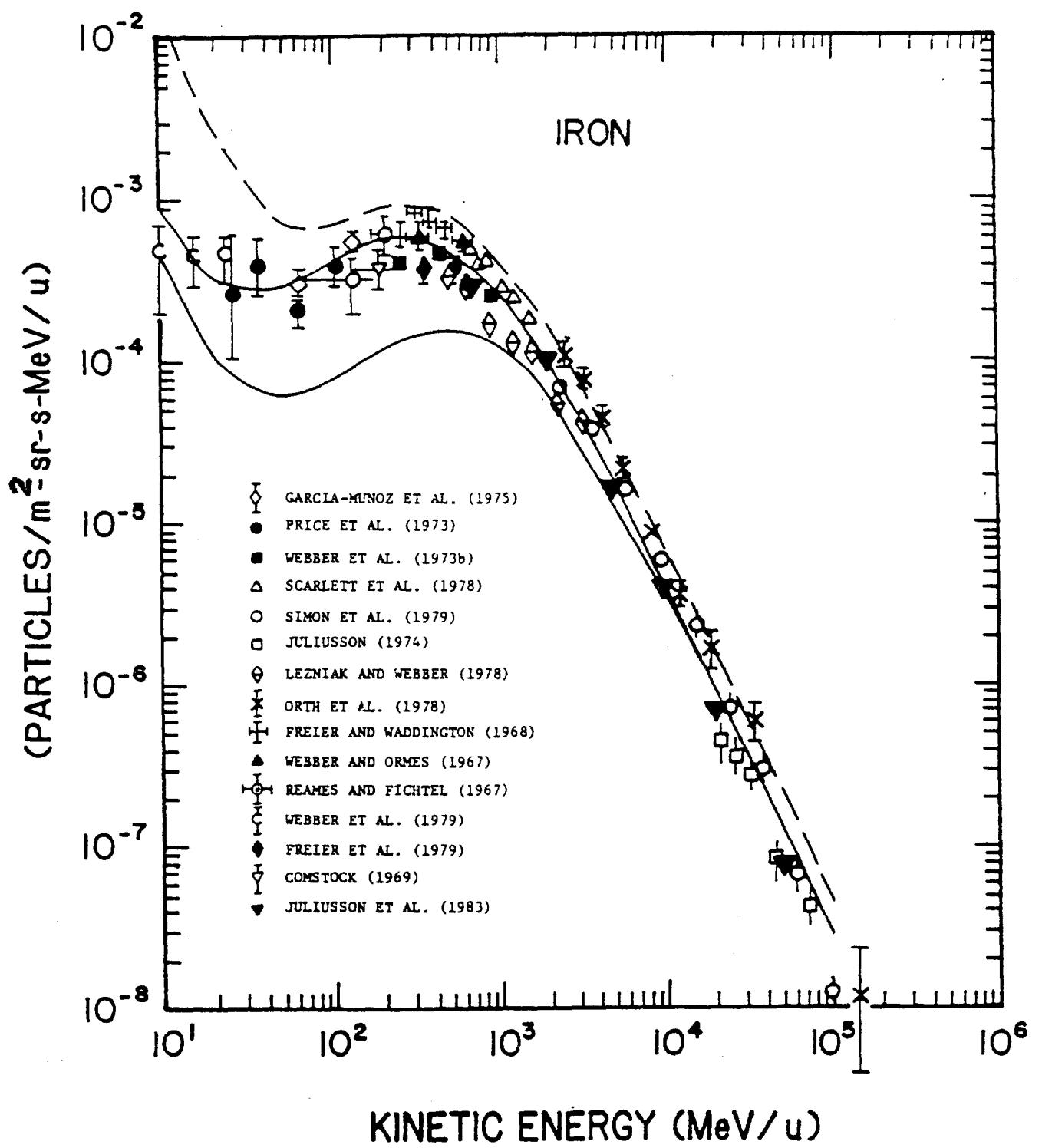


Fig. 2. The cosmic ray iron spectrum: The solid curves are for solar maximum (lower) and solar minimum (upper). The dashed curve is the 90% worst-case iron spectrum, which is implied by comparison with the cosmic ray helium spectrum.^[3]

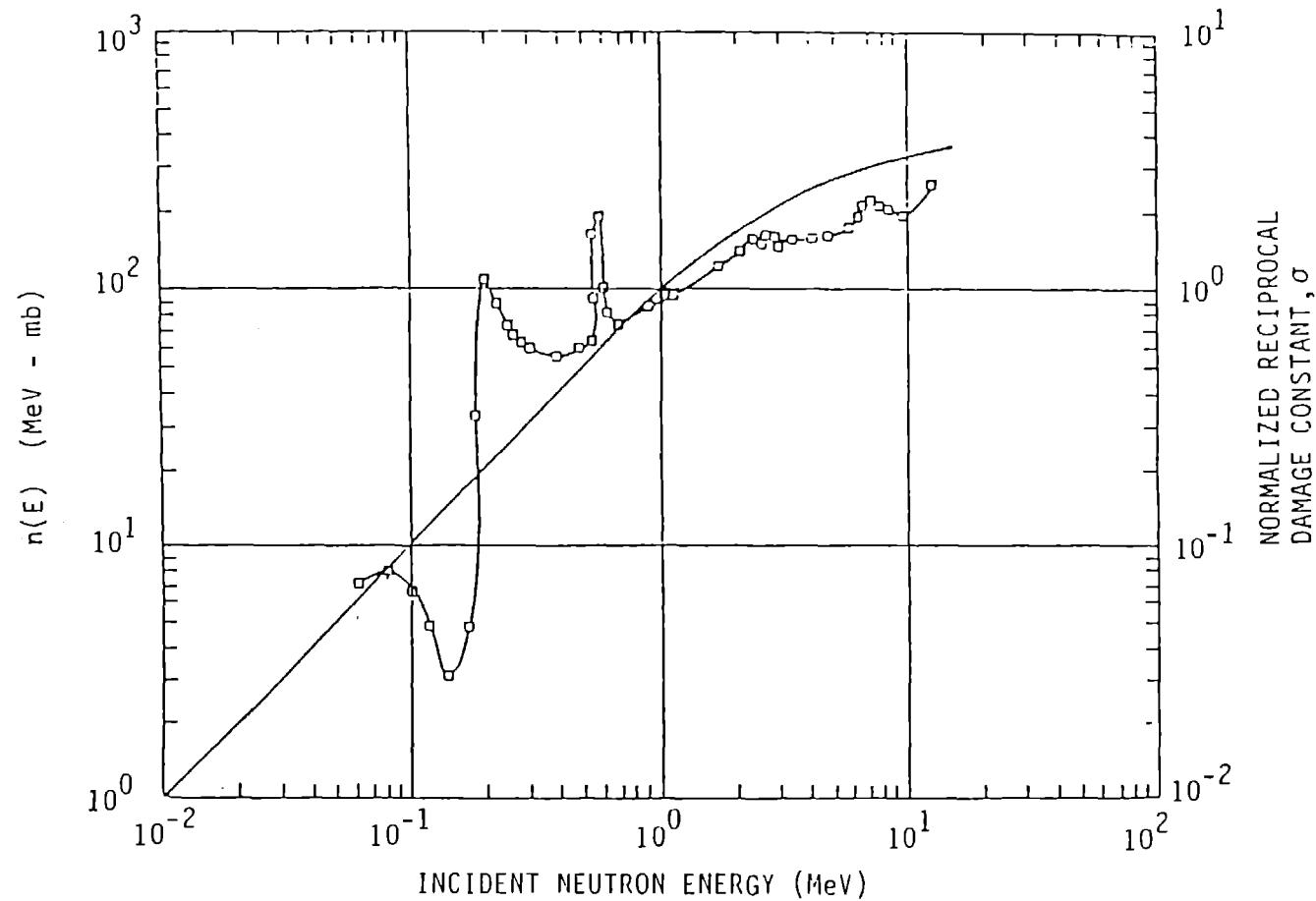


Fig. 3. Illustration of a neutron displacement damage equivalence curve. The curves (actual and fitted), in units of MeV- mb (where b is a barn or 10^{-24} cm^2), are for neutrons. The "relative displacement damage" is defined in terms of the cross section times the energy of the incident particle. As an example, for 14 MeV neutrons, the 1-MeV neutron dose damage equivalent is given by multiplying the 14 MeV dose by 2.5.[4]

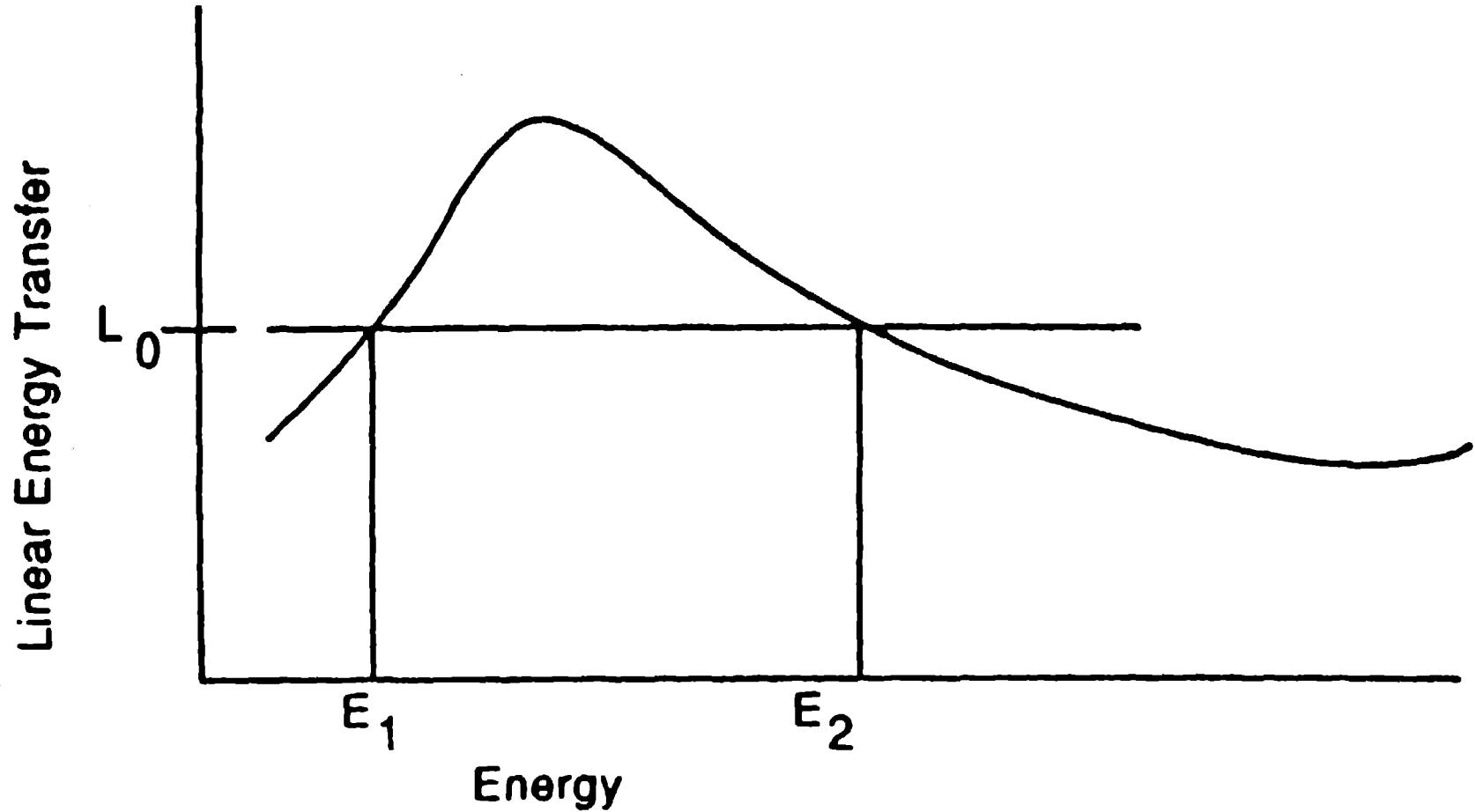


Fig. 4. Linear Energy Transfer Function (LET) versus Energy. L_0 corresponds to a constant value of LET. As illustrated, there can be multiple values of energy (E_1 and E_2 here) that correspond to the same value of LET.

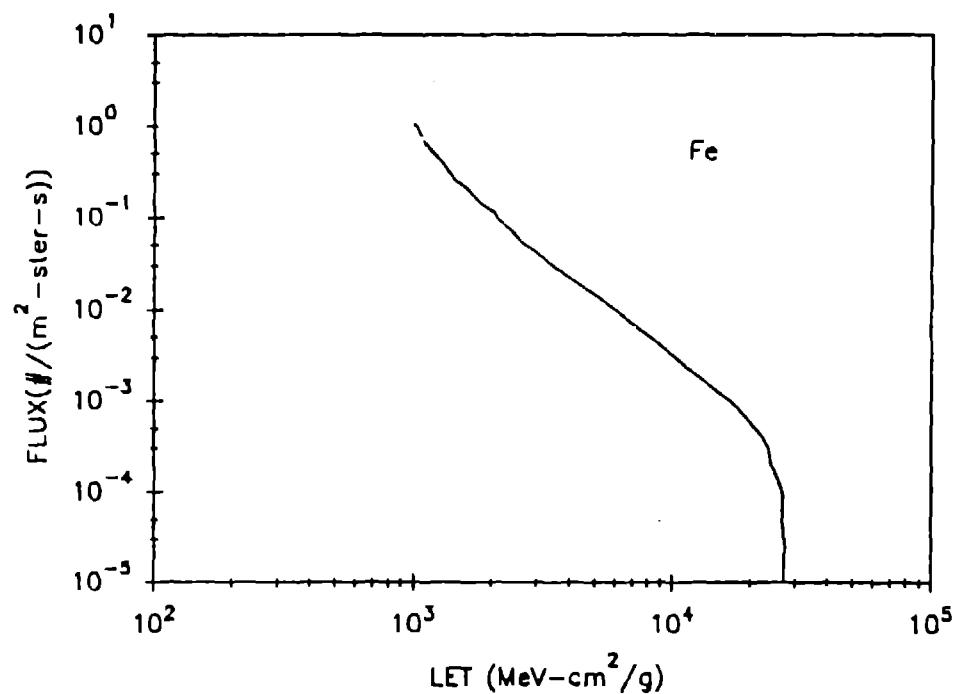


Fig. 5. Integral LET spectrum for iron.^[5]

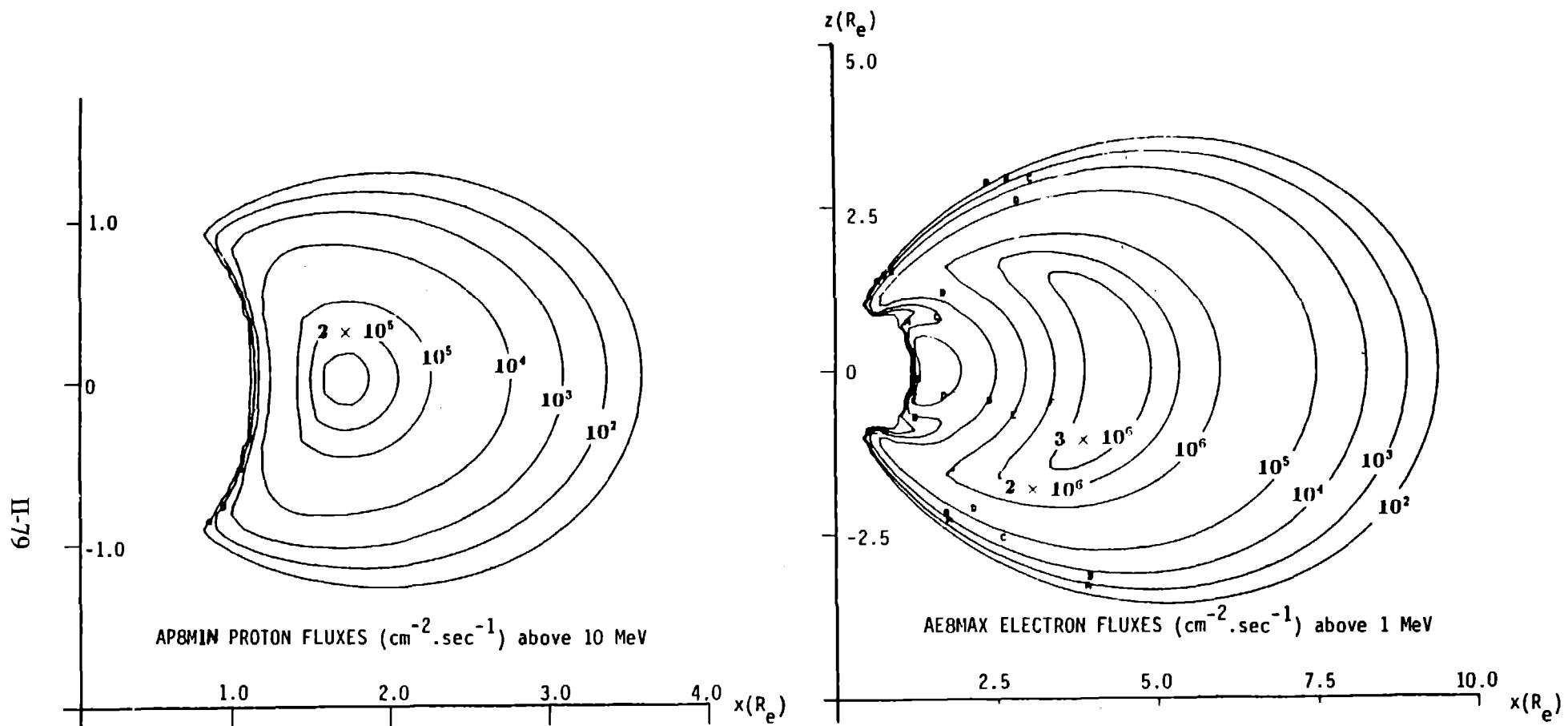


Fig. 6. The Earth's radiation belts in idealized dipole space, according to the AP8 and AE8 models. Average omnidirectional integral fluxes above energy thresholds are shown.[8]

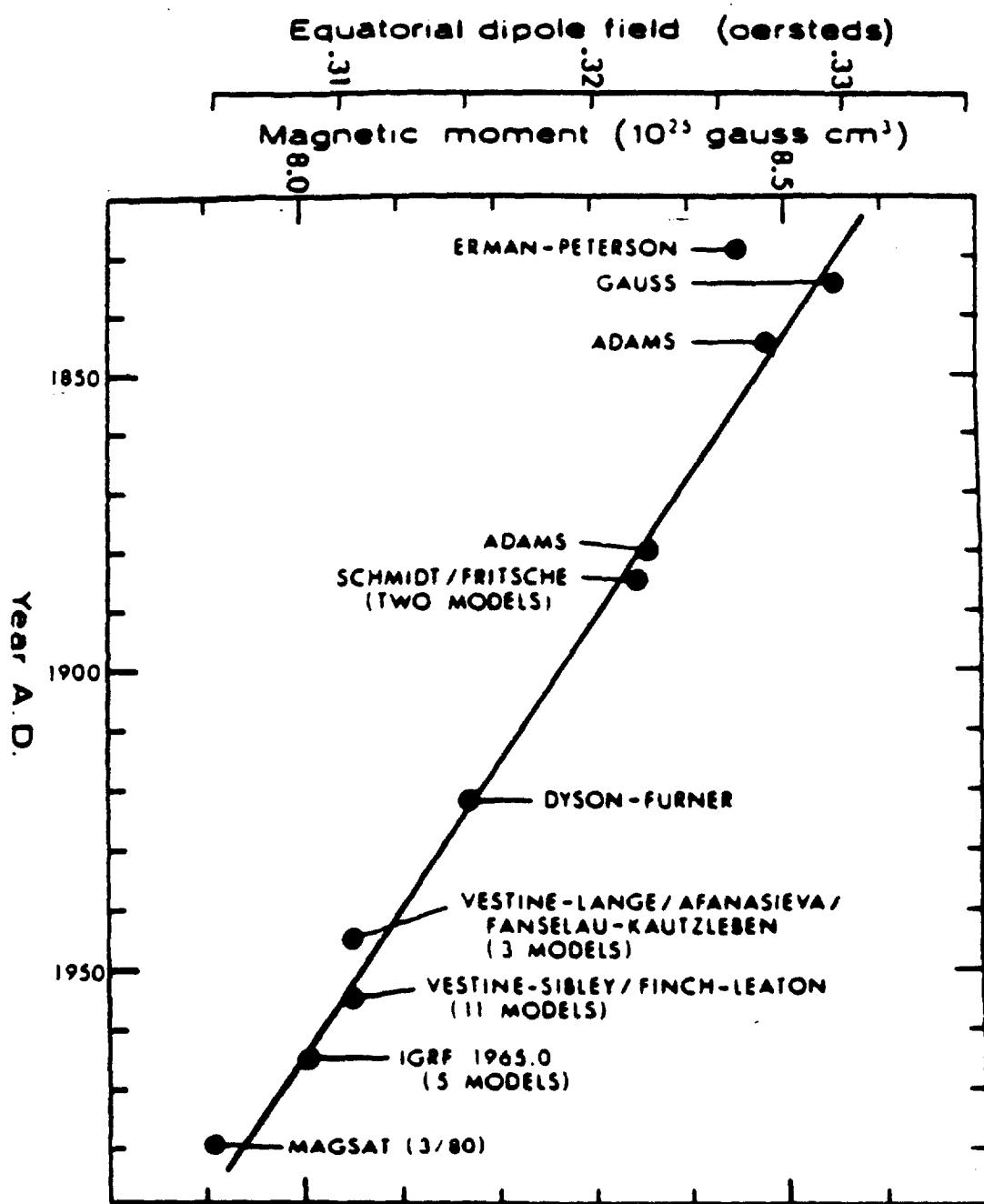
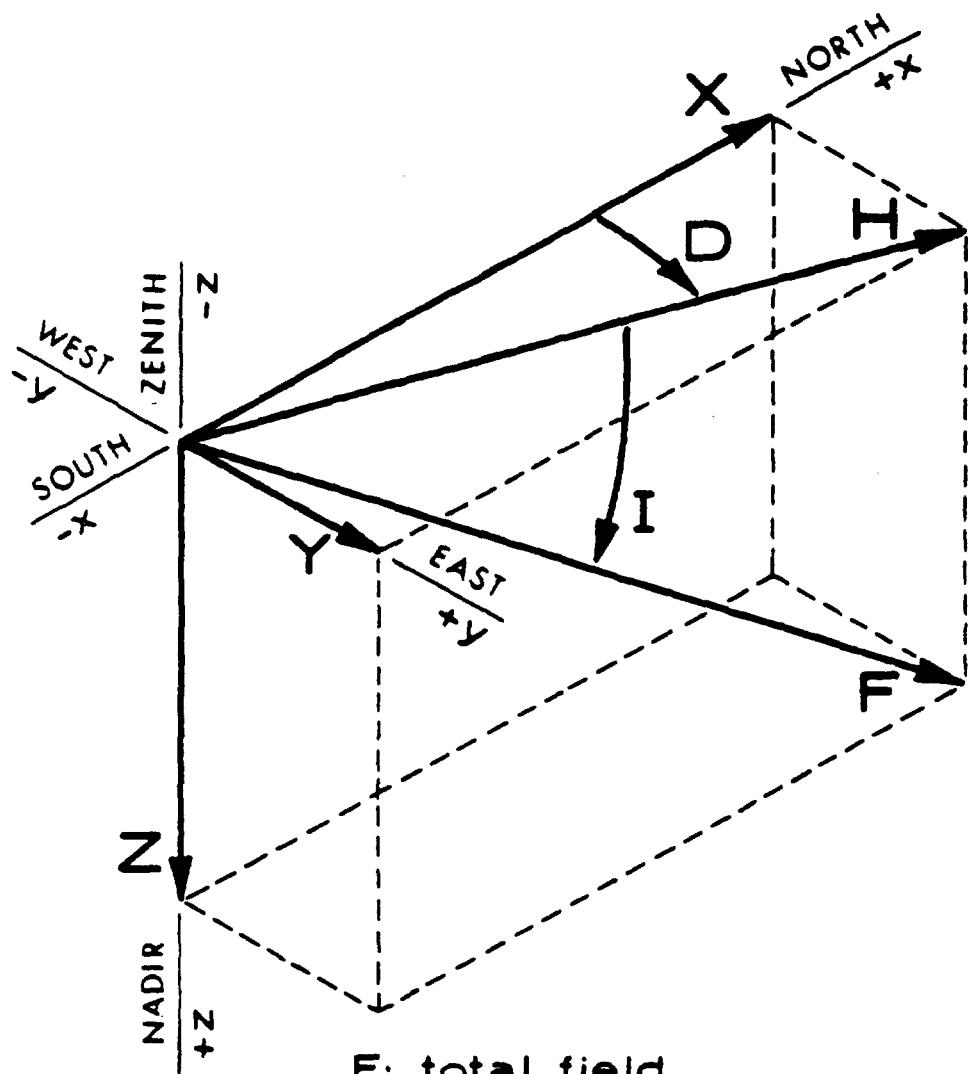


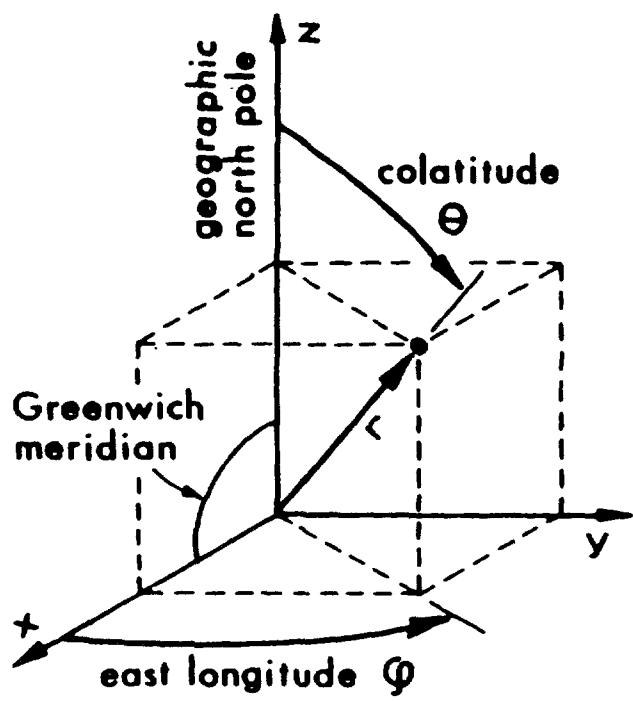
Fig. 7. Decay of the Equatorial Dipole Field Strength with Time. [9]



- F: total field
- H: horizontal component
- X: northward component
- Y: eastward component
- Z: vertical component
- D: declination
- I: inclination

Fig. 8. Geomagnetic Field Magnetic Elements^[9].

Geographic



Geomagnetic

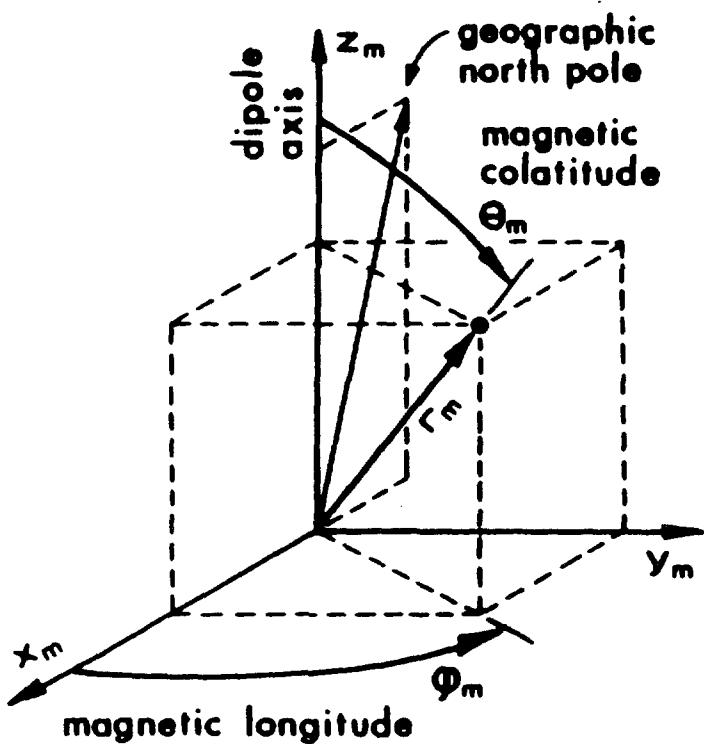
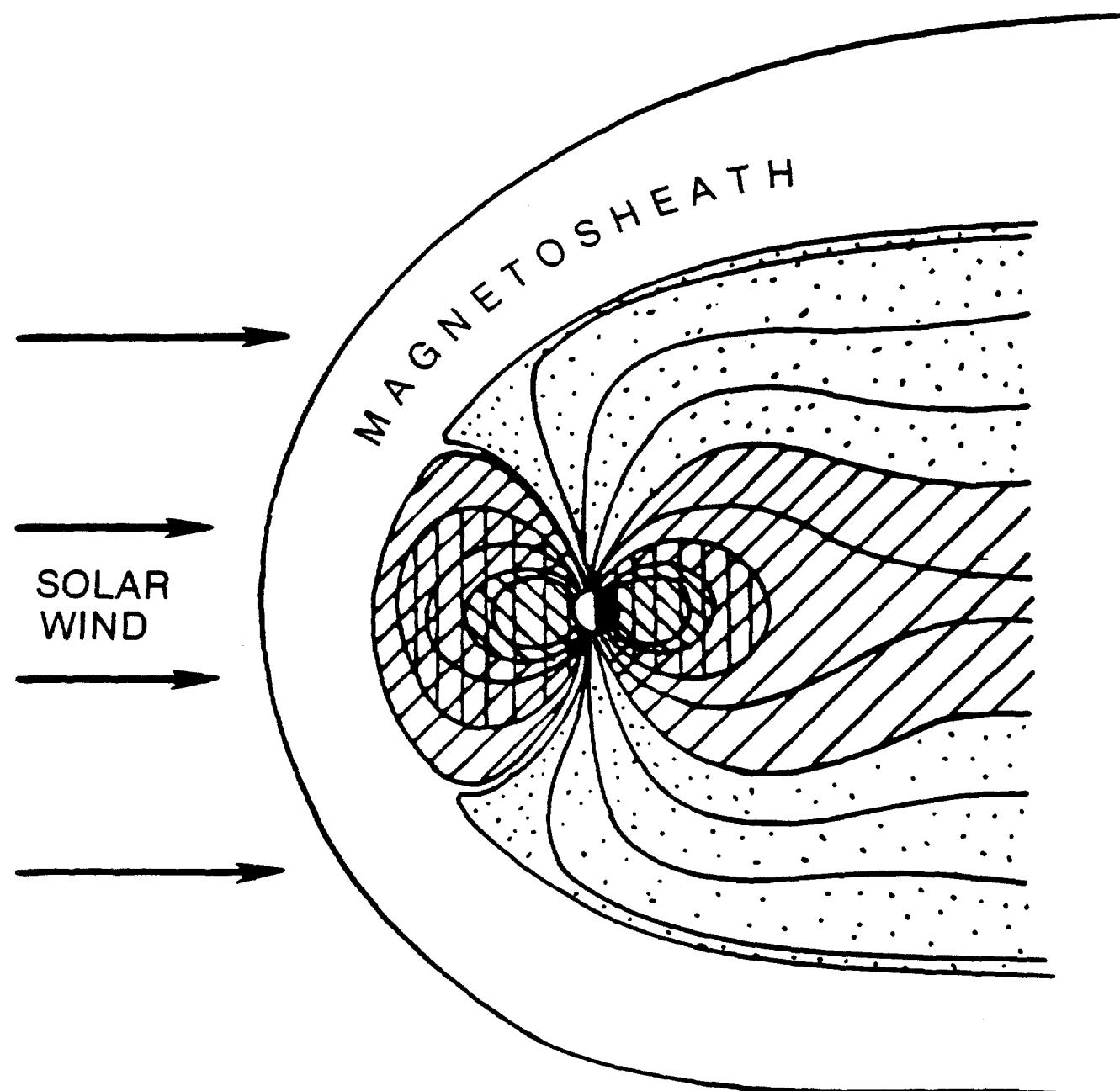


Fig. 9. Spherical Coordinate System for Geographic and Geomagnetic Coordinates. [9]



PLASMA-SPHERE PLASMA-SHEET POLAR VAN ALLEN BELTS

Fig. 10a. Cross section of the Earth's magnetic field in the noon-midnight meridian showing the structure of the field lines and the plasma regions they contain.

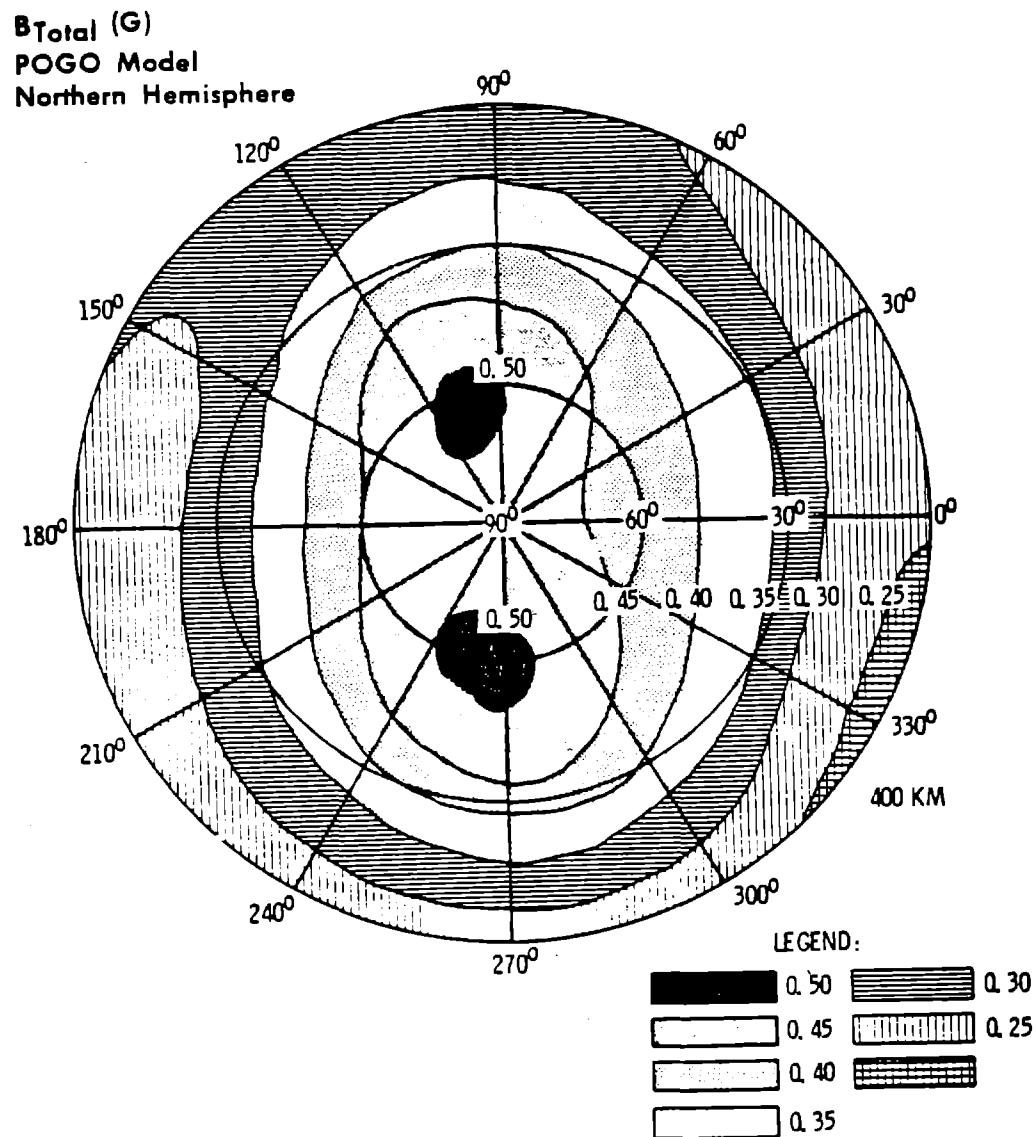


Fig. 10b. Illustration of the magnetic field amplitude over the northern hemisphere at a constant altitude of 400 km. The field amplitude varies from a minimum of 0.25 G near the equator to 0.5 G over the polar caps.

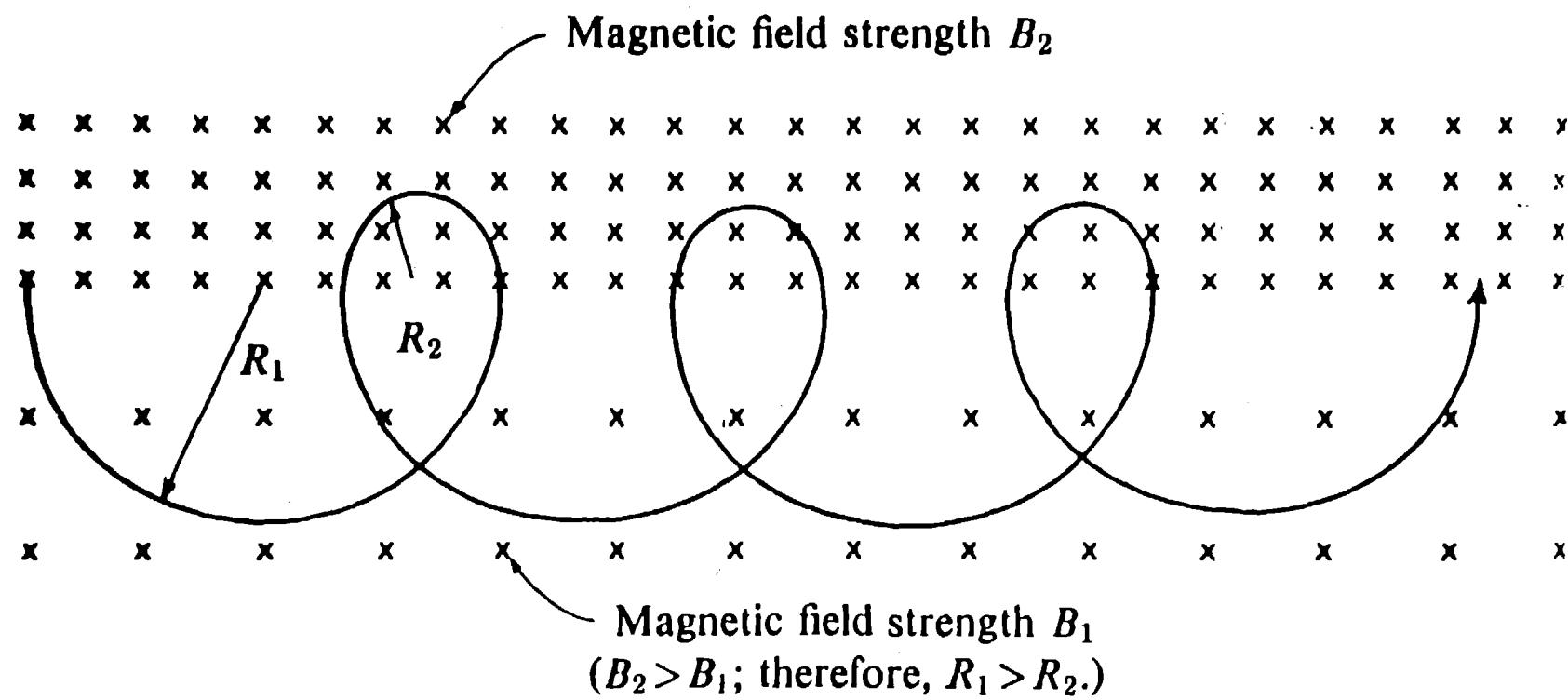


Fig. 11. Motion of a charged particle (positive in this case) in a magnetic field in the absence of an electric field. The magnetic field is constant in the lower half of the figure and equal to B_1 . It is constant and equal to B_2 in the upper part of the figure. The figure illustrates the effects of a gradient in a direction perpendicular to the direction of the magnetic field.

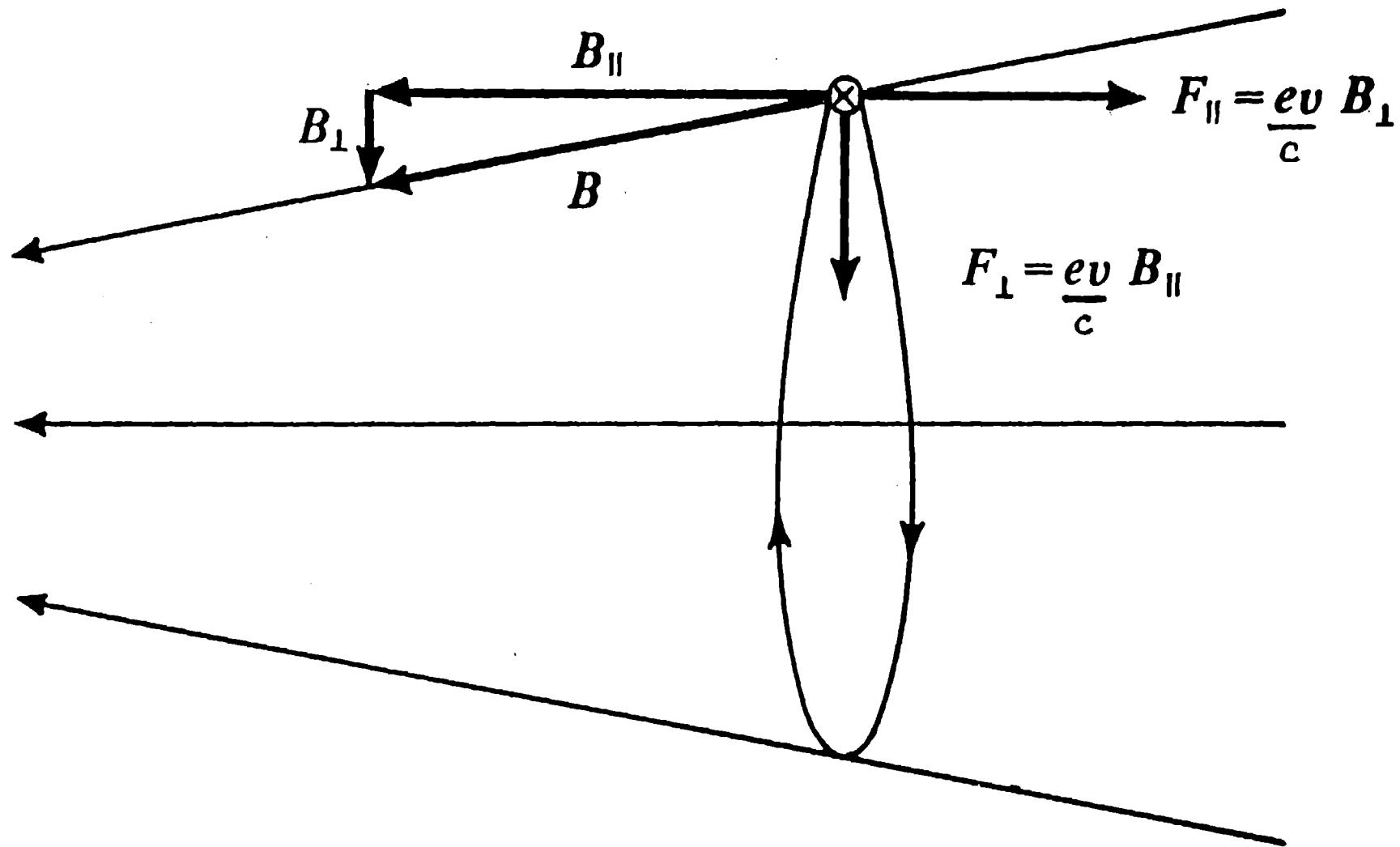


Fig. 12. Motion of a charged particle in a converging magnetic field in the absence of an electric field. F_{\perp} is the force along the magnetic field that results from the field convergence (or divergence) and is responsible for the mirroring effect (see text).

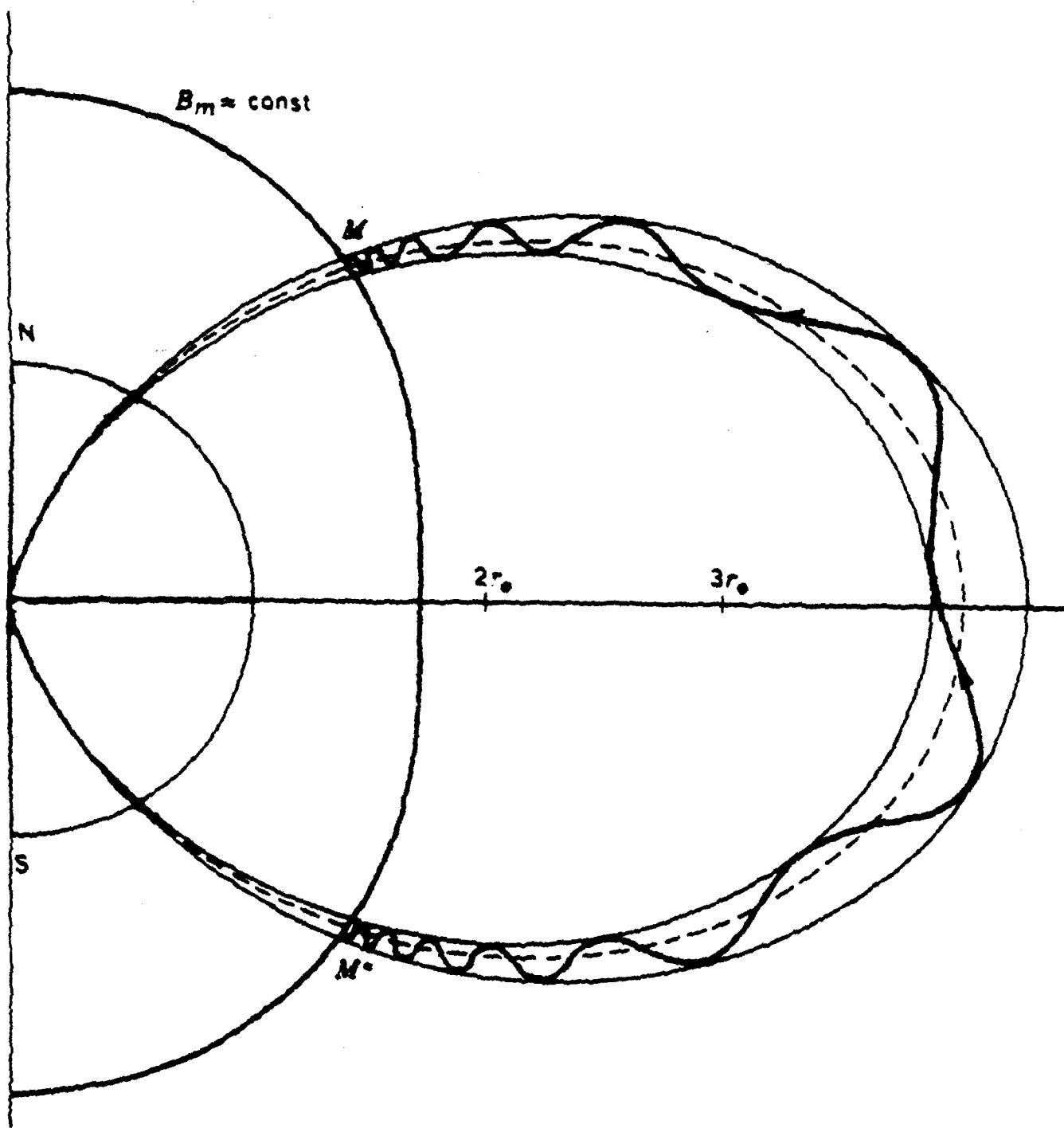


Fig. 13. Motion of a Charged Particle in a Dipole Magnetic Field.

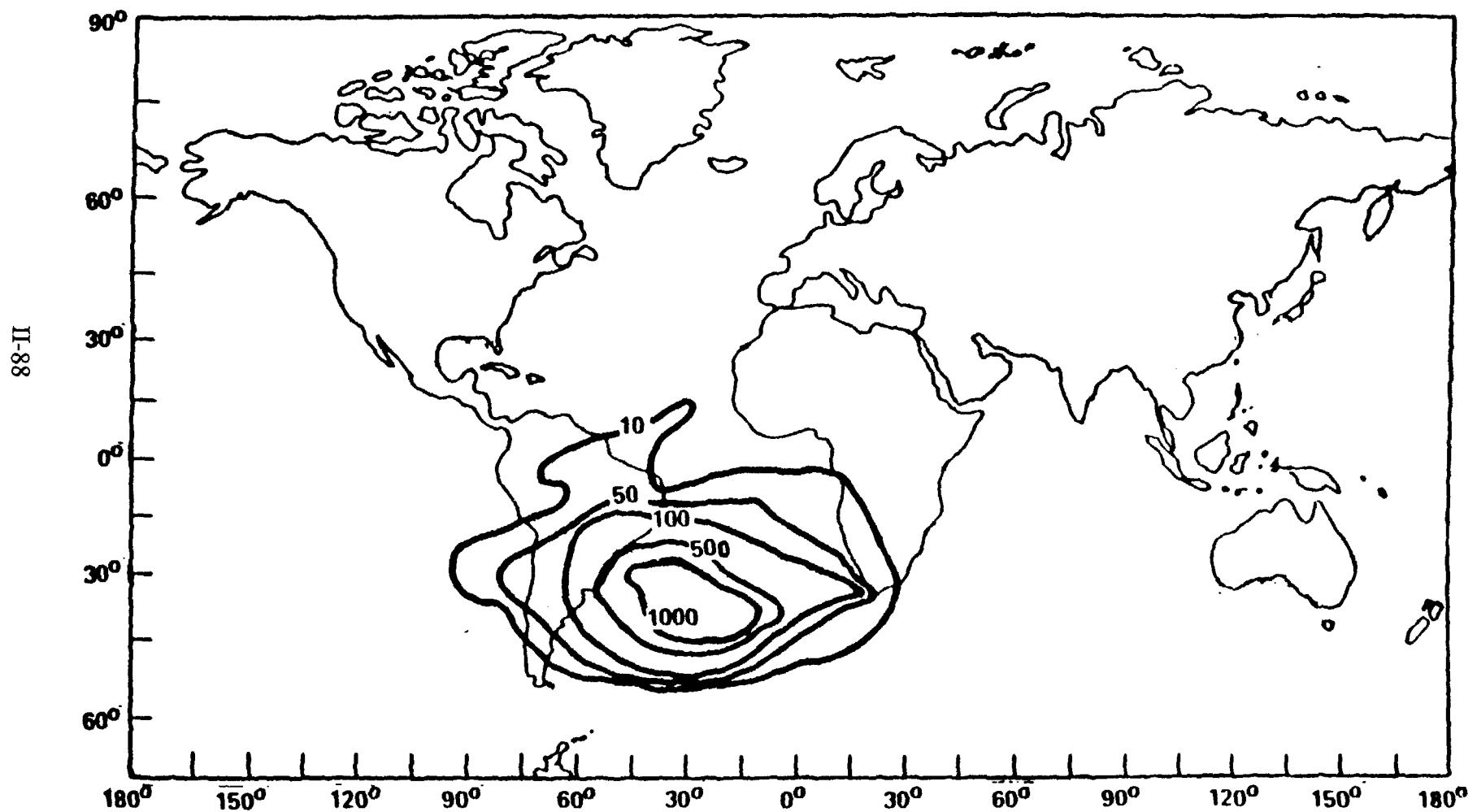


Fig. 14. Proton Flux Densities at 296 km Altitude. [19]

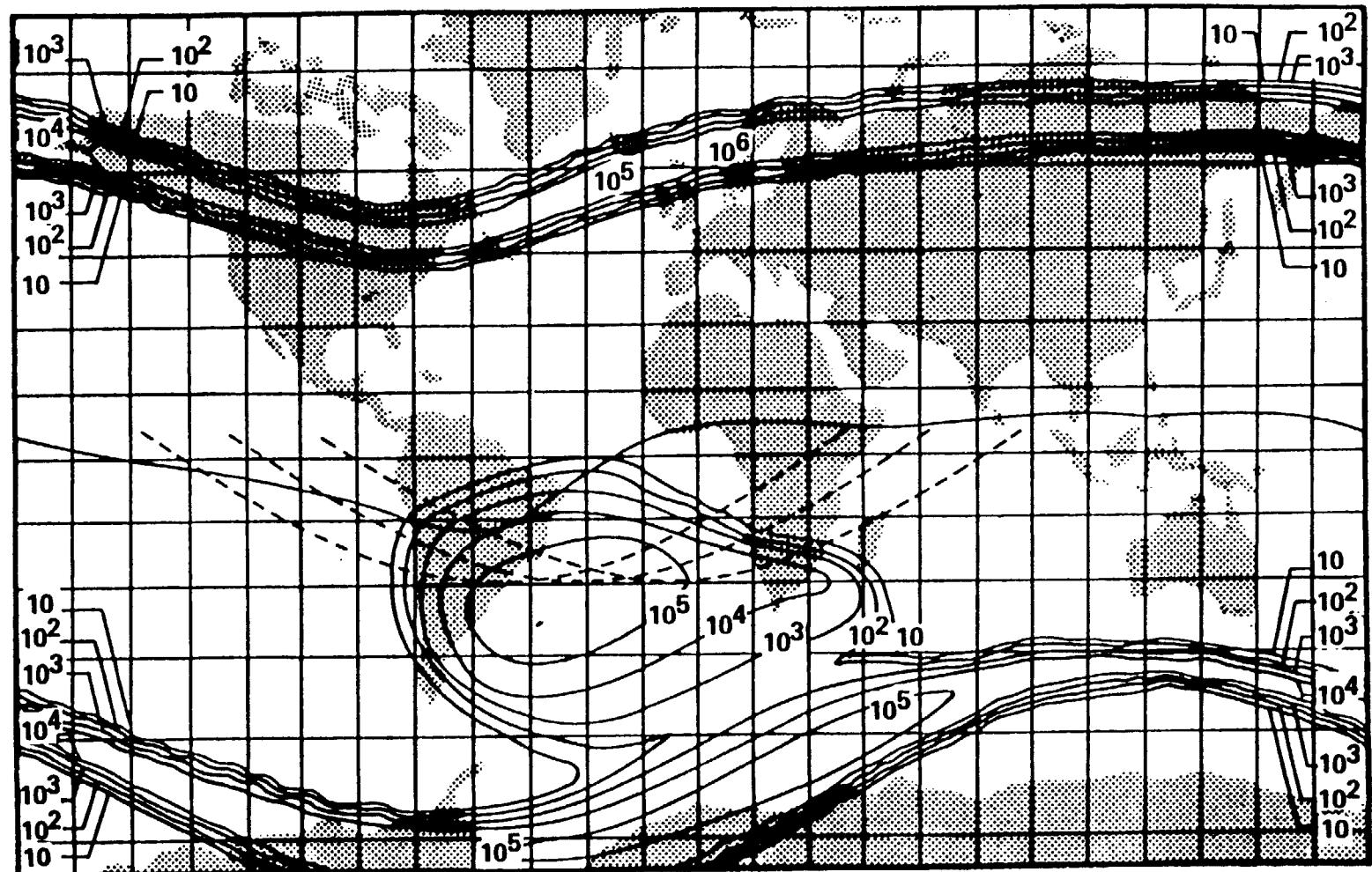


Fig. 15. Electron Constant Flux Contours at 400 km Altitude.($E > 0.5$ MeV)[19]

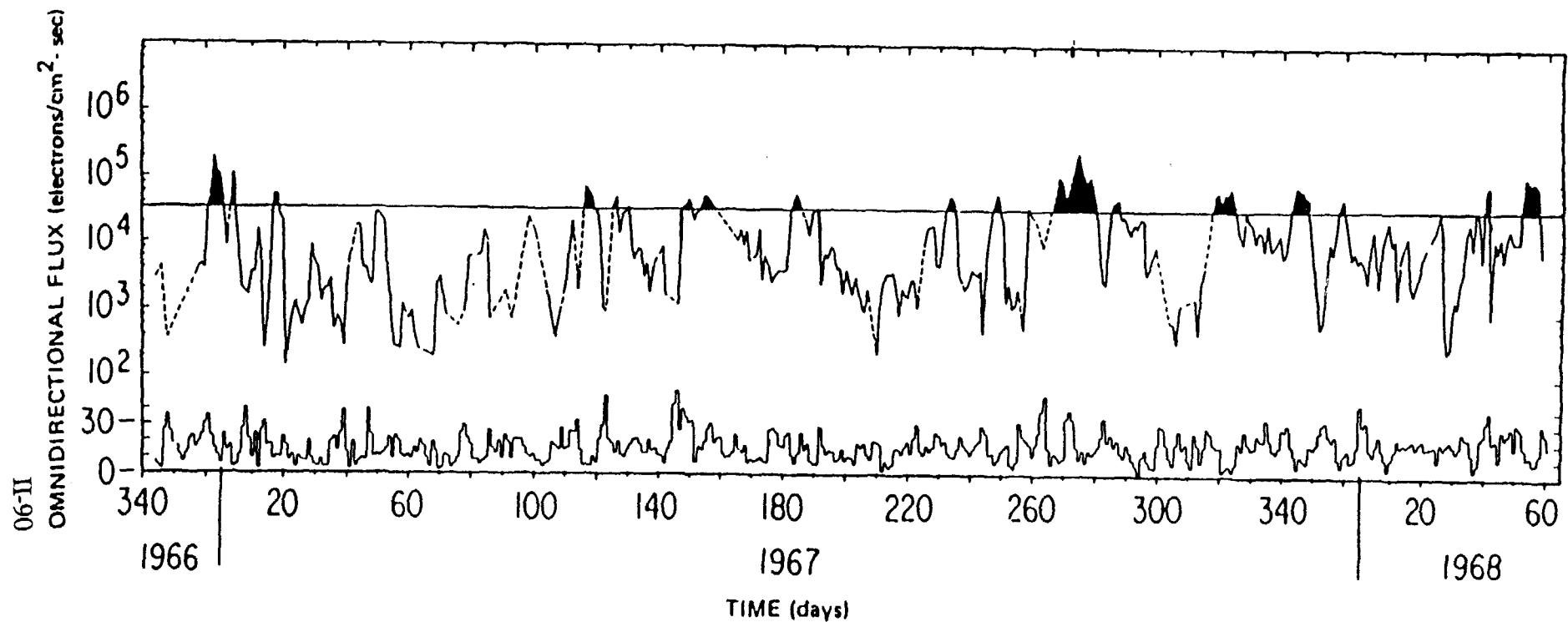


Fig. 16. Hourly averages of ATS1 1.9 MeV omnidirectional electron flux at local midnight.[20] Daily sum of K_p is plotted at the bottom of the figure. The horizontal line is the AEI-7 model equatorial flux at $L = 6.6$ and local midnight.

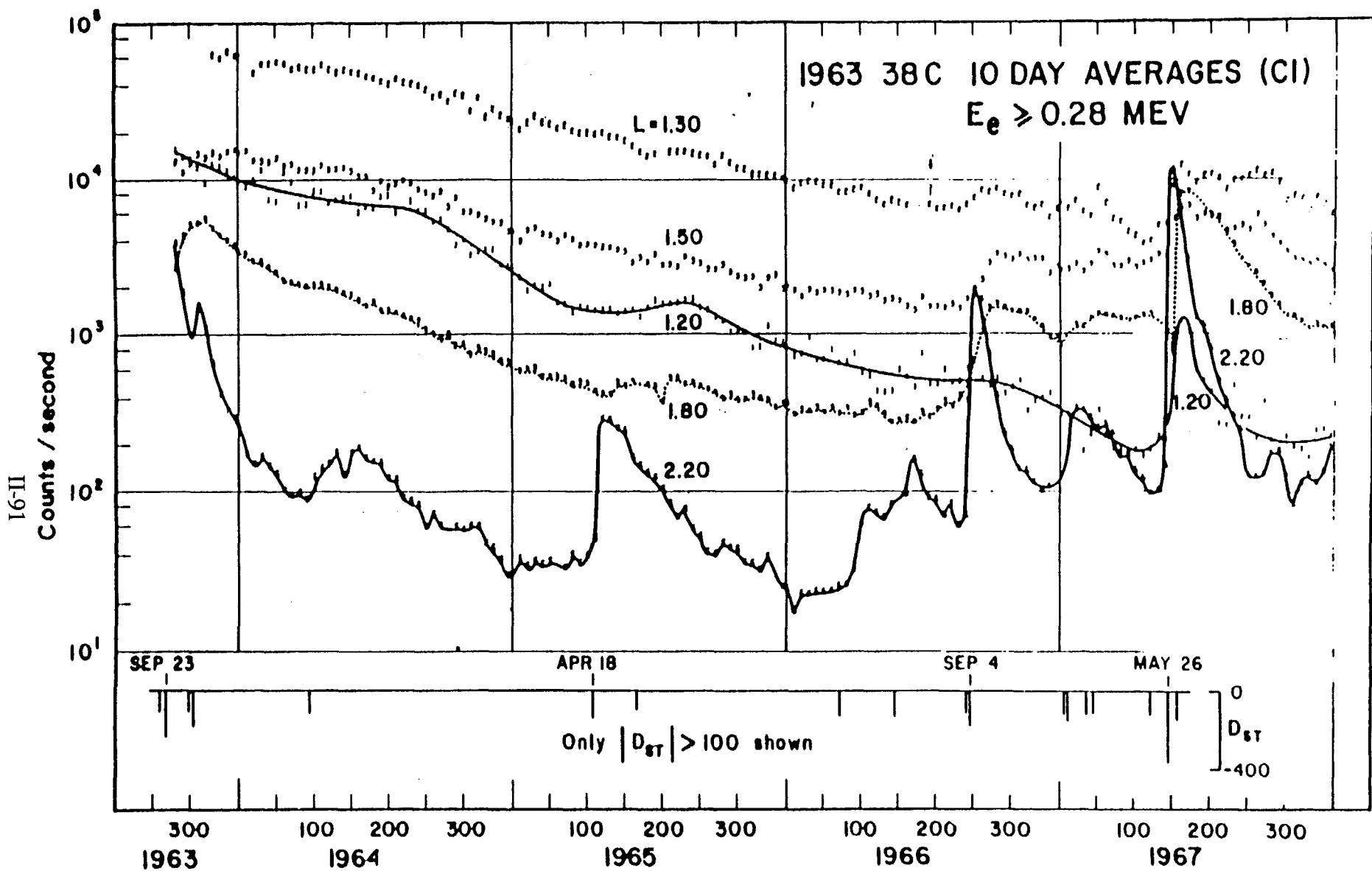


Fig. 17. 10 day averages of the inner zone electron fluxes greater than 0.28 MeV (in units of count rate) measured by the satellite 1963-38C. The effects of the decay of the Starfish nuclear debris and of the 4 major geomagnetic perturbations are evident.^[20]

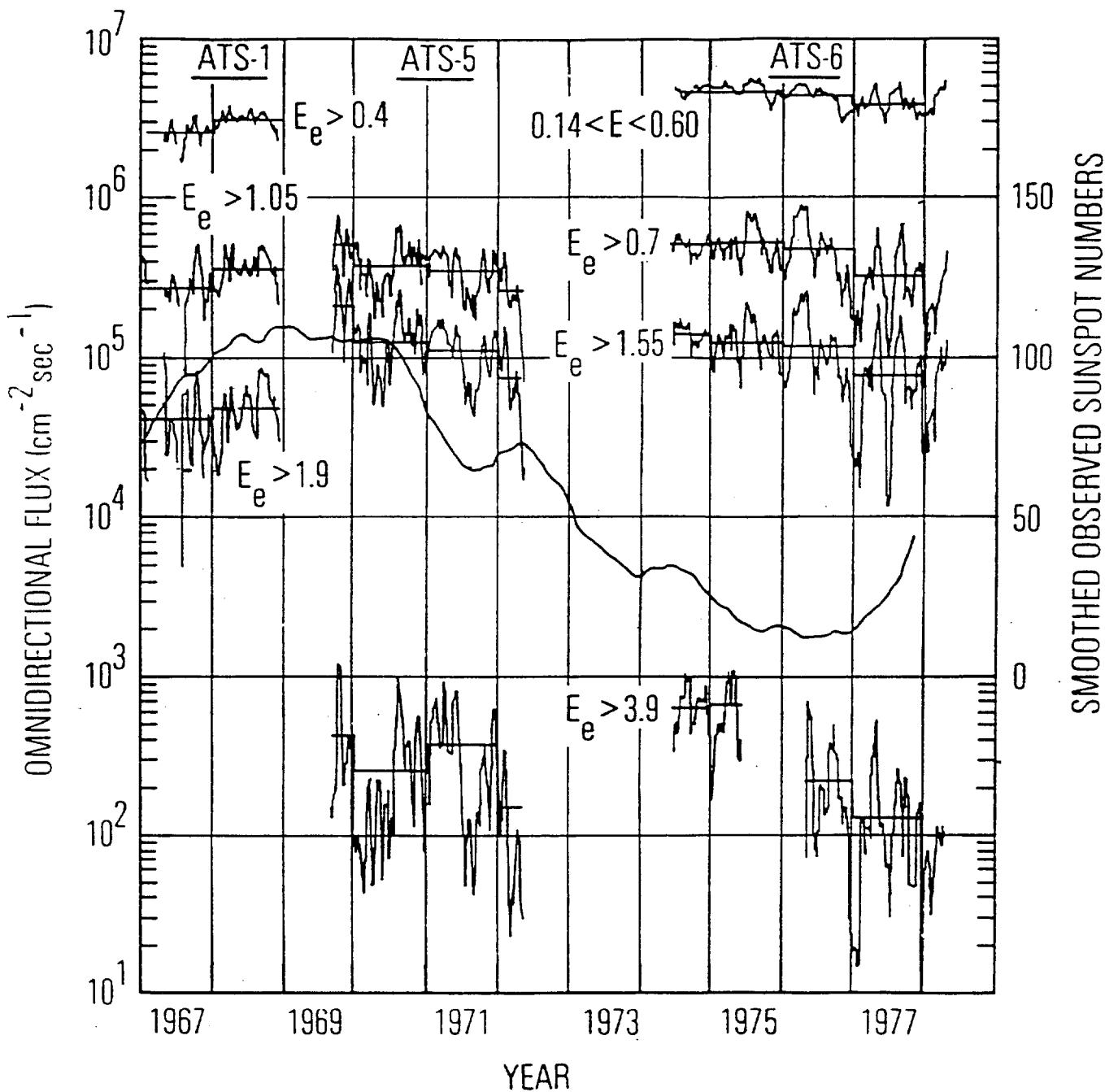


Fig. 18. ATS-1, ATS-5, and ATS-6 energetic electron fluxes (running 27-day averages) as a function of time^[21]. ATS-5 data were normalized to ATS-6 data in mid-1974. The energetic thresholds for the ATS-1, ATS-5, and ATS-6 channels are shown on the figure. The flux averages for each year are also indicated (solid horizontal lines). Superimposed on this graph are the Zurich monthly sunspot numbers.

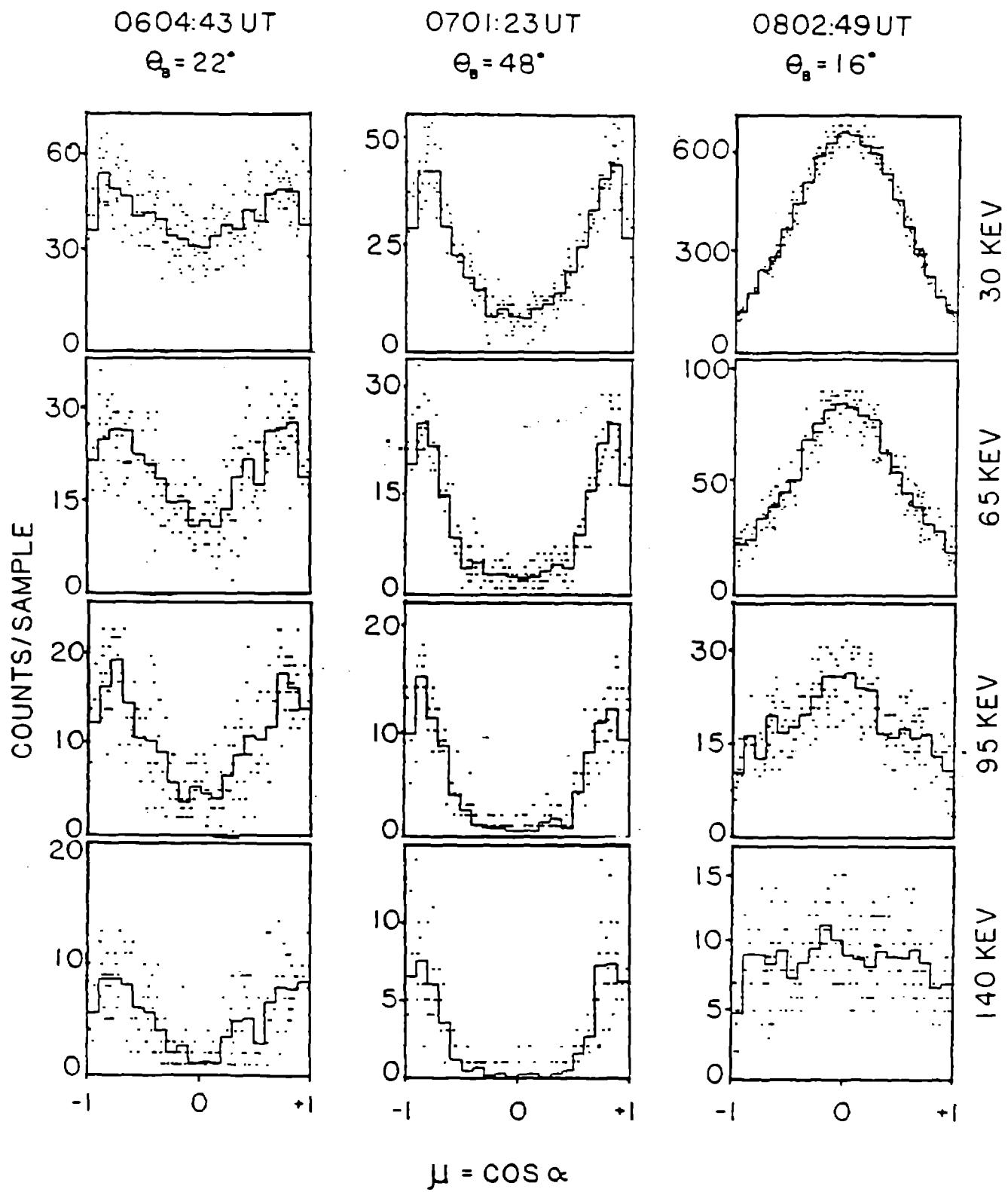


Fig. 19. Pitch angle distributions observed at selected times during the geomagnetic substorm of September 8, 1977. Electron counts per 8 ms sample plotted versus μ , the cosine of the pitch angle (α), for various energies. [89]

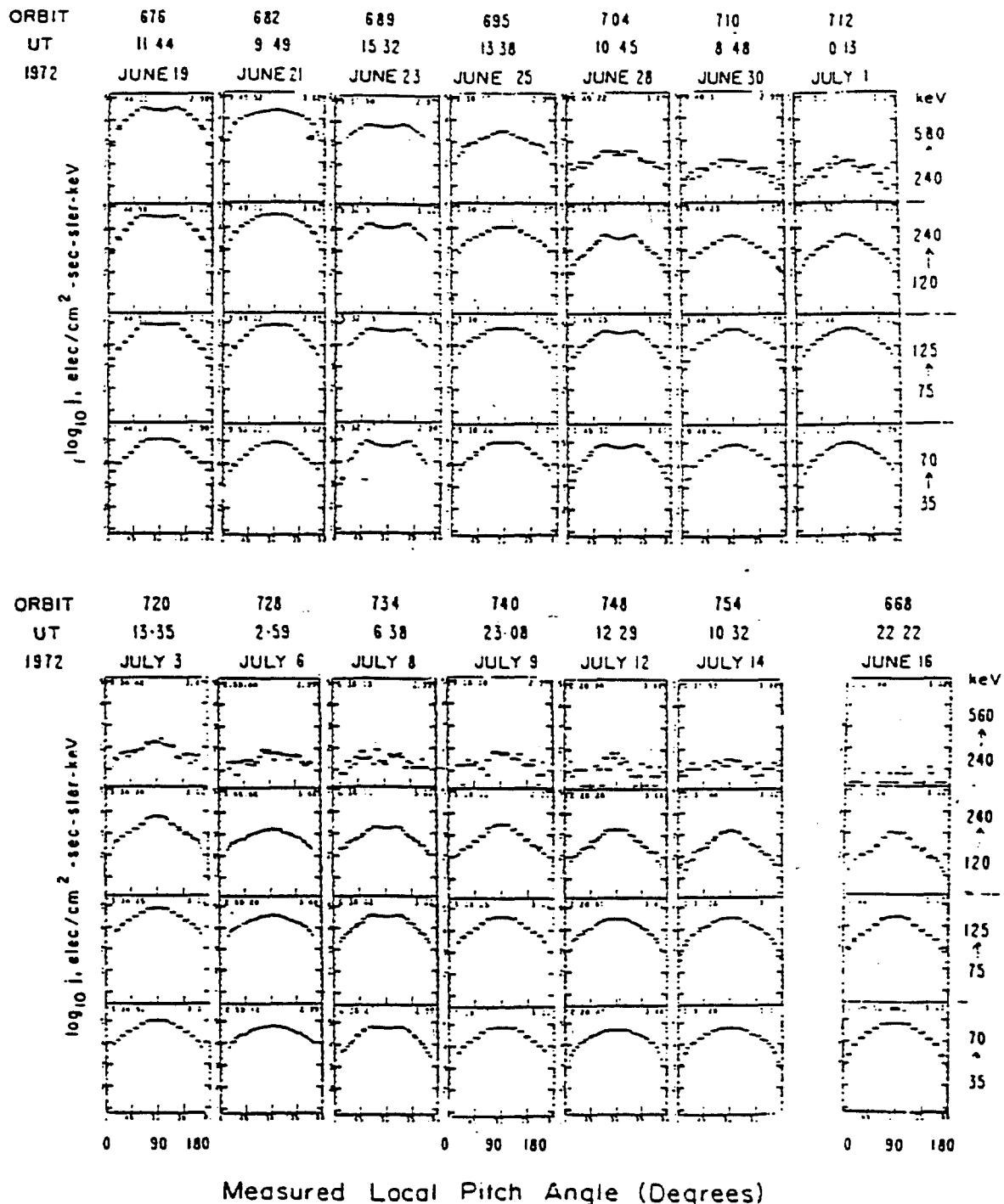


Fig. 20. Explorer 45 observations of the electron pitch angle distribution at $L = 3$, approximately once every two days, for the period following the storm of June 17, 1972. Prestorm distributions for June 16, 1972 are also shown. [22]

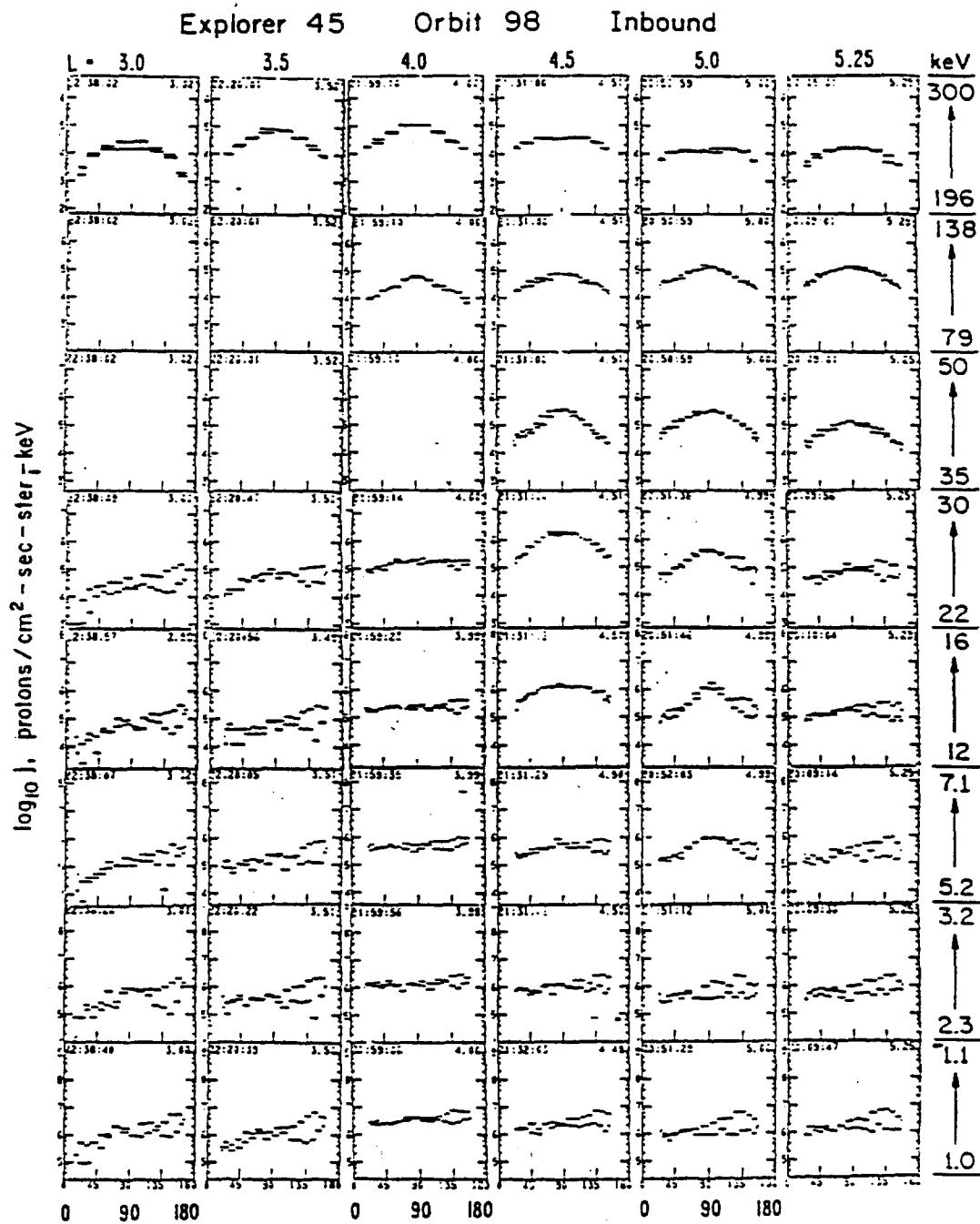


Fig. 21. Equatorial proton pitch angle distributions from Explorer 45 for every 0.5 L on inbound orbit 98. Elevated fluxes at pitch angles from 90° to 180° for the energies of < 30 keV are due to reflected sunlight. The data are for an interval following a storm on December 16, 1971. [23]

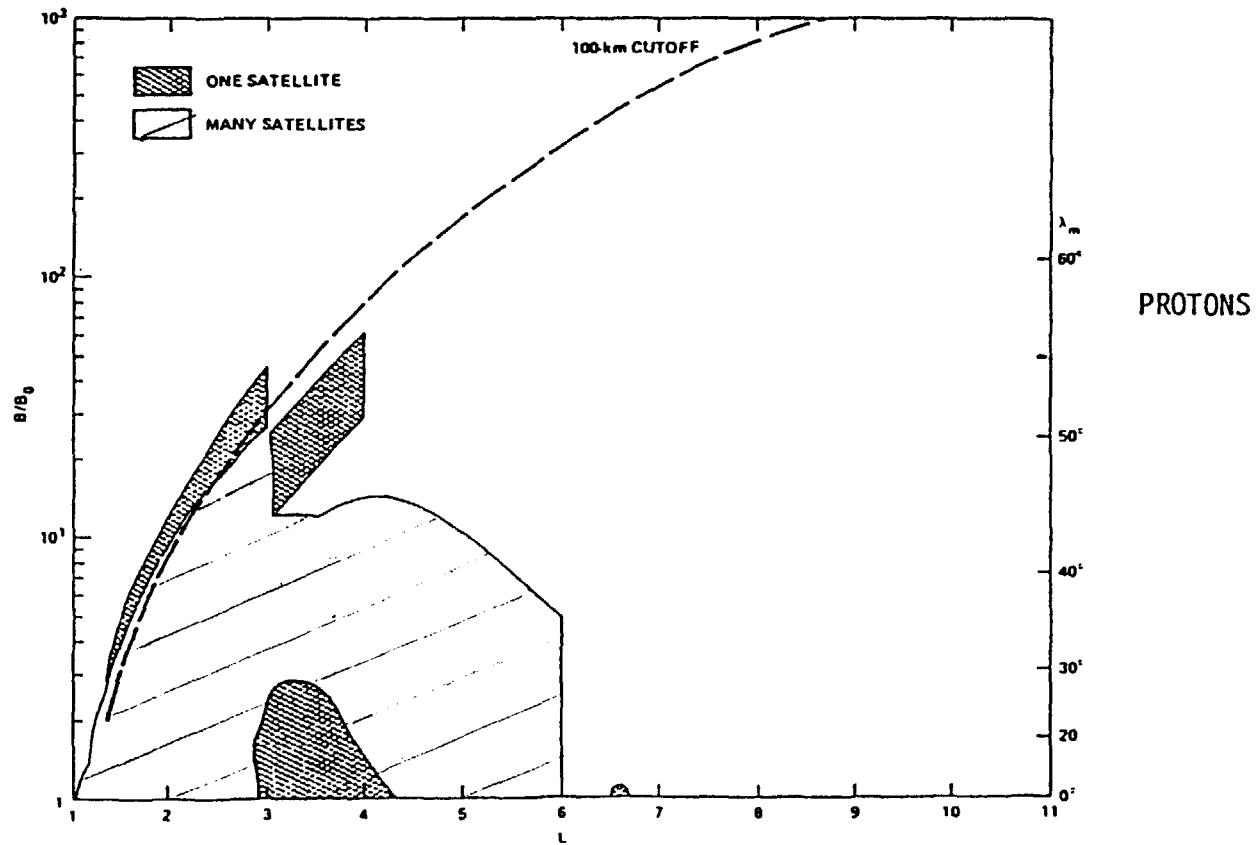


Fig. 22. B/B_0 - L coverage for experiments used in constructing the AP-8 model. The model and the data used have a cutoff at $L = 6.6$. [15]

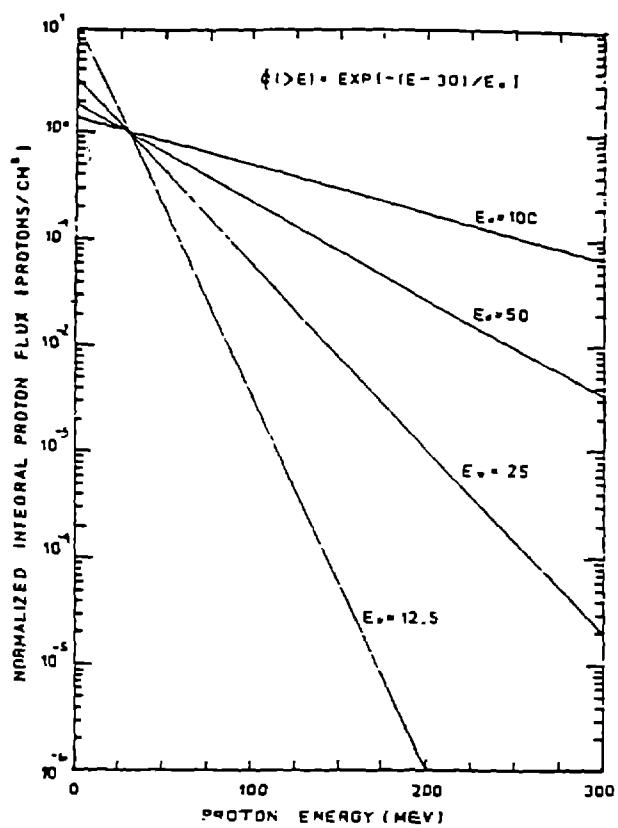


Fig. 23. Normalized integral proton flux as a function proton energy. The range of spectral shapes is what is to be expected given the current knowledge of natural variations in this population.[30]

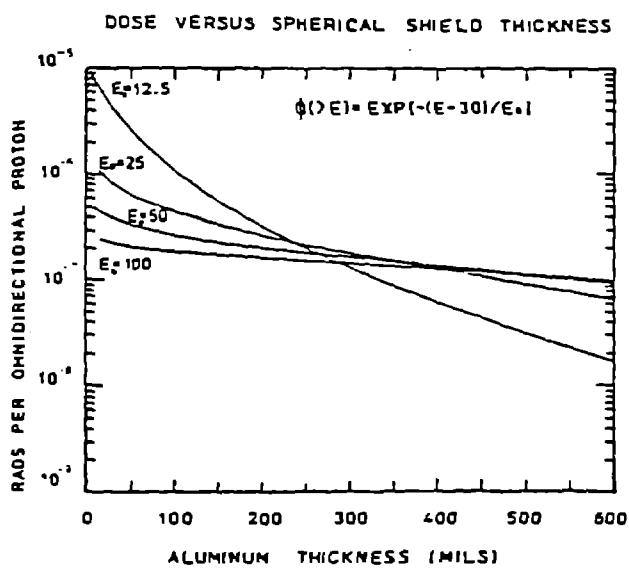


Fig. 24. Depth-dose per omnidirectional proton as a function of spherical aluminum thickness based on the spectra in Figure 23.[30]

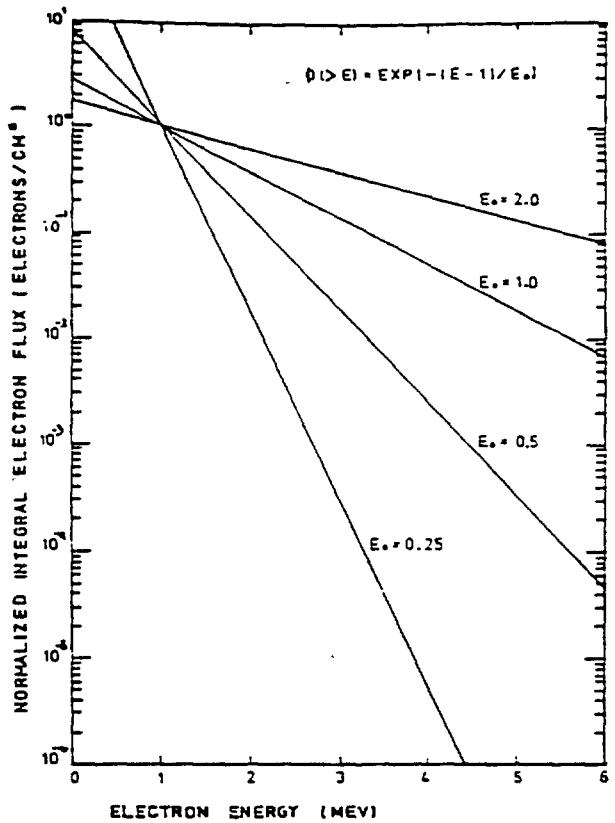


Fig. 25. Normalized integral electron flux as a function electron energy. The range of spectral shapes is what is to be expected given the current knowledge of natural variations in this population.[30]

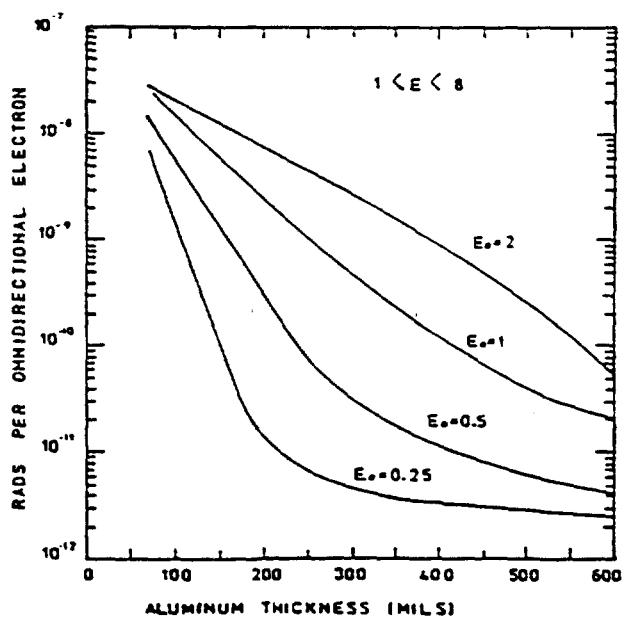
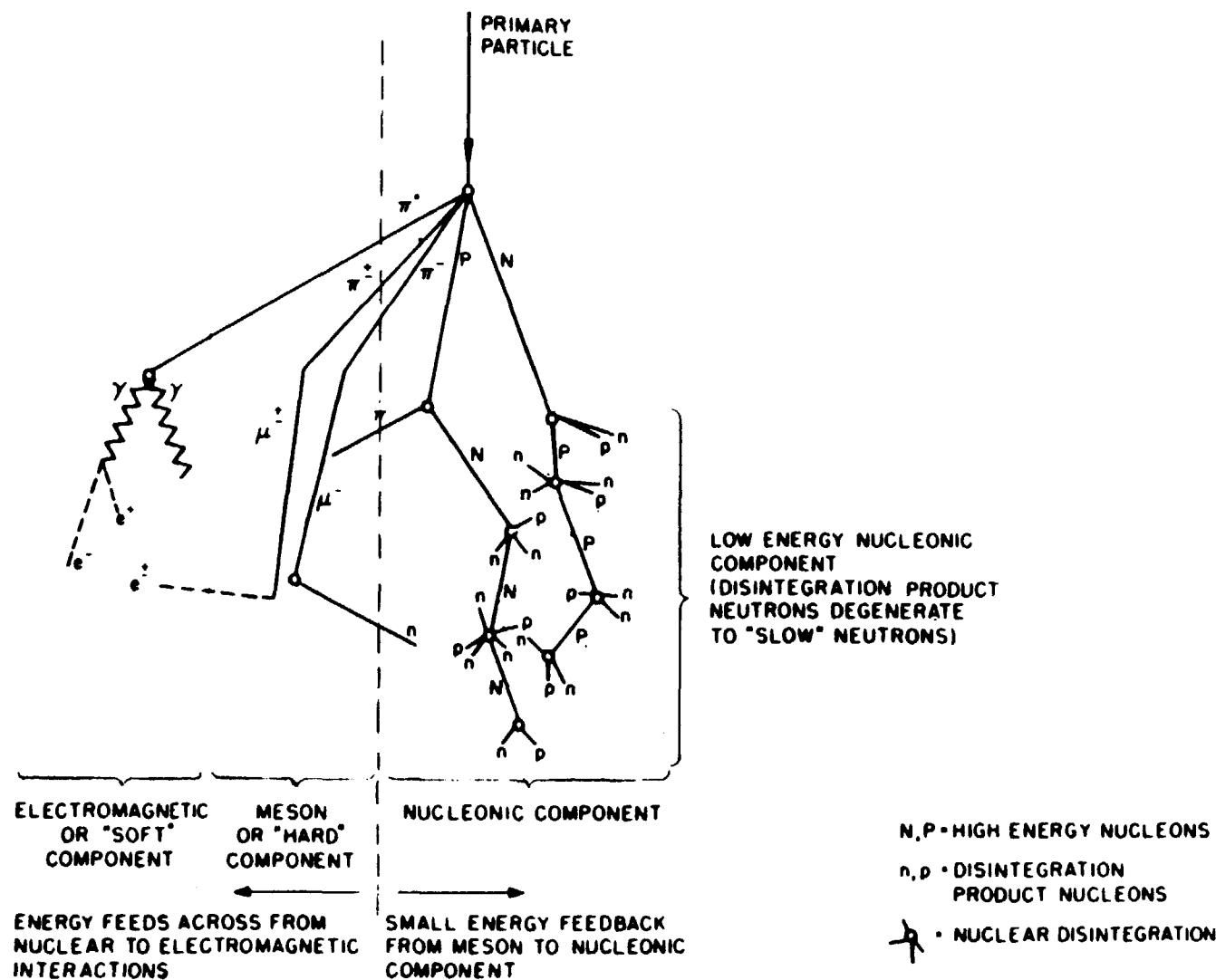
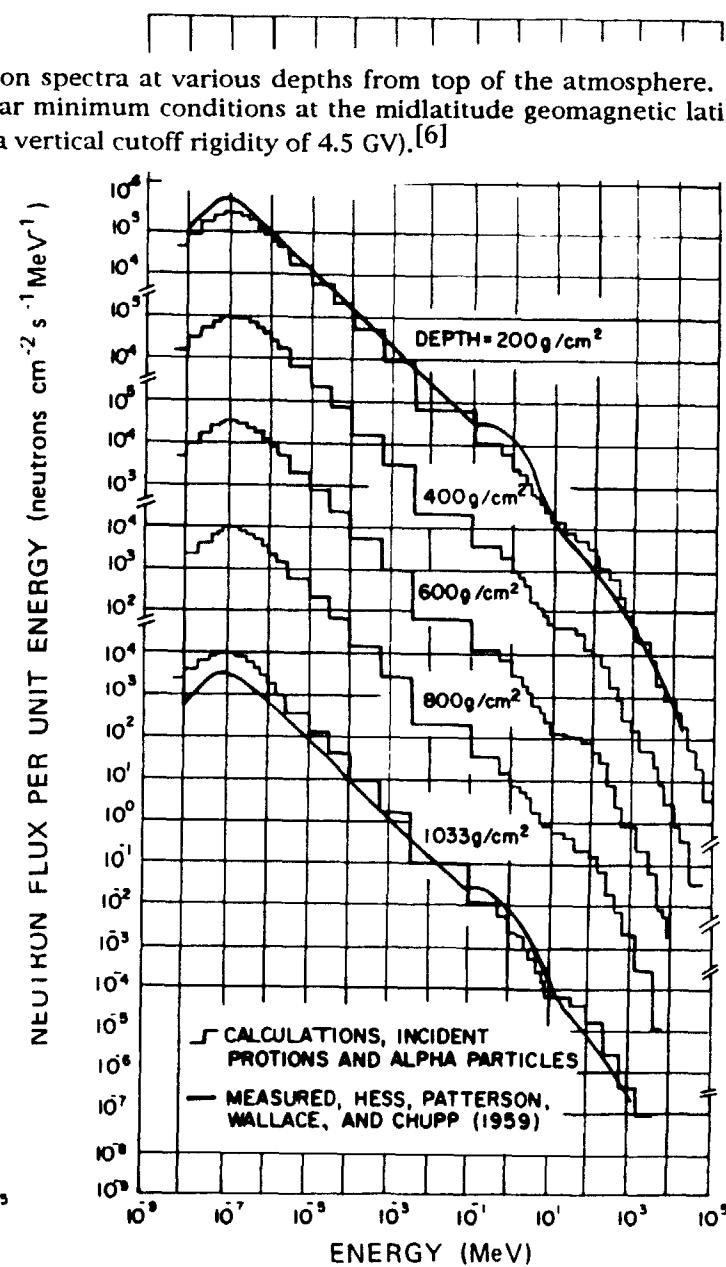
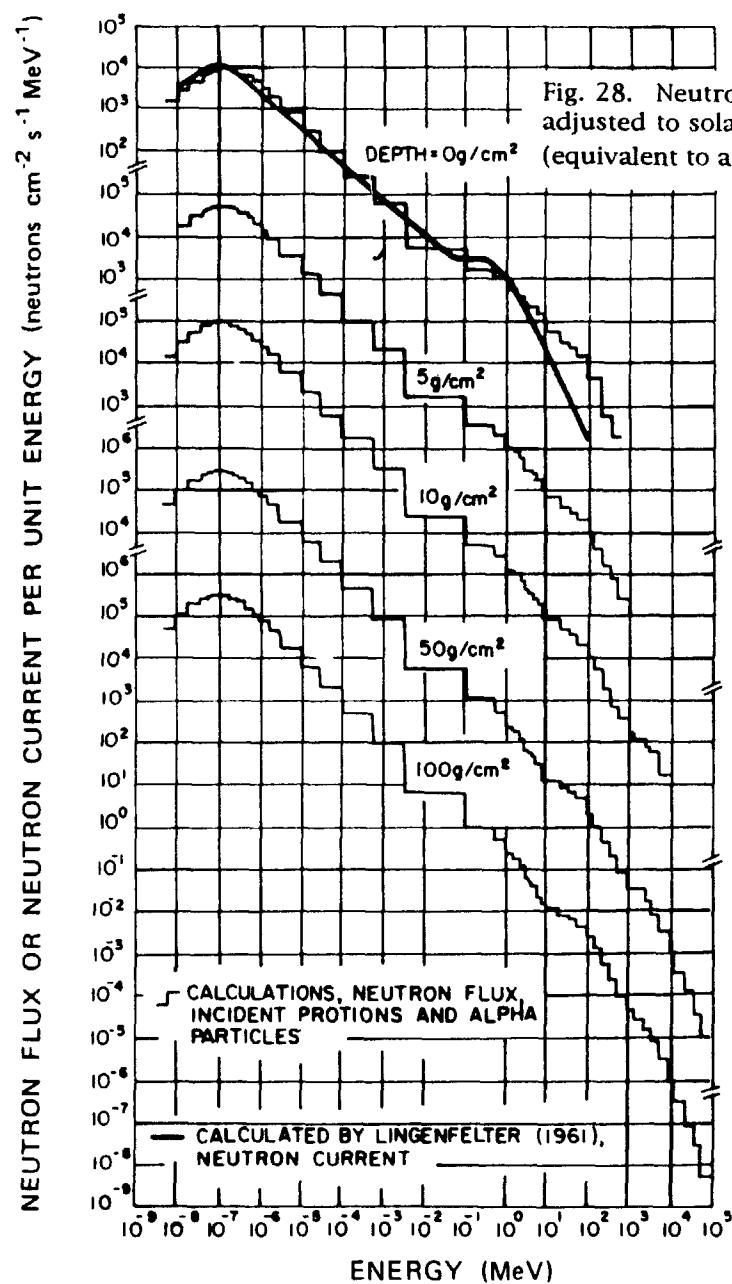


Fig. 26. Depth-dose per omnidirectional electron as a function of spherical aluminum thickness based on the spectra in Figure 25.[30]

Fig. 27. Schematic drawing of a cosmic ray shower. N and P are high energy neutrons and protons while the lower case n and p are used to denote disintegration product neutrons and protons. Pions, mesons, electrons, and gamma rays are indicated by conventional symbols. [6]





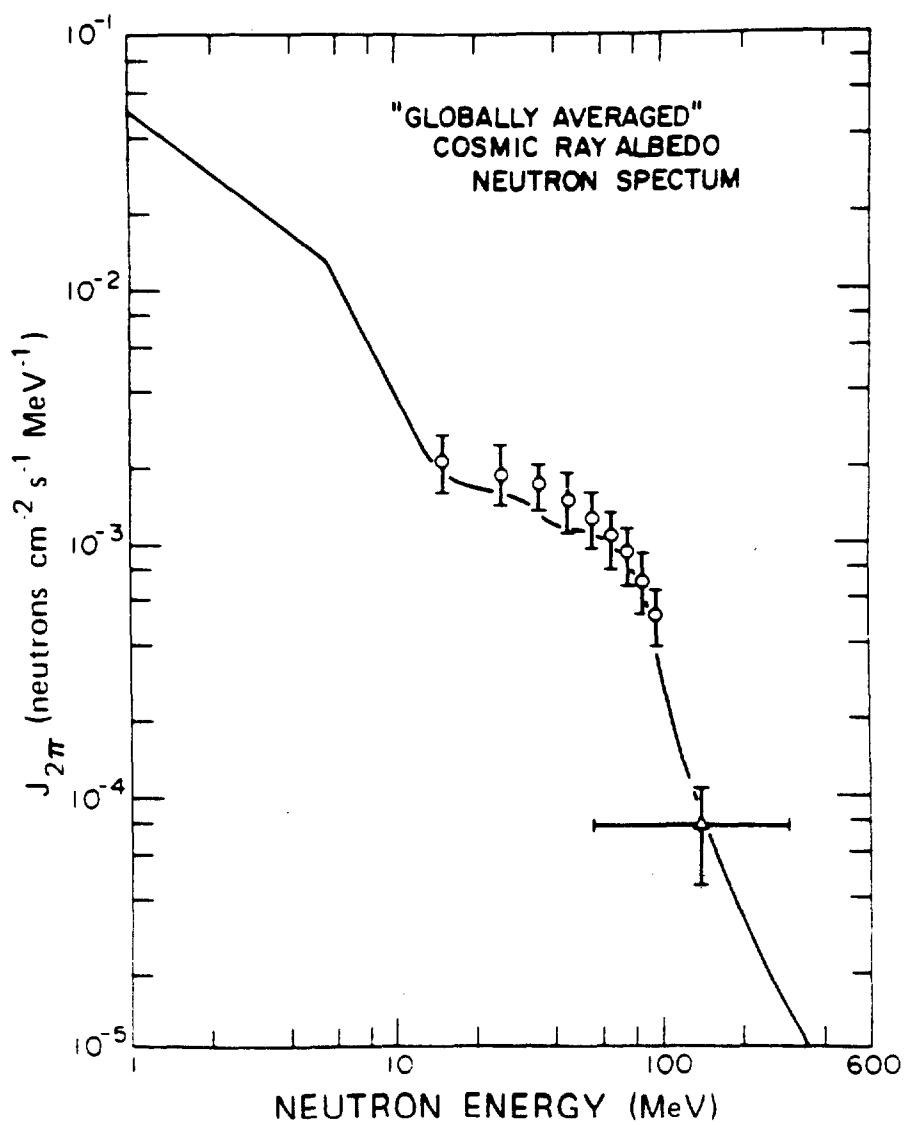
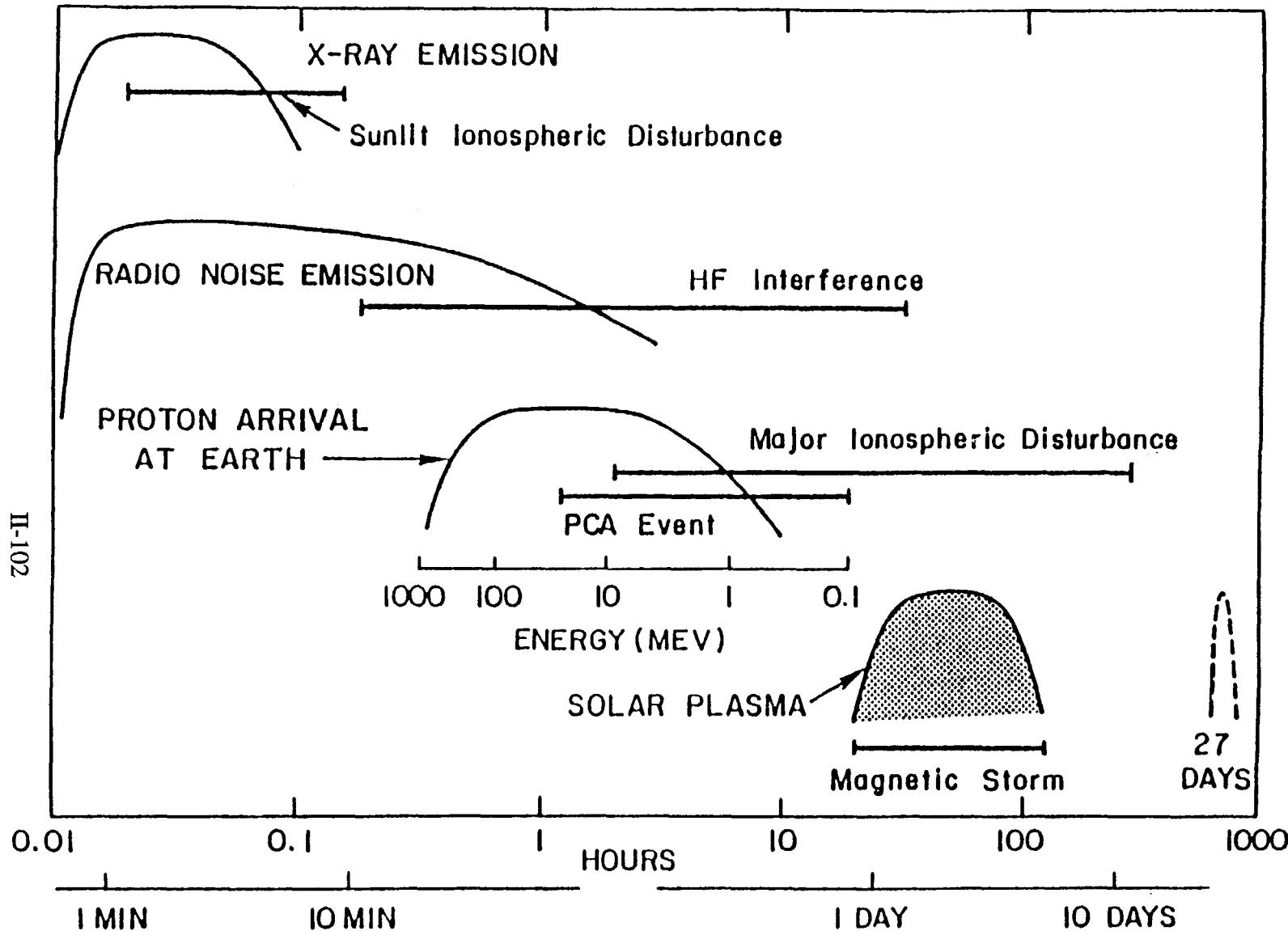


Fig. 29. A model "global averaged" albedo neutron spectrum. [90]

II-102



TIME SCALE FOR SOLAR FLARE EFFECTS

Fig. 30. Schematic plot of the relative variations in time of the amplitudes of the X-ray, radio noise, high energy particle, and solar plasma fluxes for a "typical" large solar flare.

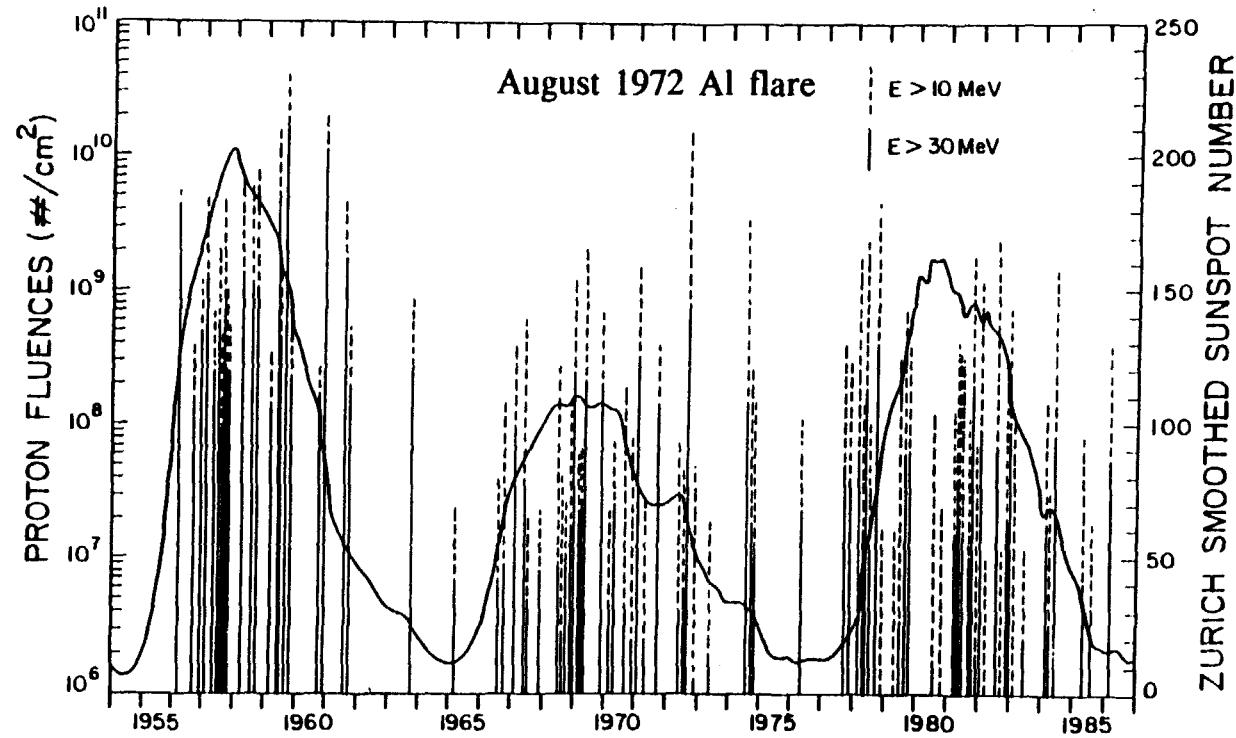


Fig. 31. Solar-flare proton fluences (vertical lines) superimposed on the sunspot cycle (continuous curve). Solid lines correspond to $E > 3$ MeV, dashed lines to $E > 1$ MeV proton fluences.

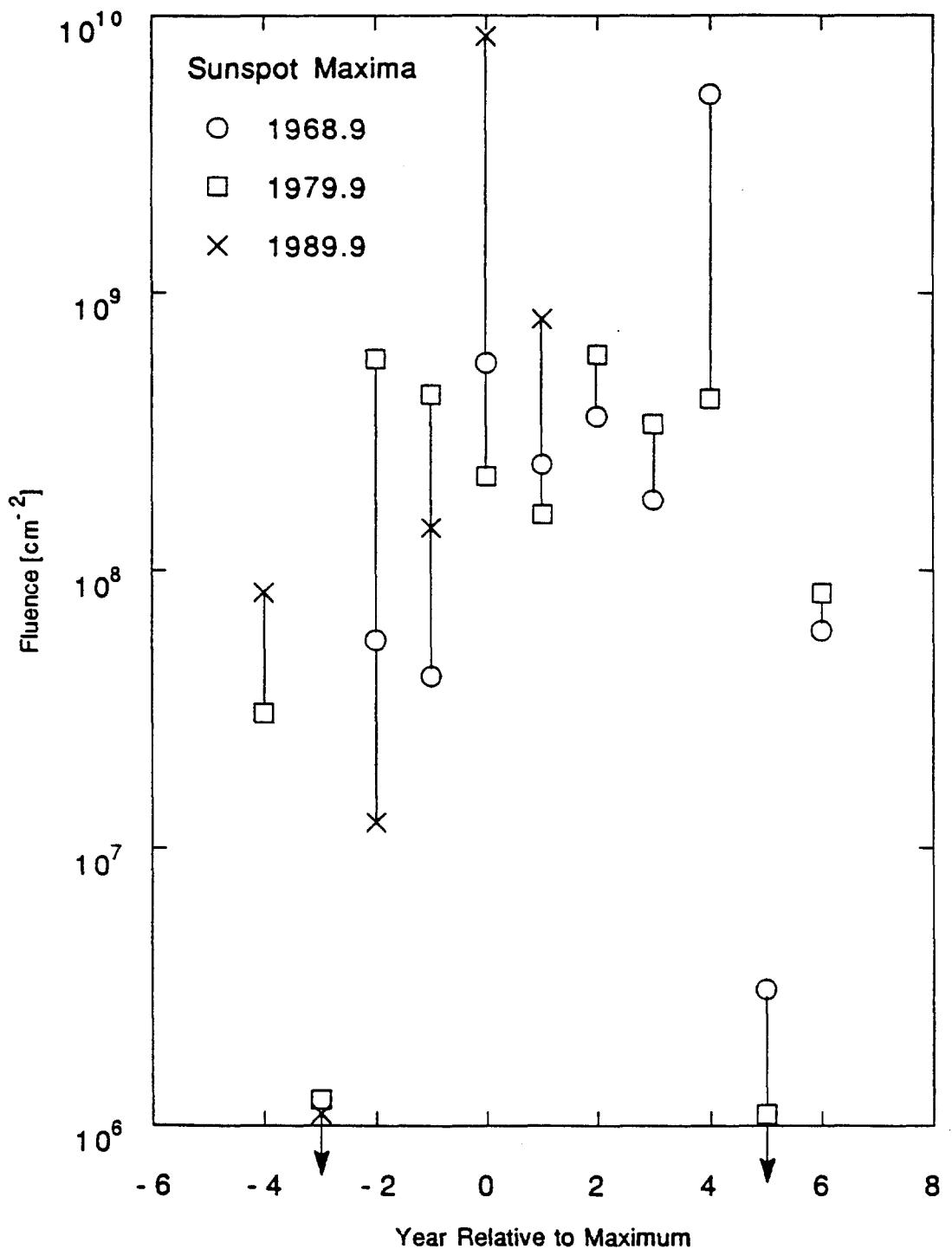


Fig. 32. Yearly Event Fluences for Protons of Energy >30 MeV vs Year Relative to Sunspot Maximum.[31]

AUGUST 1972, OCTOBER 1989 SOLAR PROTON EVENTS

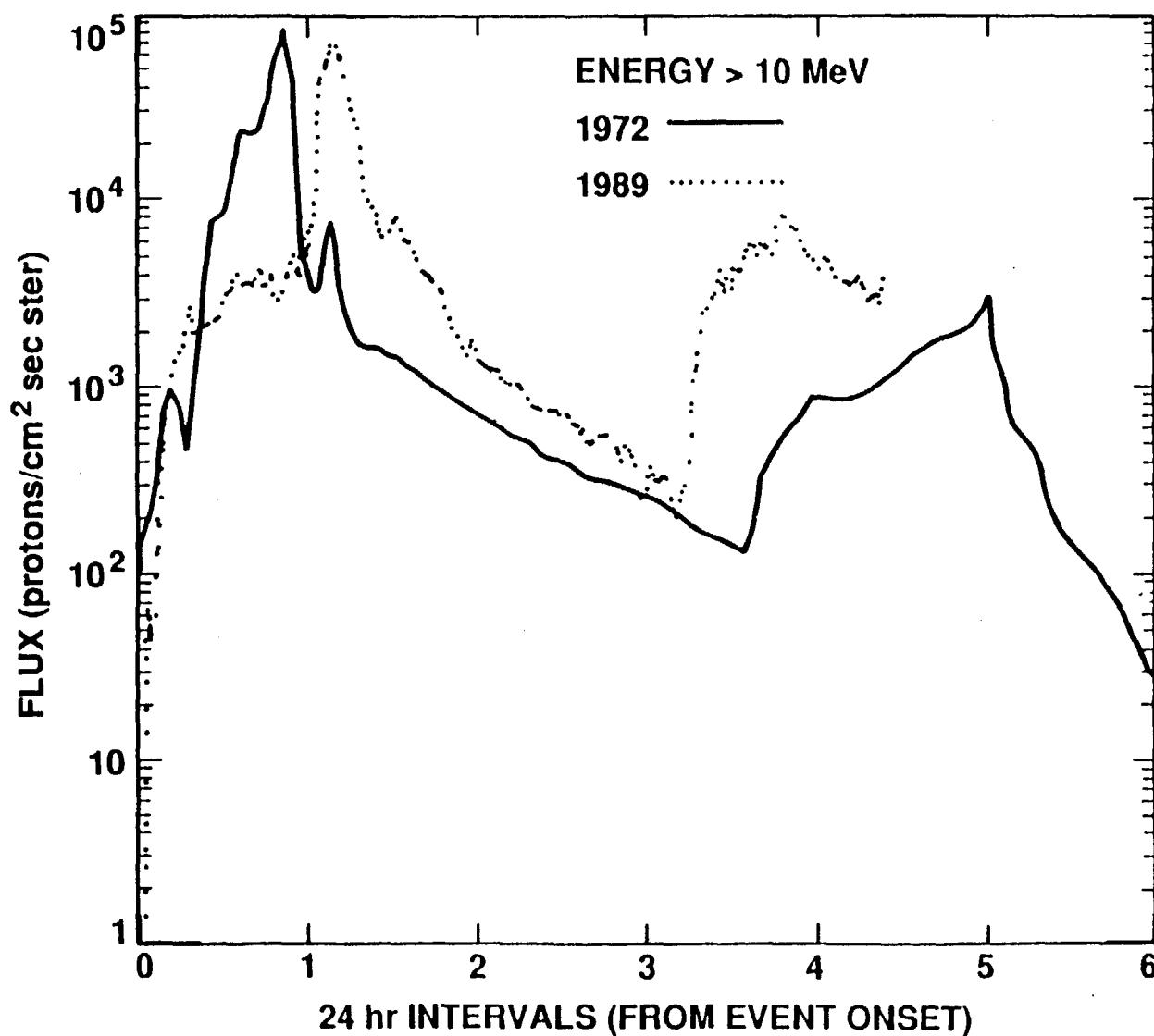


Fig. 33. Comparison of the $E > 10$ MeV proton flux as a function of time for the great flares of August 1972 and the October 1989.

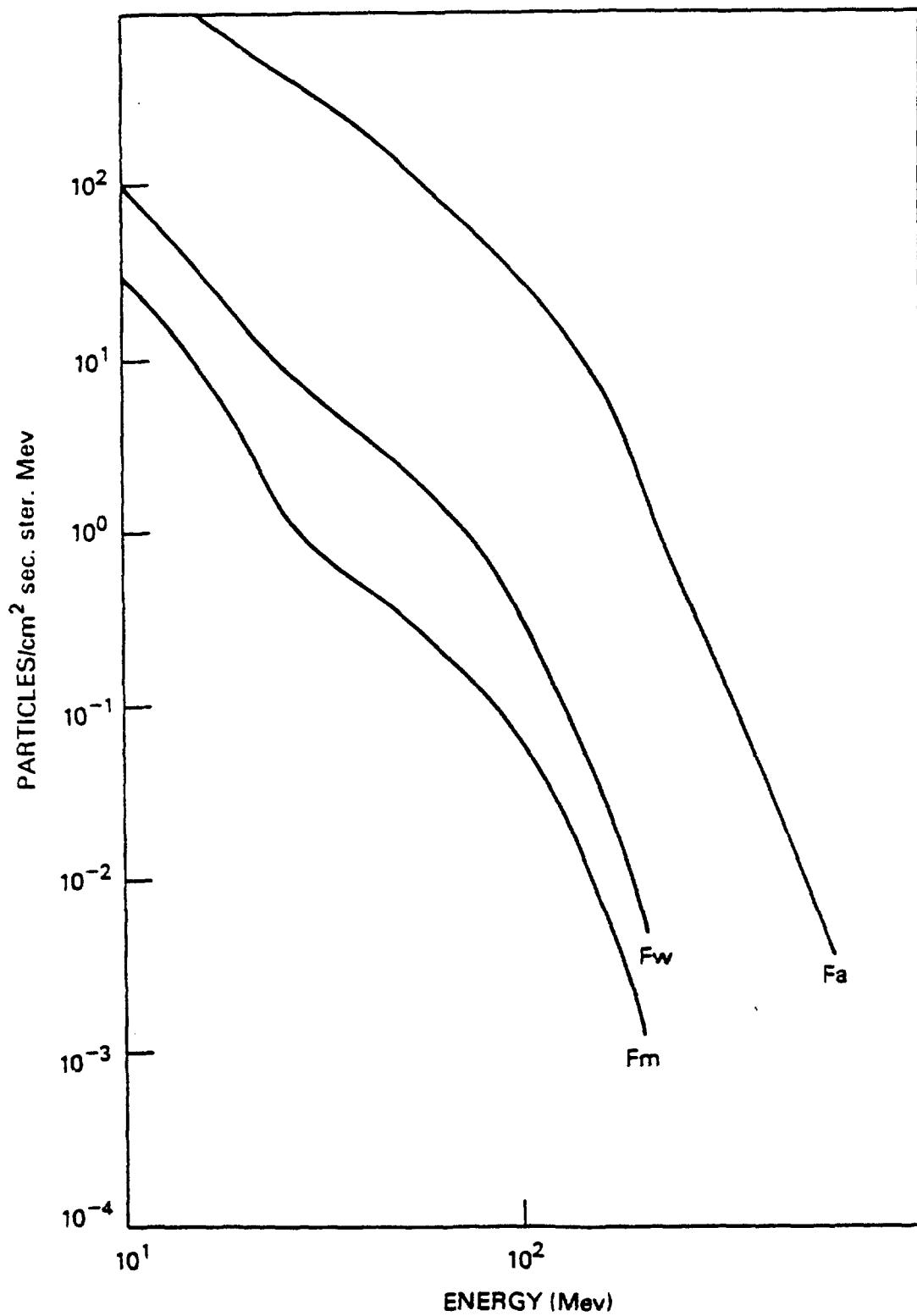


Fig. 34. The peak proton flux differential energy spectra for: F_m , a "typical" event; F_w , a "worst-case" event (90% confidence level); and F_a , an "anomalously large" event. [28]

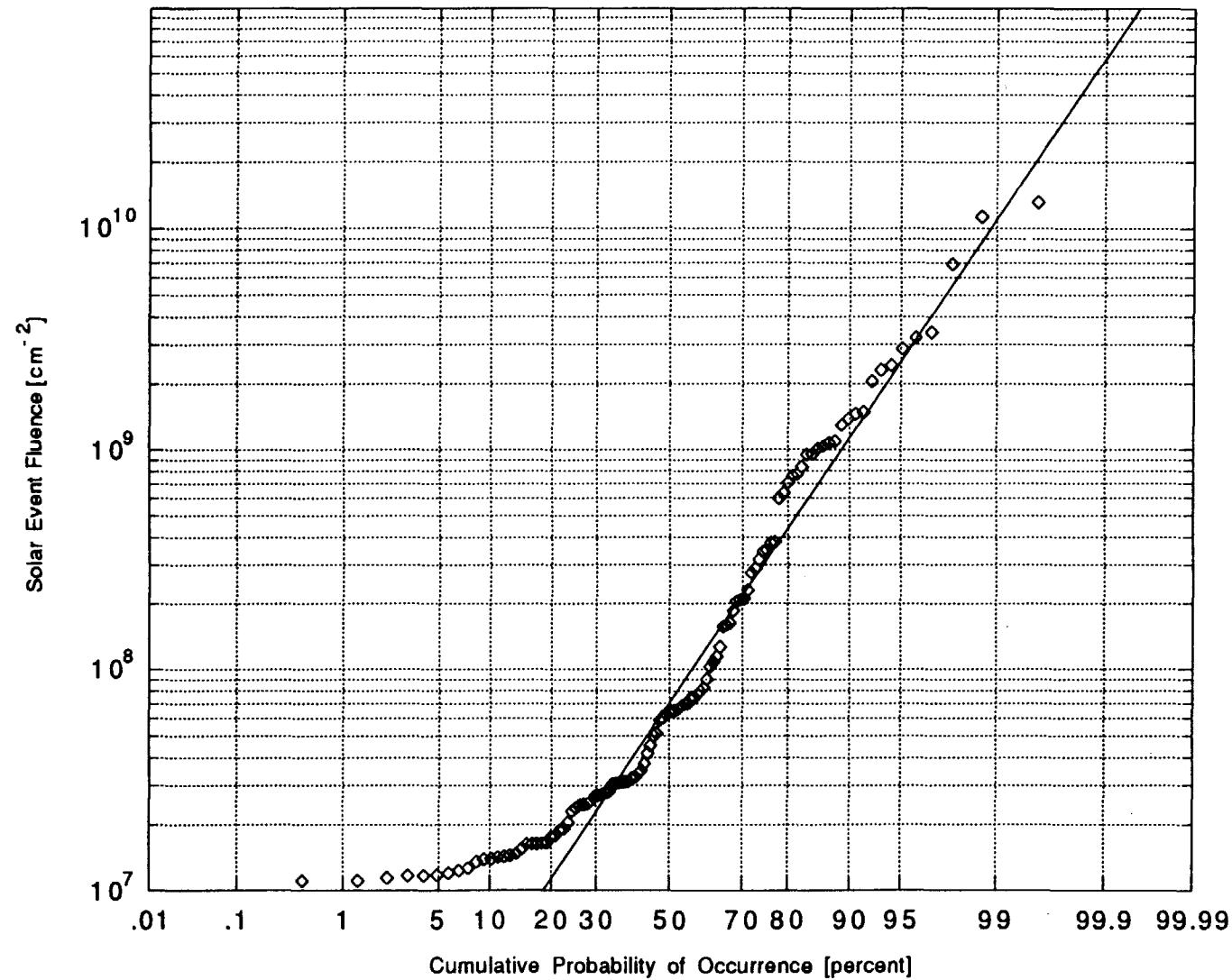


Fig. 35. Distribution of Solar Event Fluences for Solar Active Years between 1963 and 1991 for Protons of Energy > 10 MeV for which Daily Averaged Flux exceeds 1.0 cm⁻²·s⁻¹·sr⁻¹. The straight line is the selected log normal distribution.[31]

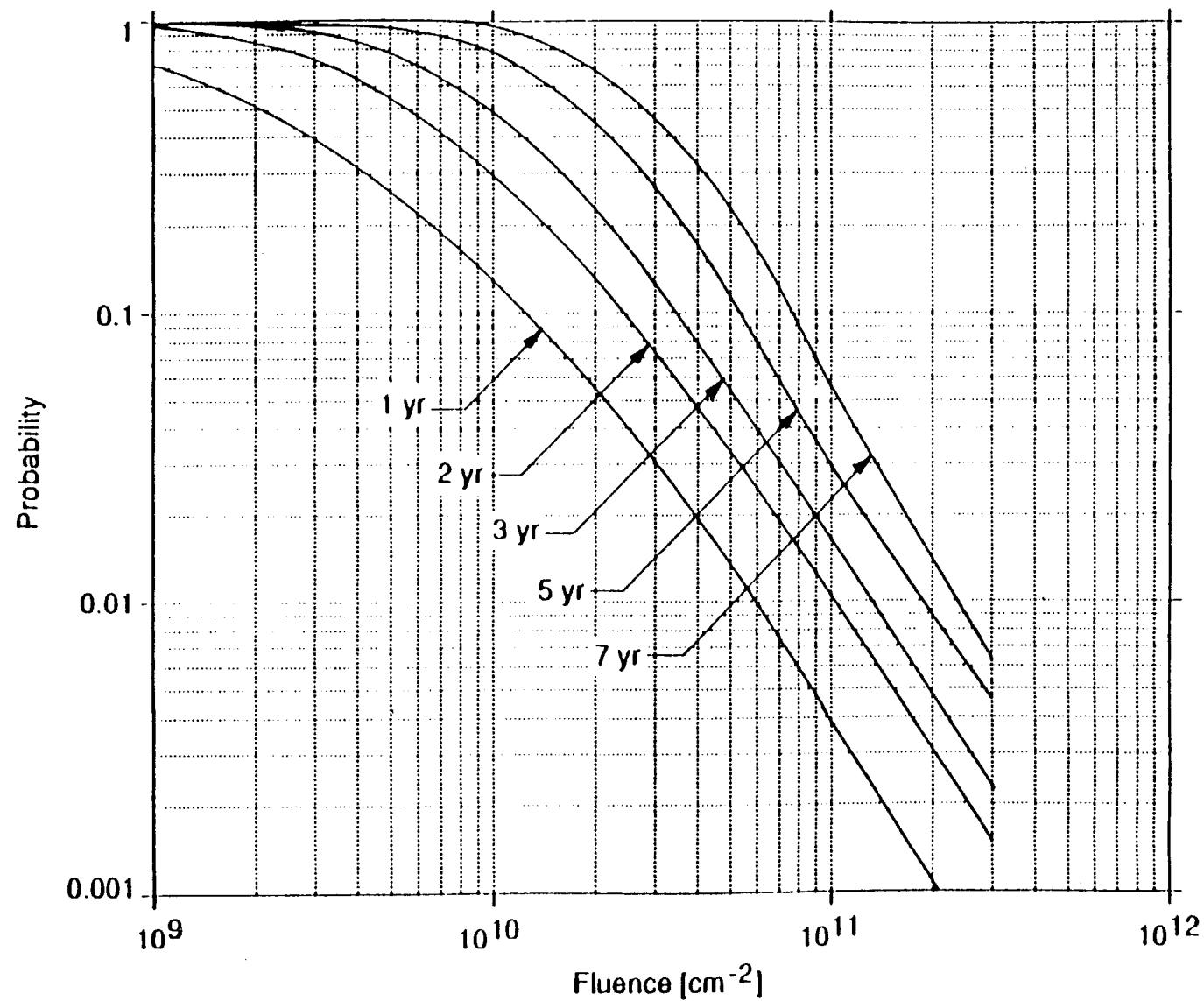


Fig. 36. Fluence Probability Curves for Protons of Energy Greater than 10 MeV for Various Mission Lengths.[31]

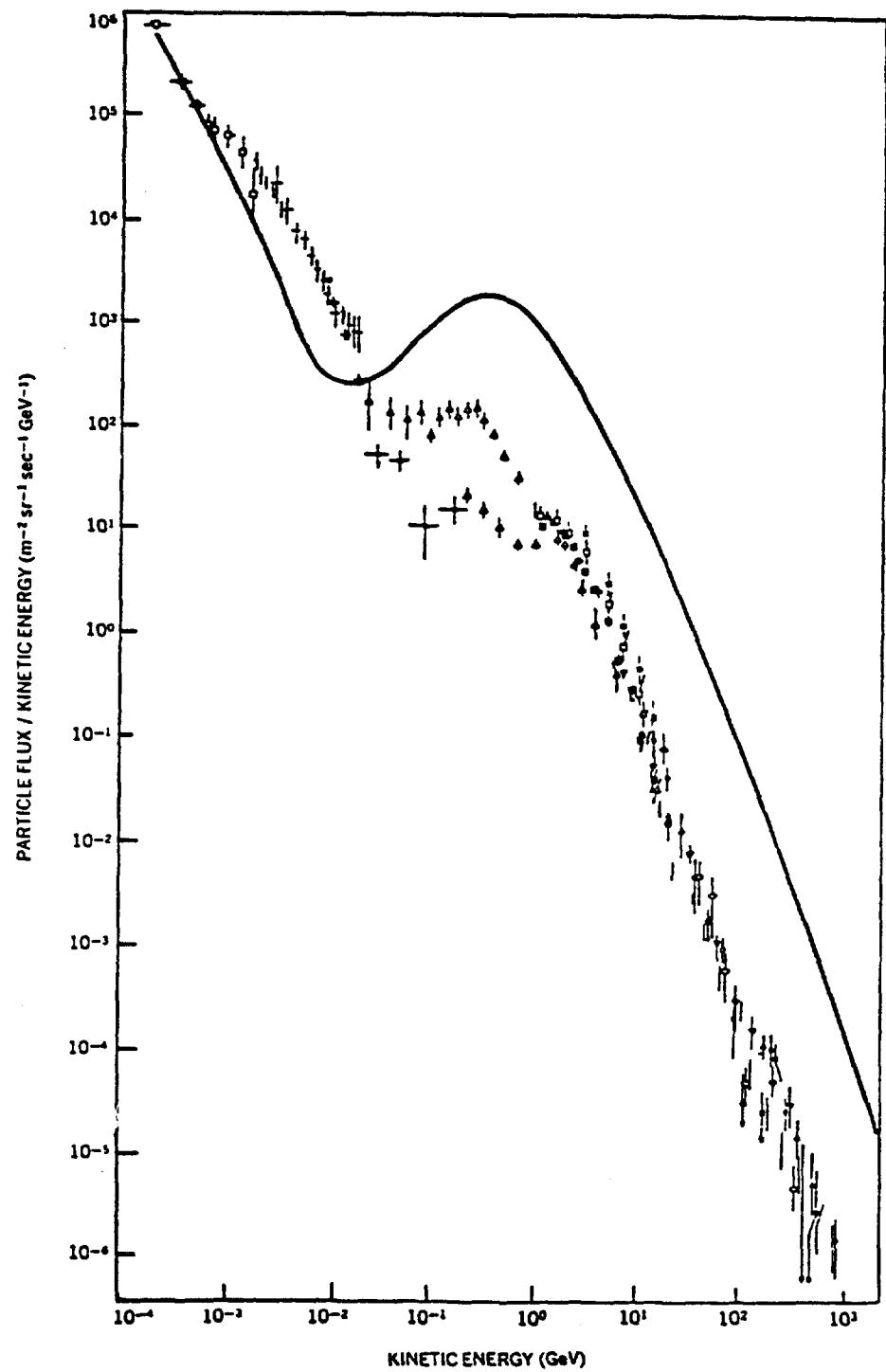


Fig. 37. The energy spectra of cosmic-ray protons (line) and electrons (points) as measured near Earth.^[39] Below a few GeV, interstellar spectra are strongly influenced by the Sun.

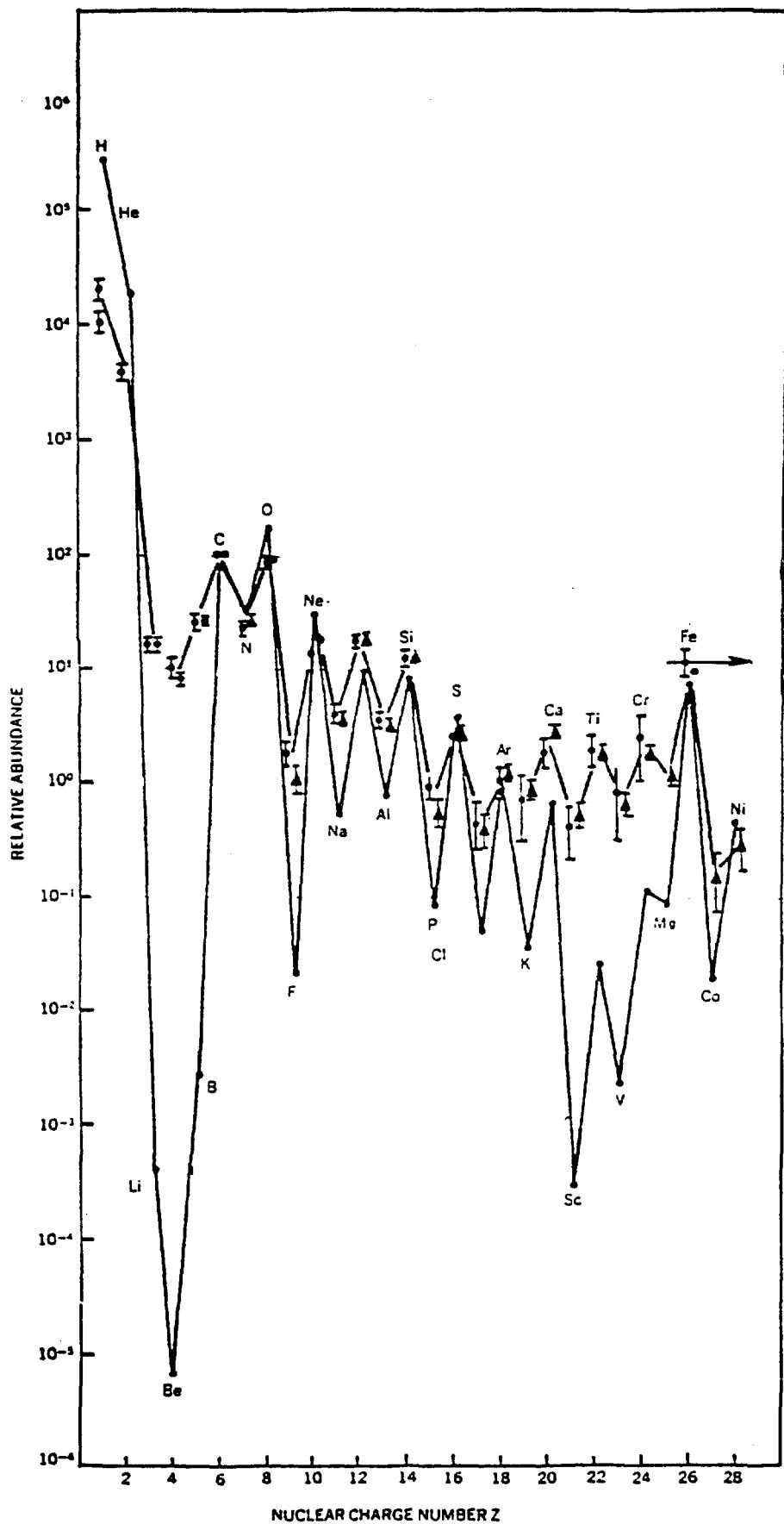


Fig. 38. Relative abundance of the elements from hydrogen to the iron group, normalized to that of carbon ($C=100$)^[39]. The barred points represent the cosmic-ray abundances measured near Earth. The dots (without error bars) represent the elemental abundances in the solar system.

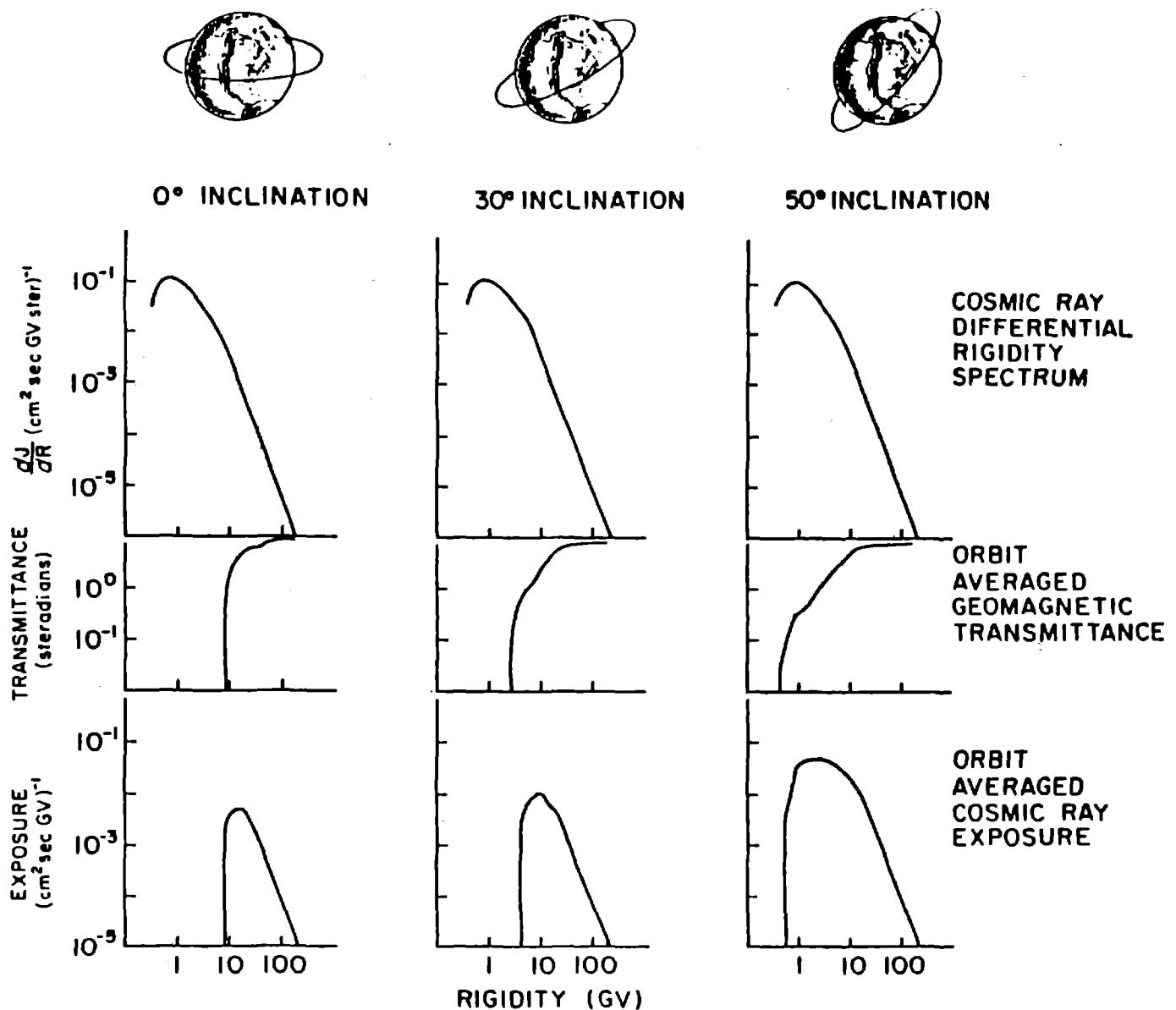


Fig. 39. Schematic representations^[40] of the effects of the Earth's magnetic field on the Cosmic Ray environment as functions of particle rigidity and orbital inclination. The top 3 spectra represent the Cosmic Ray spectrum outside the Earth's field. The second set of figures represents an orbit average geomagnetic transmittance factor which is a function of rigidity and orbital inclination. The third plot is the result of convolving the 2 upper sets of curves and is the orbit average Cosmic Ray exposure as a function of rigidity for 0°, 30°, and 50° inclination orbits.

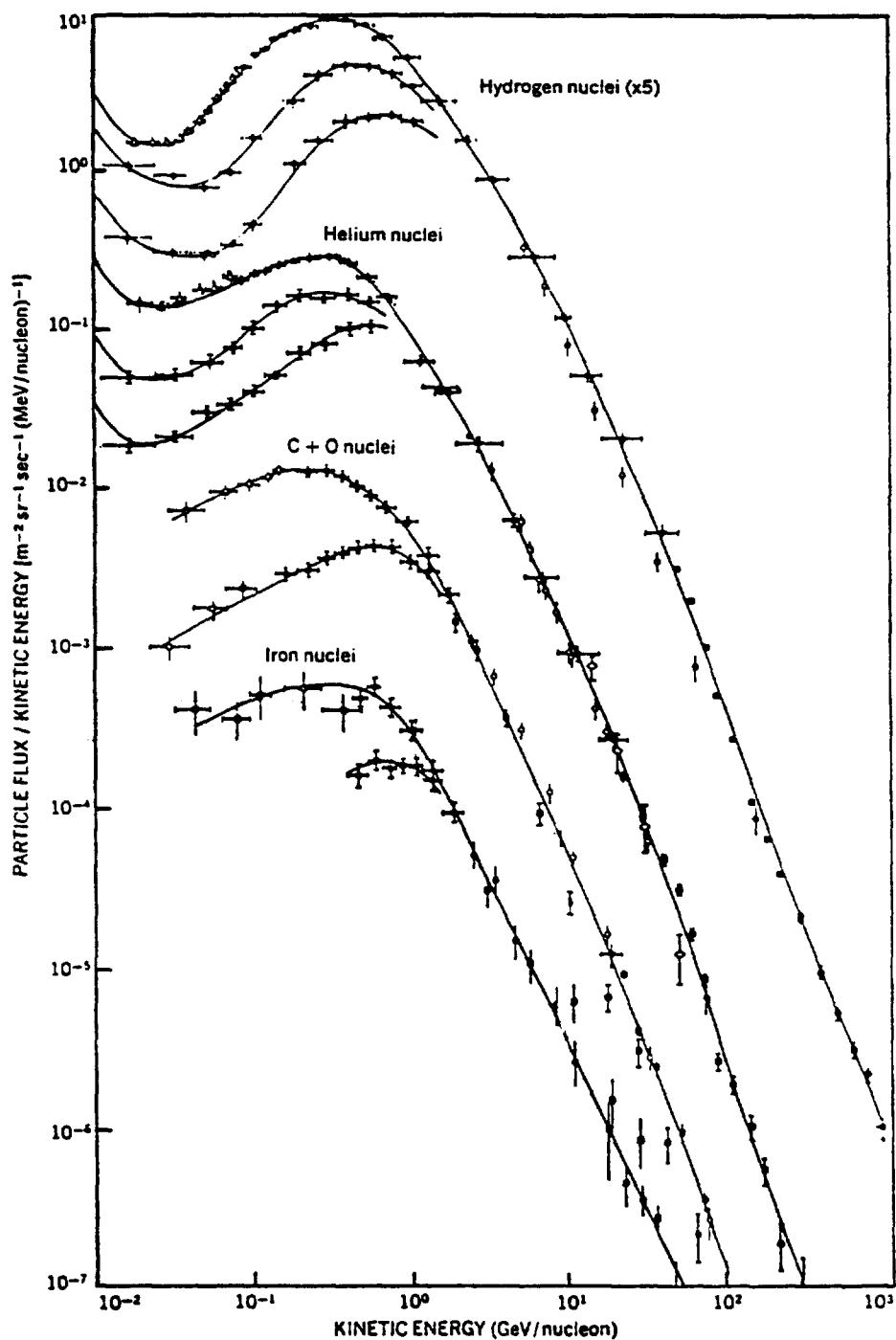


Fig. 40. Cosmic-ray energy spectra^[39] of the more abundant nuclear species as measured near Earth. Below a few GeV/nucleon, these spectra are strongly influenced by modulation within the solar system. The different curves for the same species represent measurements at various levels of general solar activity, the lowest intensity being observed at the highest activity level.

MAGNETOSPHERES

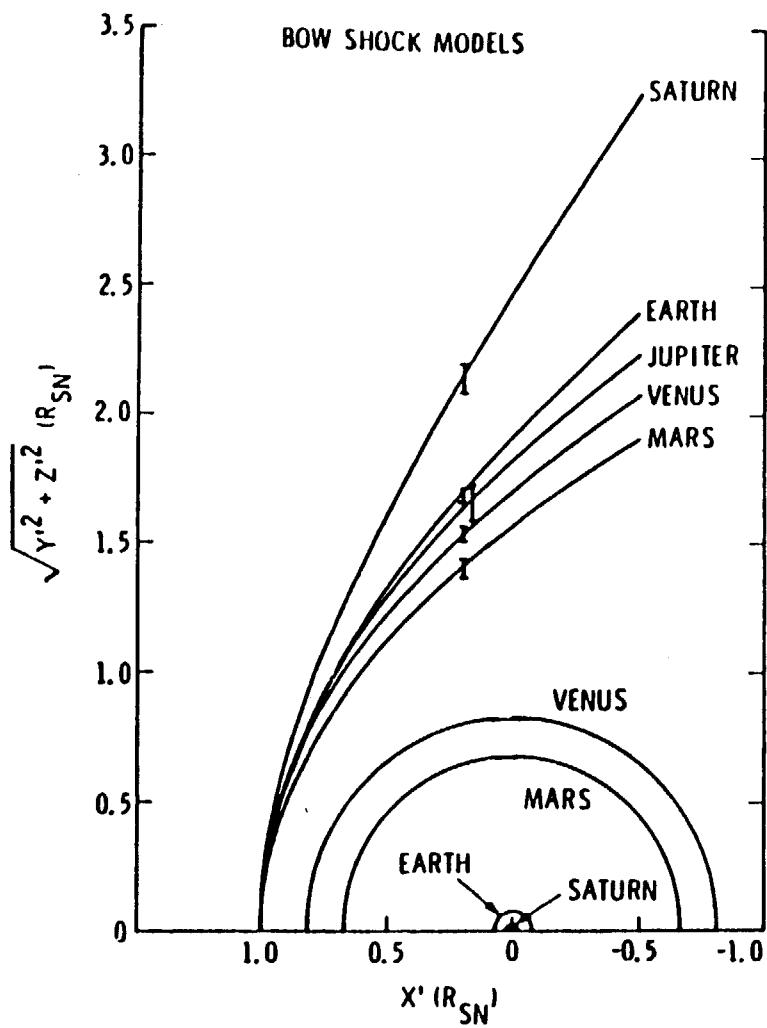


Fig. 41. Comparison of planetary-magnetospheres. [44, 91]

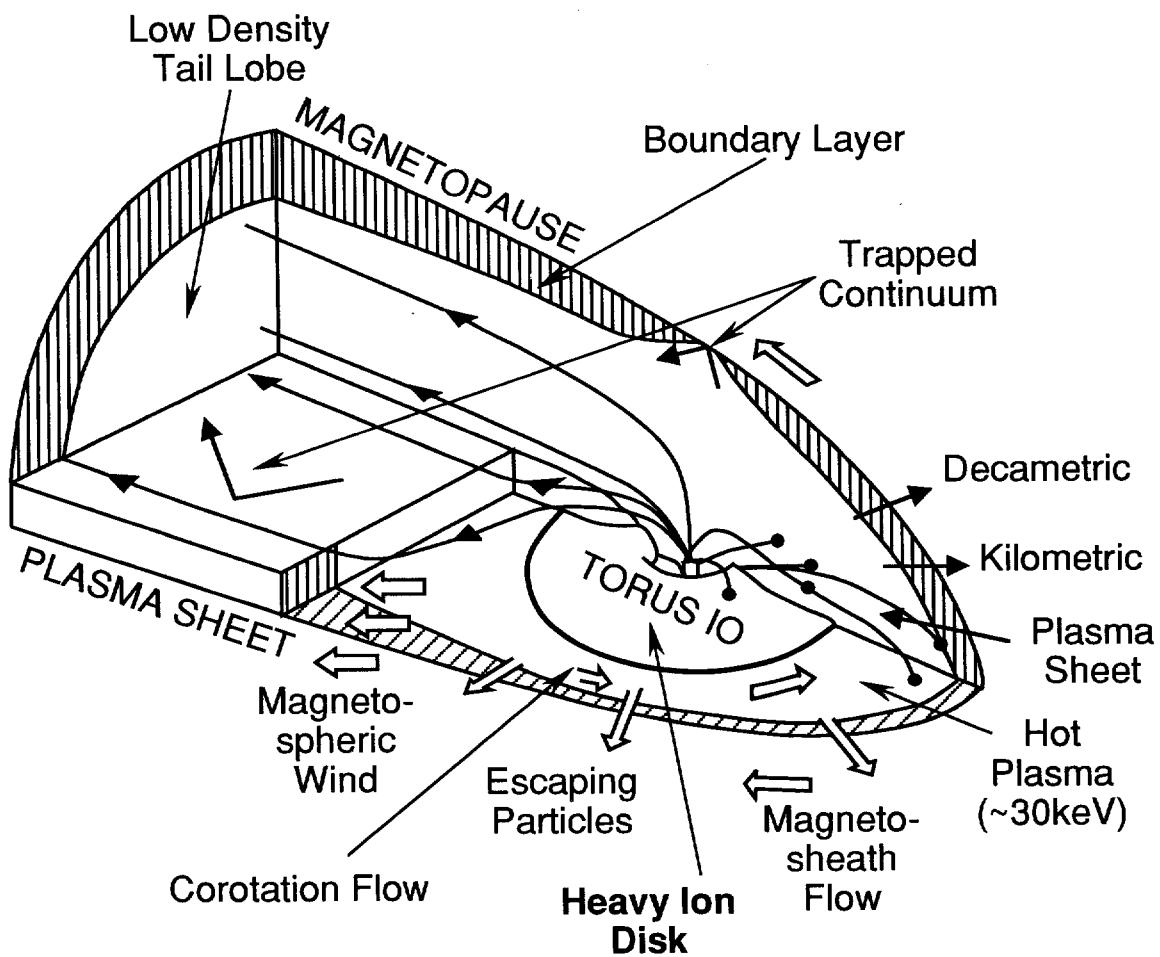


Fig. 42. Schematic representation of Jupiter's magnetosphere illustrating the various plasma regions and particle flows. [54]

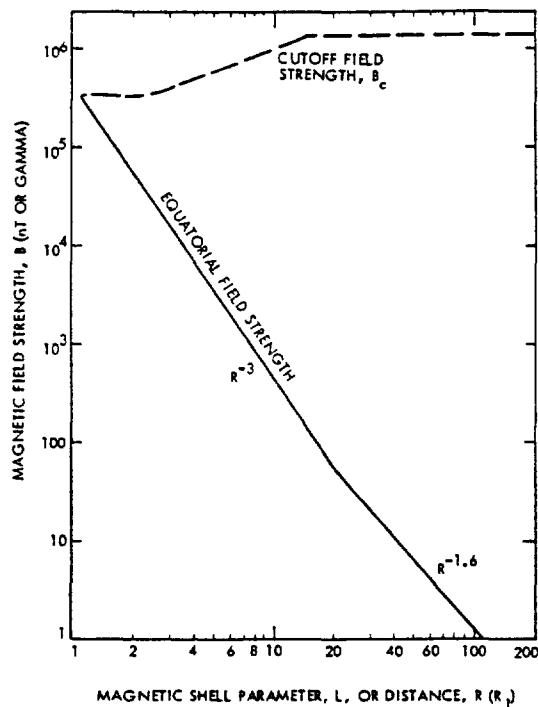


Fig. 43. Equatorial and cutoff magnetic field strengths for the Jupiter magnetic field and charged particle models.[45]

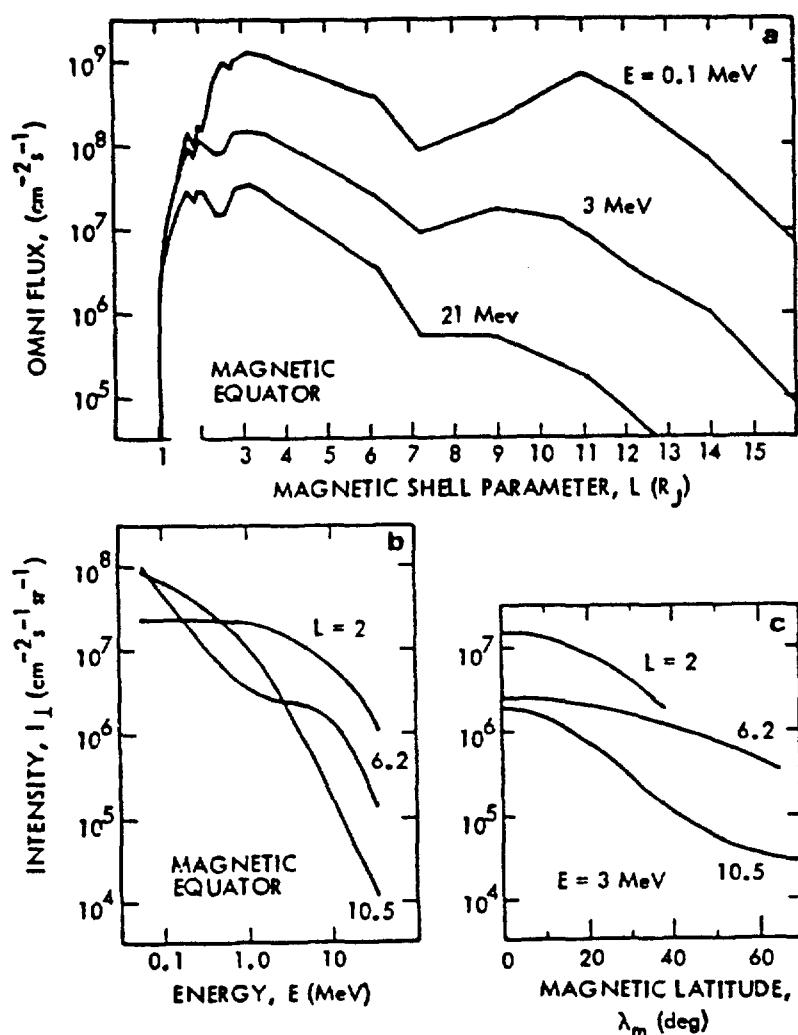


Fig. 44. Sample dependencies^[45] for Jupiter energetic electron model. (a) Omnidirectional integral flux at three energies as a function of L shell. (b) Integral intensity at 90° pitch angle at the equator and for three L shell values as a function of energy. (c) Integral ($>3 \text{ MeV}$) intensity at 90° pitch angle for different L shells as a function magnetic latitude.

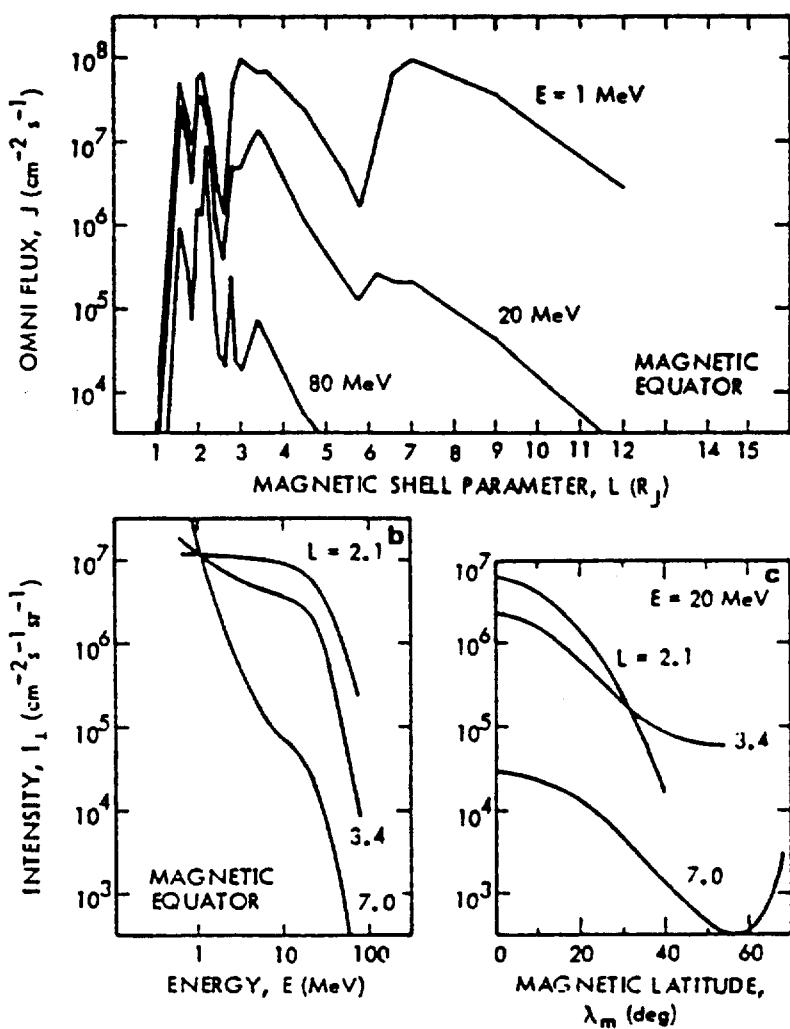


Fig. 45. Sample dependencies^[45] for Jupiter energetic proton model. (a) Omnidirectional integral flux at three energies as a function of L shell. (b) Integral intensity at 90° pitch angle at the equator and for three L shell values as a function of energy. (c) Integral ($E > 20 \text{ MeV}$) intensity at 90° pitch angle for different L shells as a function magnetic latitude. The increase of the intensity at $L = 7.0$ and $\lambda_m > 55^\circ$ illustrates the kind of model defect that can occur when carefully crafted algebraic forms are used to extrapolate beyond the range of available data, and this increase is unphysical.

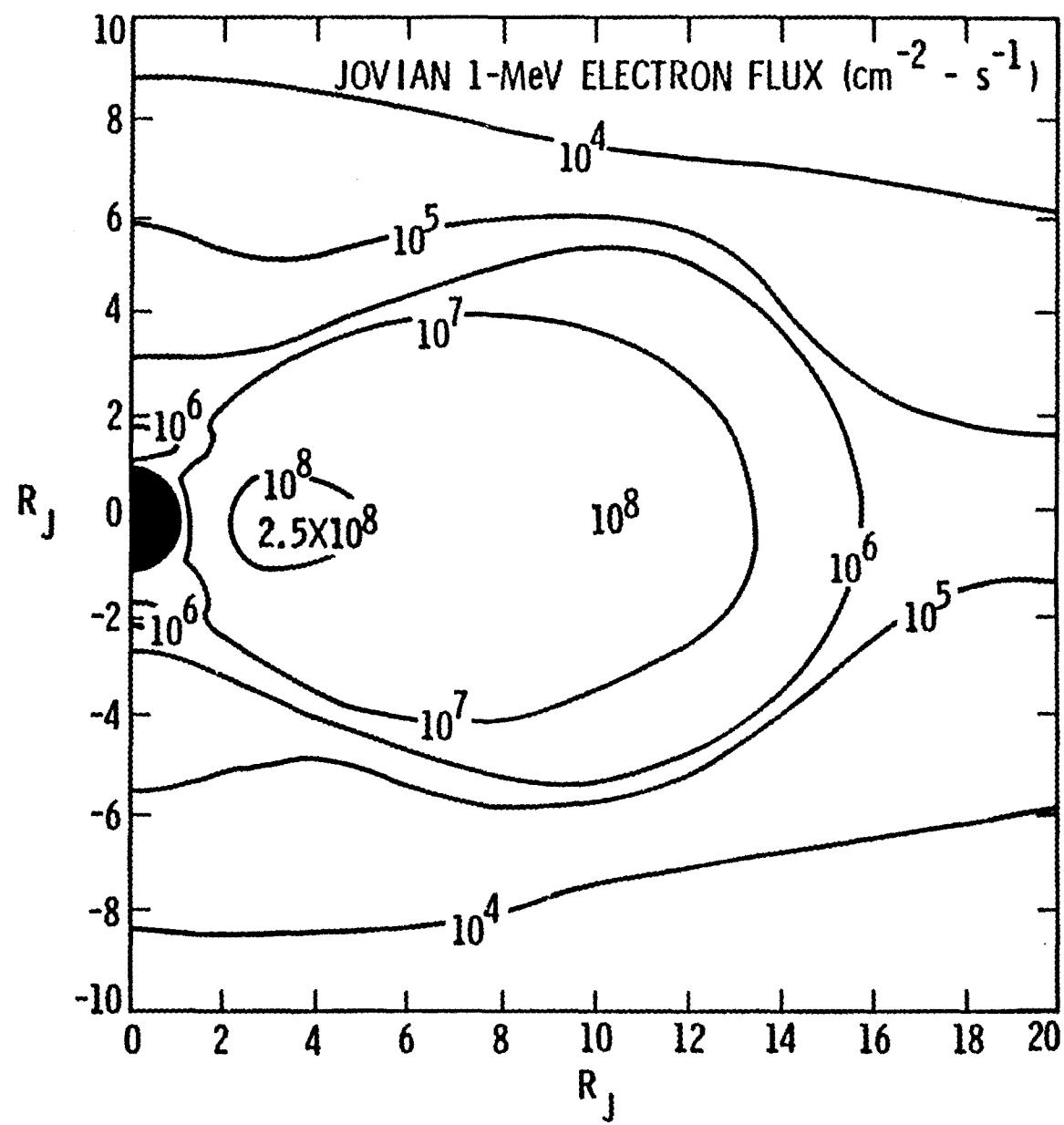


Fig. 46. Contours for electron fluxes above 1 MeV at Jupiter.^[45]

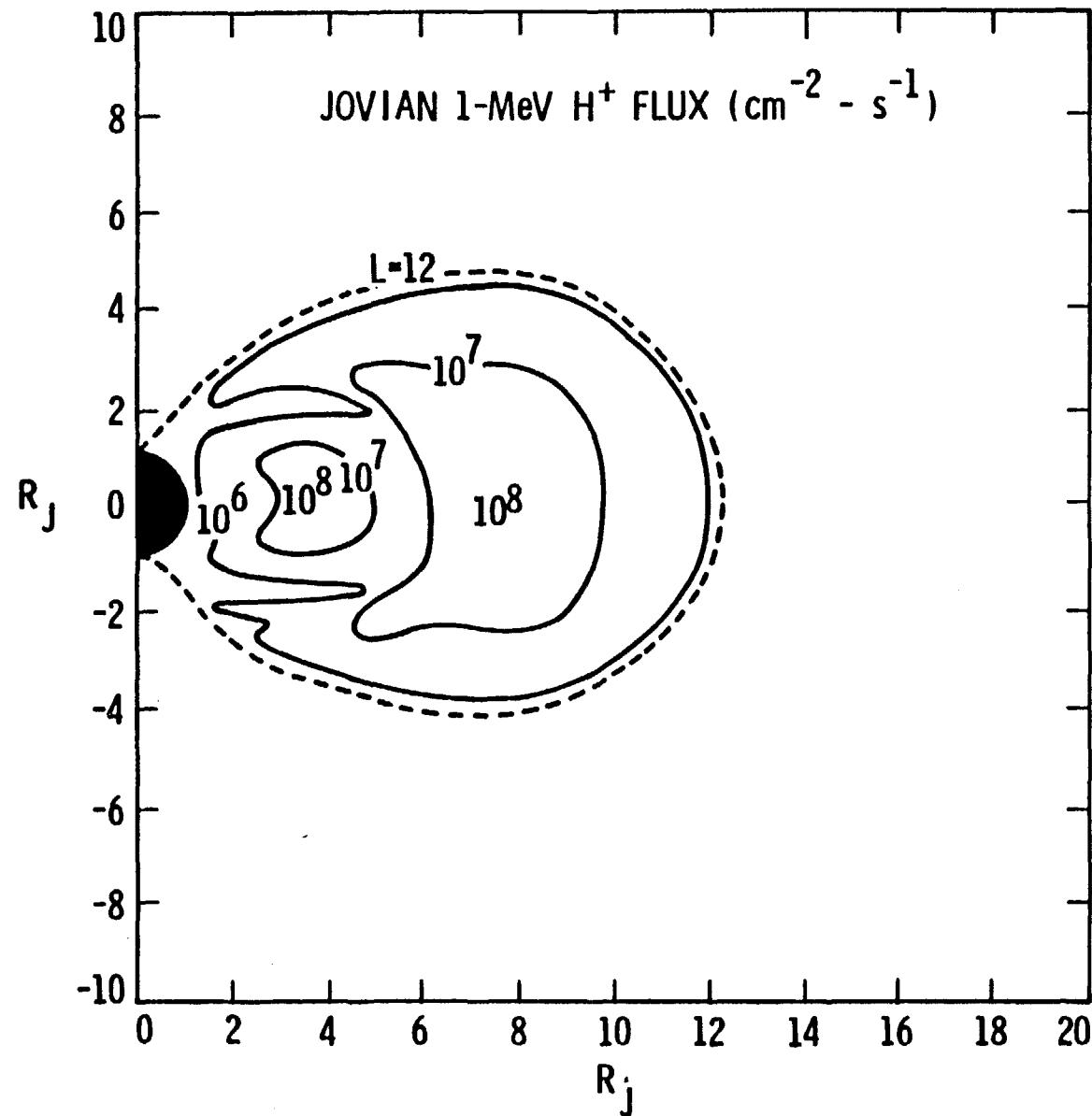


Fig. 47. Contours for proton fluxes above 1 MeV at Jupiter.^[45]

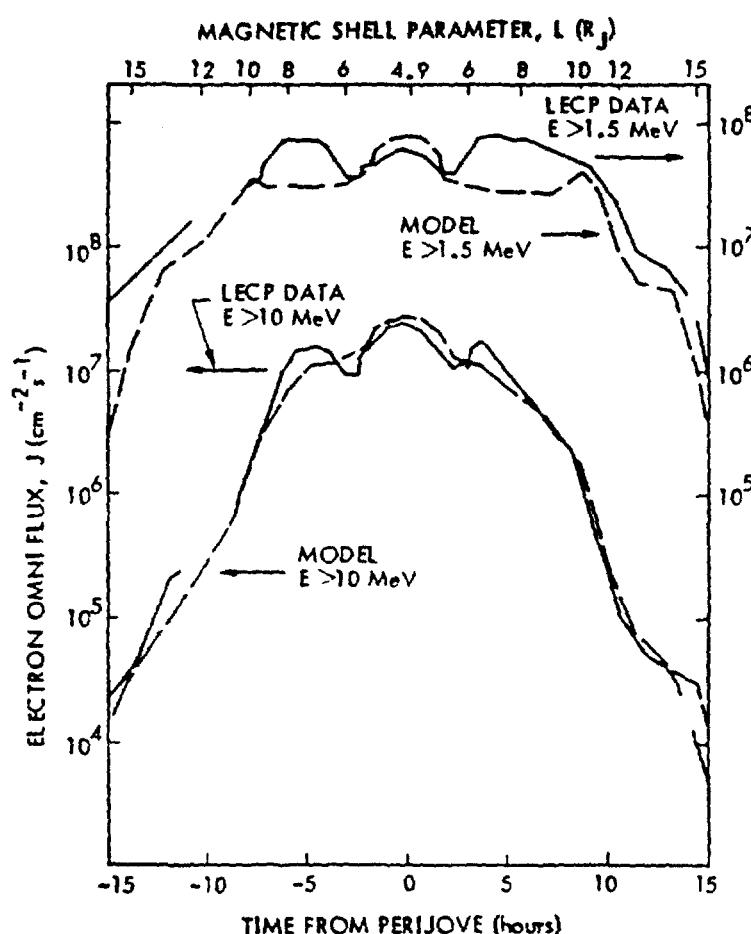


Fig. 48. Comparison^[45] of Voyager 1 LECP electron flux profiles with predictions from the Jupiter energetic electron model. Arrows indicate left or right flux scales.

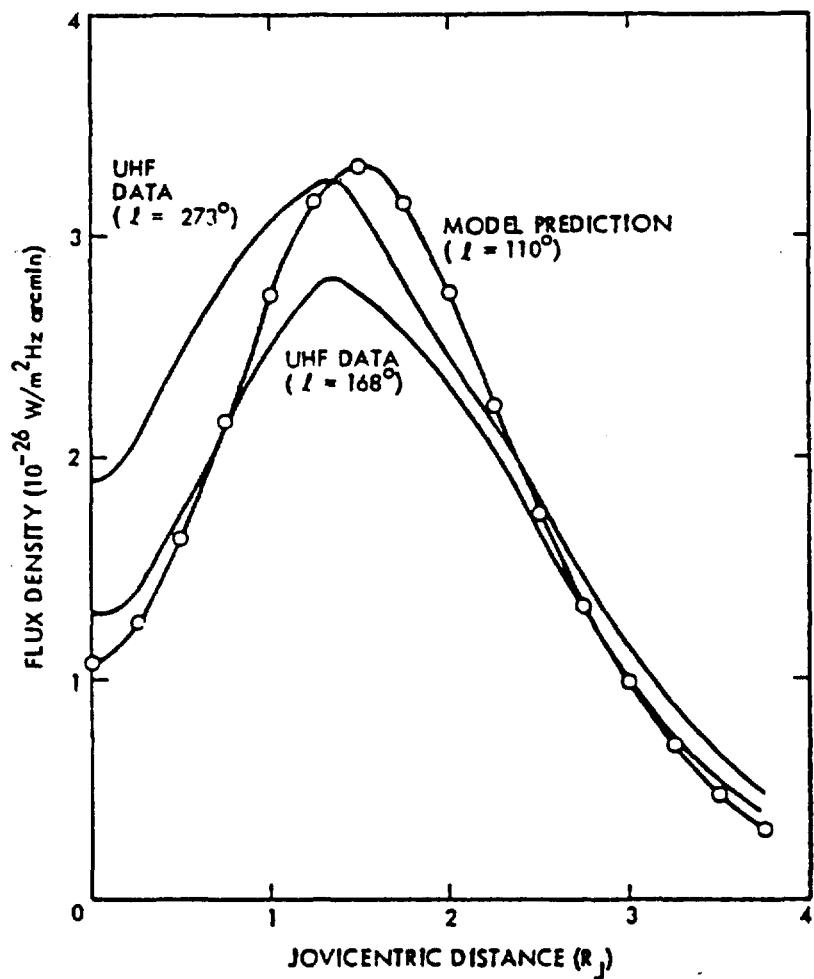


Fig. 49 Comparison of UHF strip brightness observed from Earth at 21-cm wavelength with predictions from the Jupiter energetic electron model. [45]

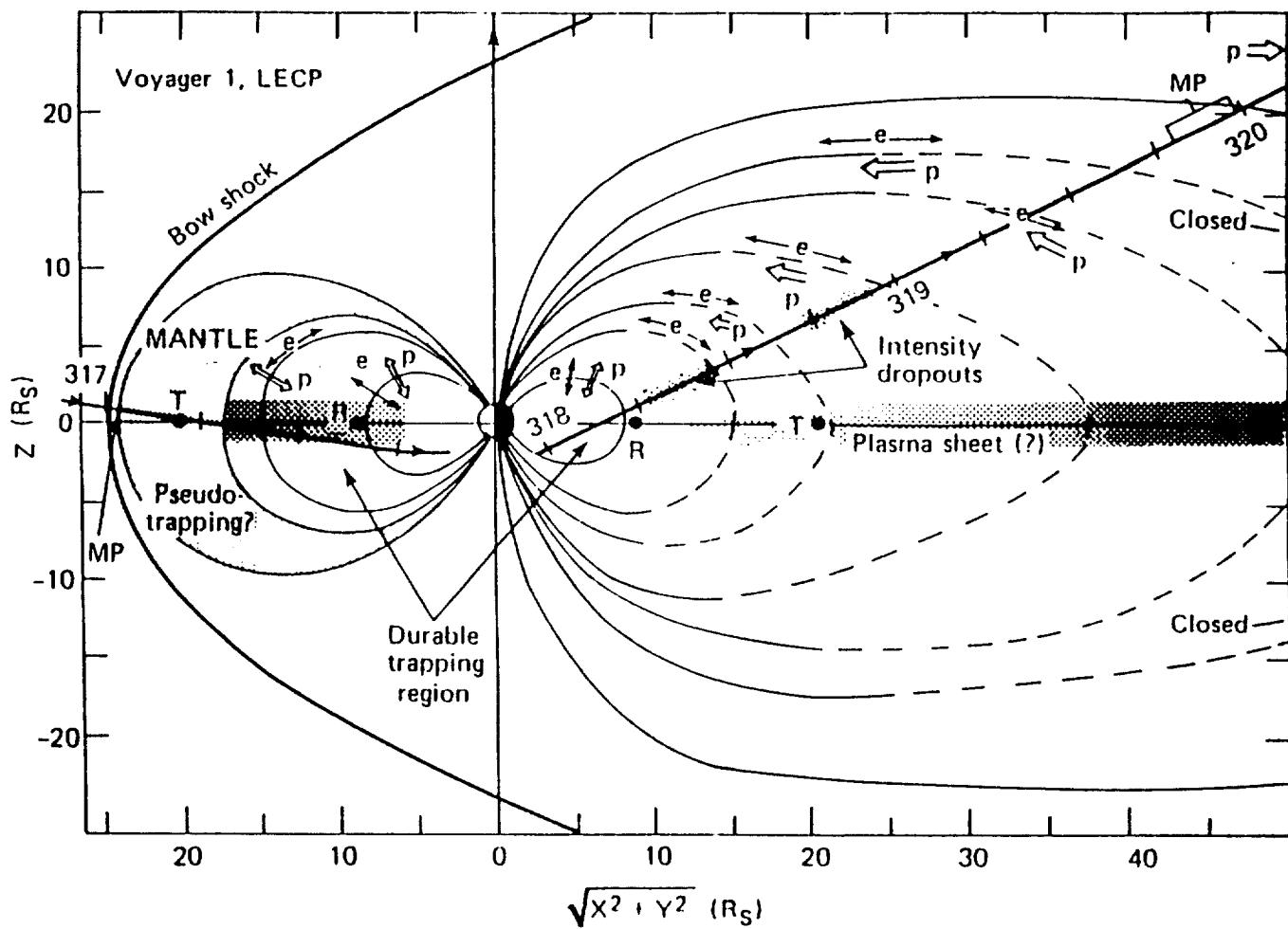


Fig. 50. A schematic representation of Saturn's magnetosphere in the r - z (trajectory) plane as revealed by the LECP data[59]. Tick marks on the trajectory are at 6-hour intervals from day 318 to 320. Note the Titan-associated mantle region outside $\sim 17 R_S$, and the presence of closed field lines in the tail lobe region. The phase of the second-order anisotropies relative to the magnetic field (B) is shown for both electron (e) and protons (p). R, Rhea; T., Titan; MP, magnetopause.

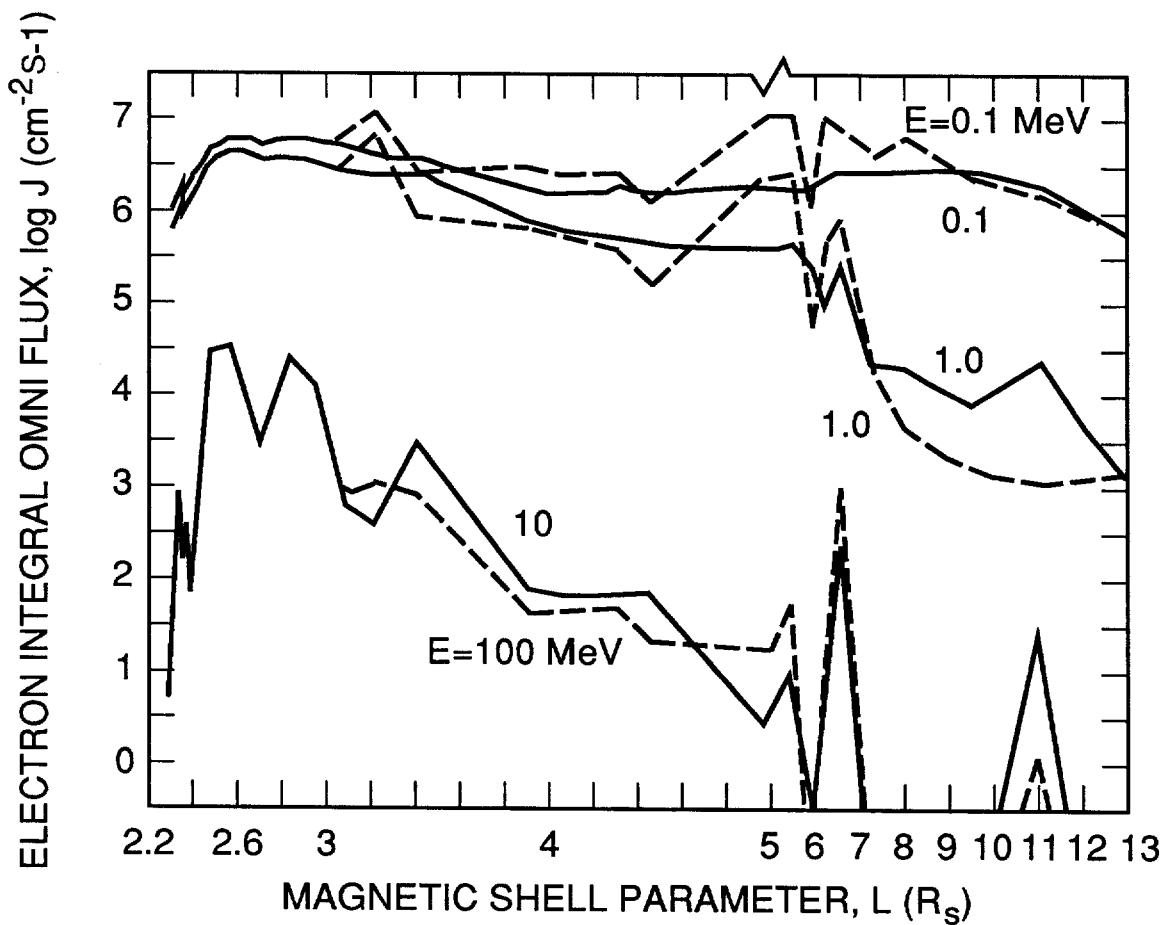


Fig. 51. Integral omnidirectional flux for the Saturn electron model [60] at three energy thresholds. The fluxes at the magnetic equator ($B_e/B=1.0$) and at high magnetic latitude ($B_e/B=0$) are represented respectively by the solid and dashed lines. The horizontal axis changes scale at $L=5$.

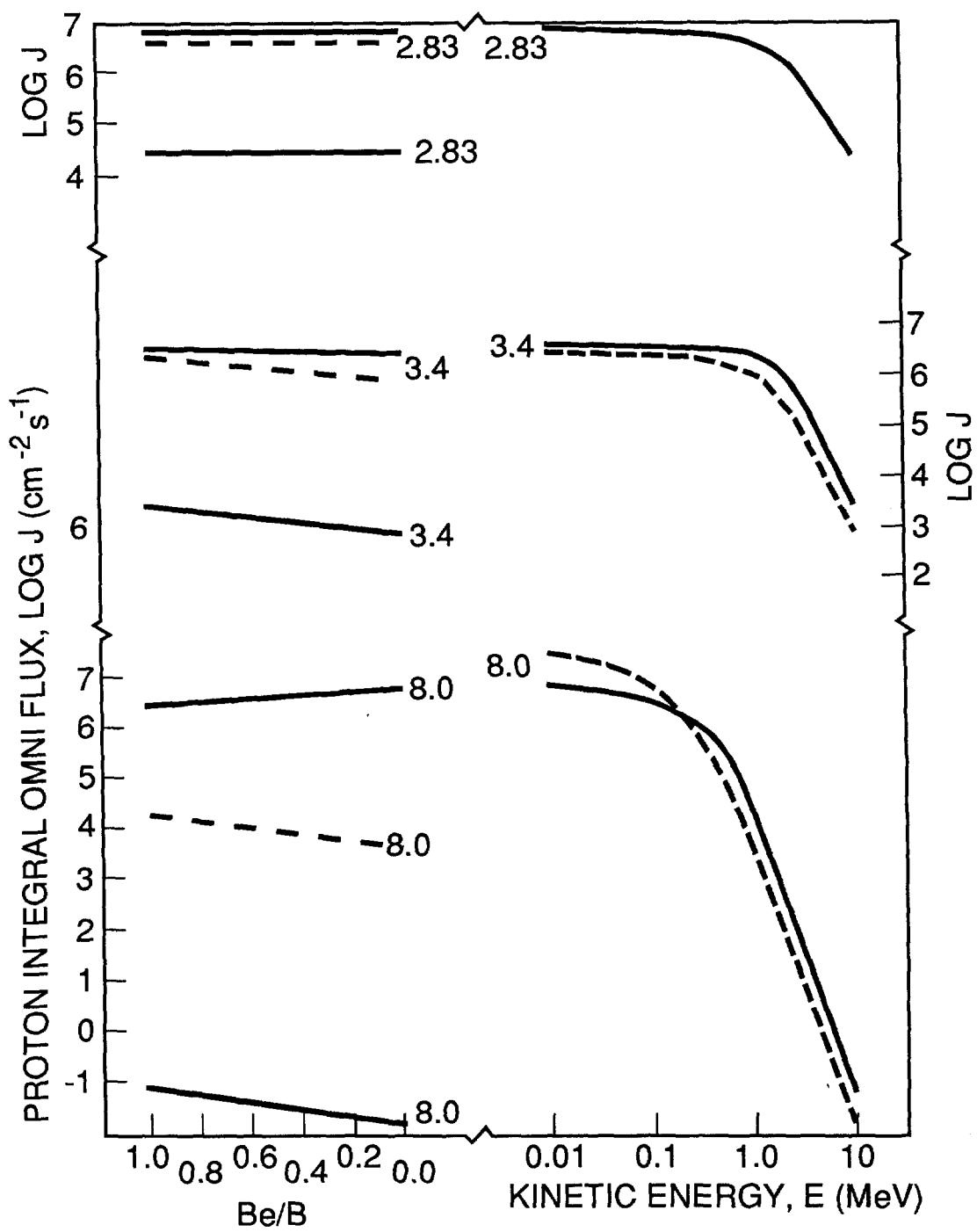


Fig. 52. Integral omnidirectional flux for the Saturn electron model [60] at three sample values of L (numbers adjacent to the curves). On the left, the dependencies on local magnetic field strength B are shown for energy thresholds of 0.1, 1.0, and 10 MeV (solid, dashed and solid lines) at each L . On the right spectra are shown for the magnetic equator ($B_e/B=1.0$, solid lines) and for high magnetic latitude ($B_e/B=0$, dashed lines) at each L .

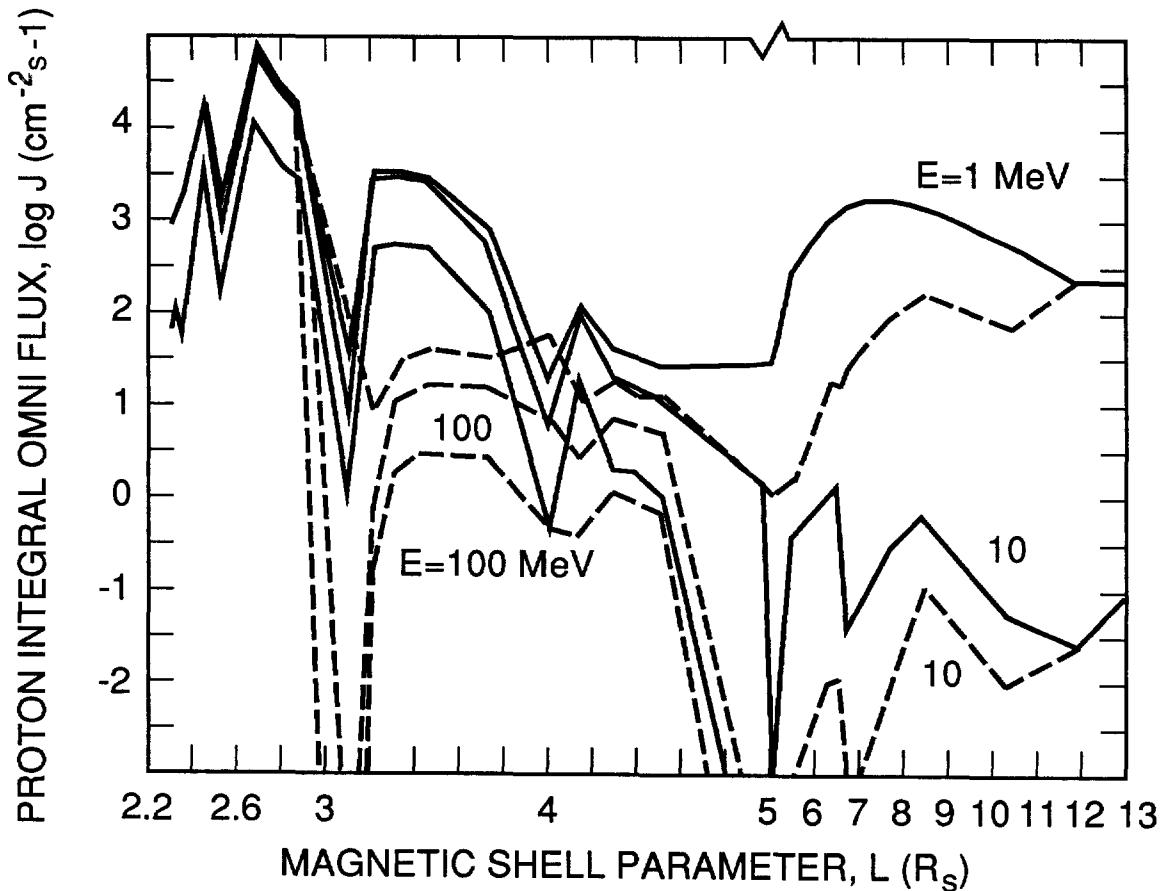


Fig. 53. Integral omnidirectional flux for the Saturn proton model [60] at three energy thresholds. The fluxes at the magnetic equator ($B_e/B=1.0$) and at high magnetic latitude ($B_e/B=0$) are represented respectively by the solid and dashed lines. The horizontal axis changes scale at $L=5$.

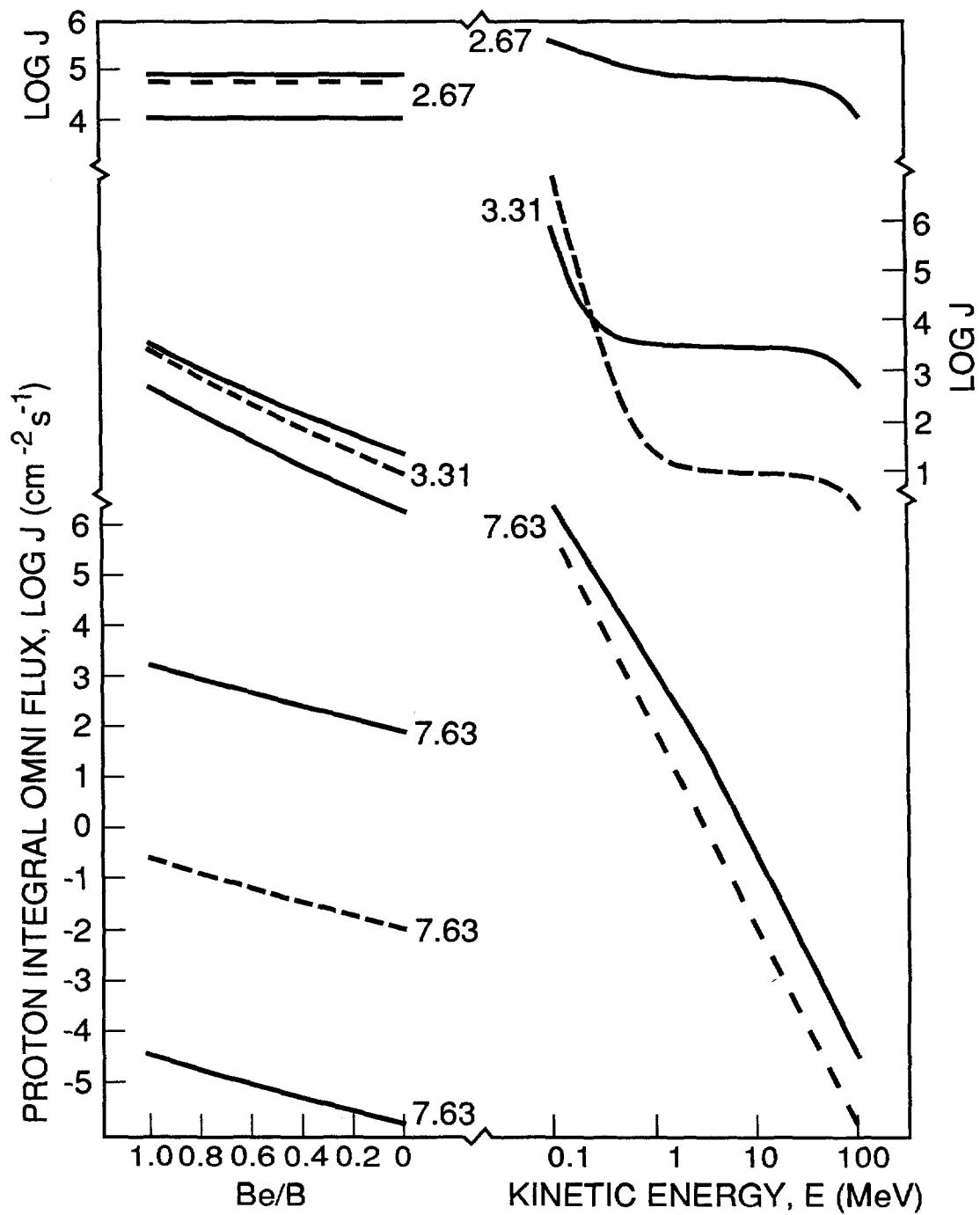


Fig. 54. Integral omnidirectional flux for the Saturn proton model [60] at three sample values of L (numbers adjacent to the curves). On the left, the dependencies on local magnetic field strength B are shown for energy thresholds of 1, 10, and 100 MeV (solid, dashed and solid lines) at each L . On the right spectra are shown for the magnetic equator ($B_e/B=1.0$, solid lines) and for magnetic latitude ($B_e/B=0$, dashed lines) at each L .

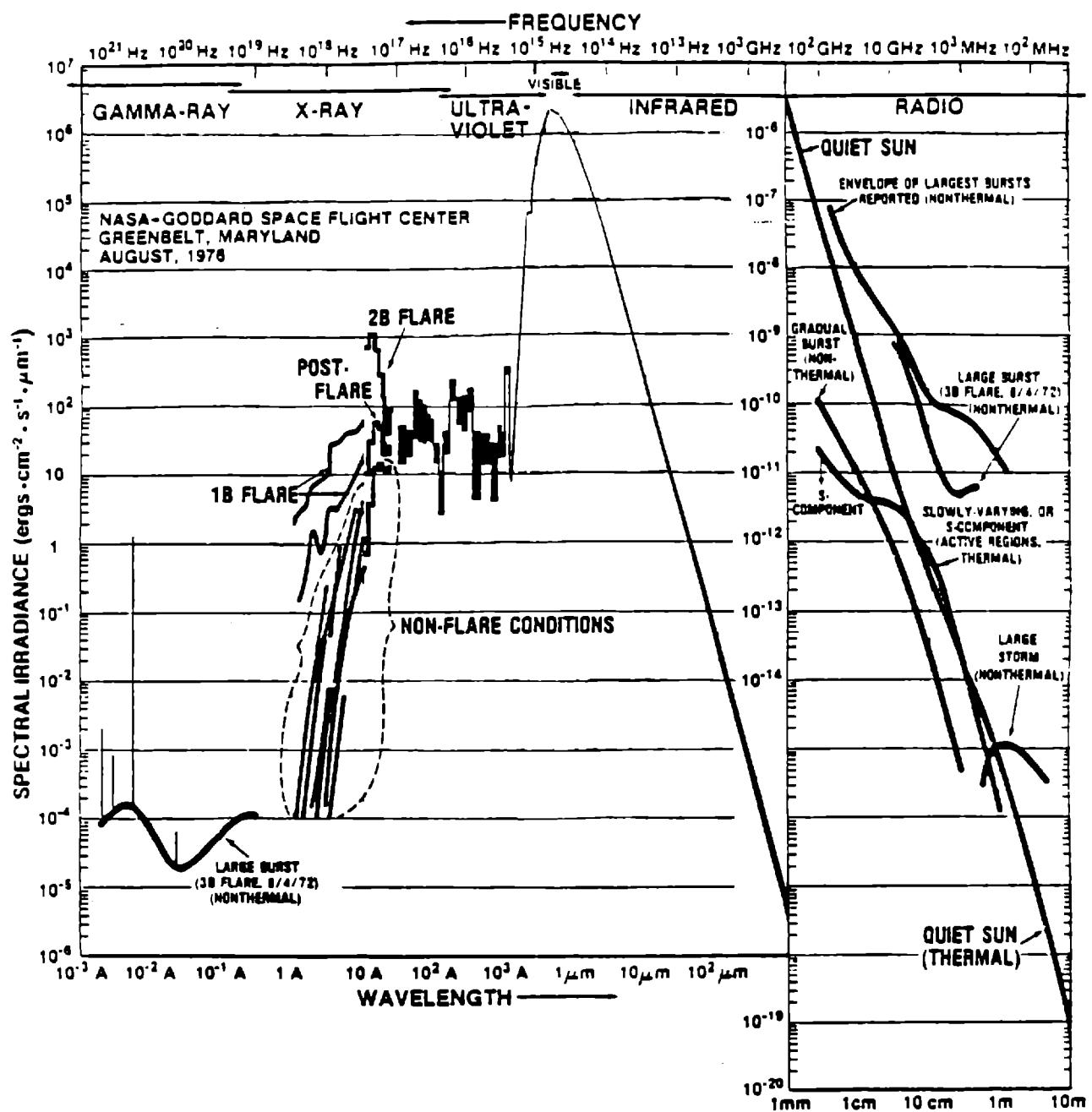


Fig. 55. The spectral distribution of solar irradiance^[61] from a wavelength of 10⁻³ Å to 10 m.

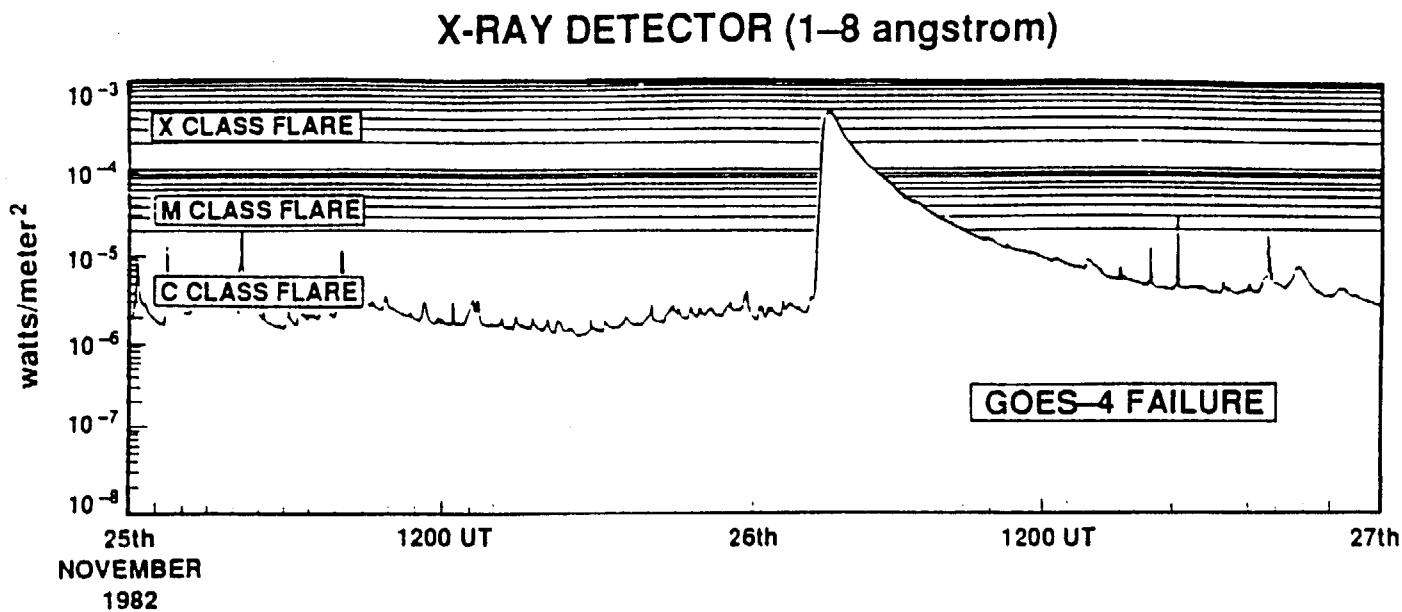


Fig. 56. X-ray flux (watts/m²) between 1 and 8 Å versus time for the November 26, 1982 solar flare as measured by the NOAA GOES spacecraft at geosynchronous orbit.^[63] Also shown are the various classes of X-ray flares.

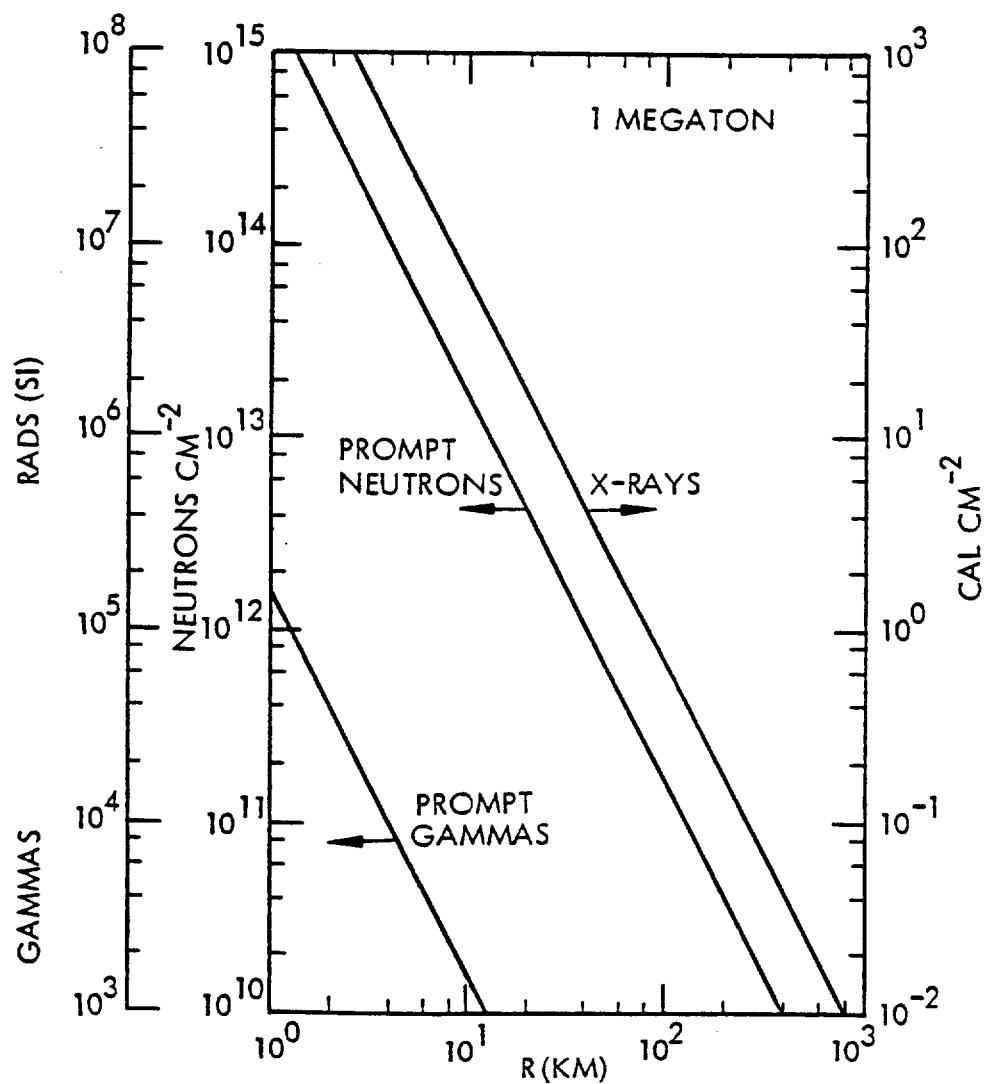


Fig. 57. The prompt radiation environment in space for a 1 megaton nuclear weapon as a function of distance from the detonation.^[64] The neutron environment is in terms of particle fluence (n/cm^2). The γ -ray environment is in dosage (rads(Si)). The X-ray environment is in terms of energy fluence (cal/cm^2).

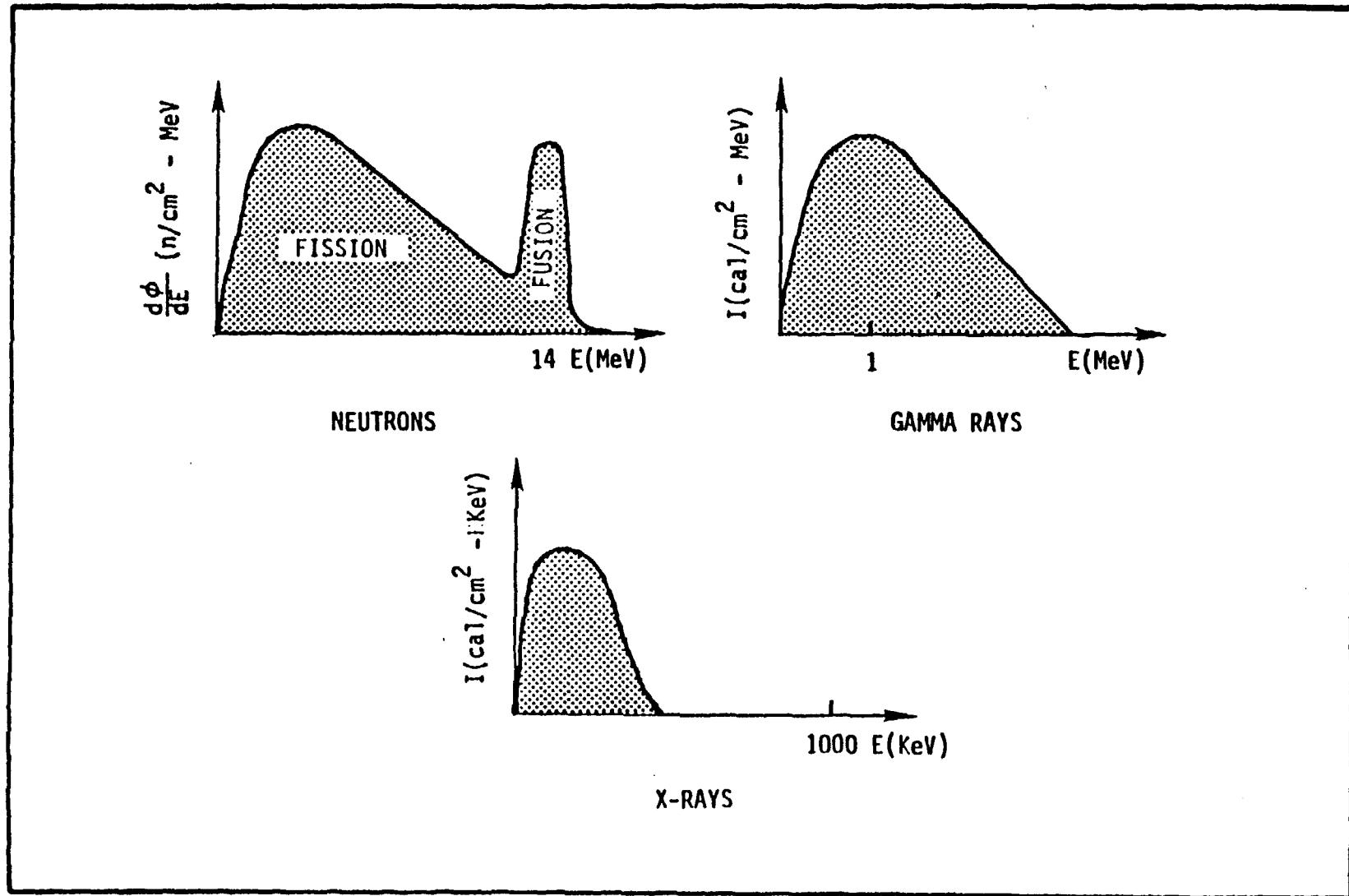
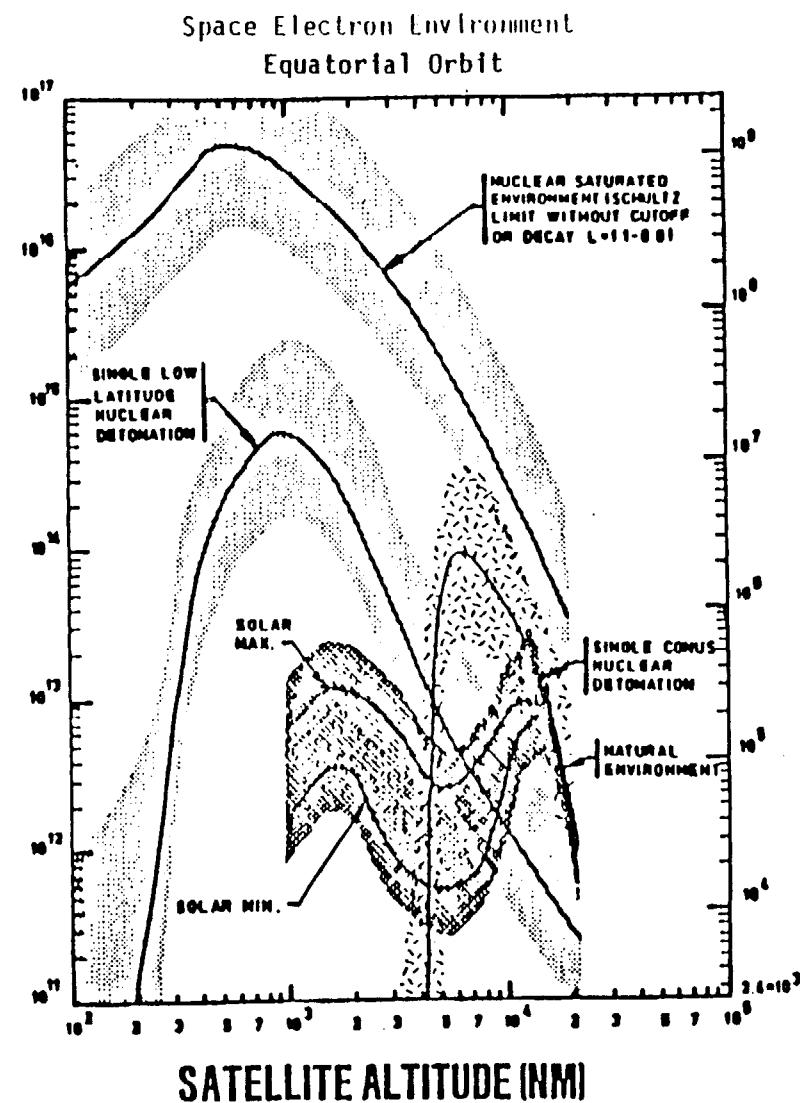


Fig. 58. Schematic representations of the prompt radiation differential spectra for nuclear weapon generated neutrons, γ -rays, and X-rays. [65, 92]

DOSE [RADS-UNSHIELDED SILICON]



TEN DAY FLUENCE ($E>1$ MeV) (e^-/cm^2)

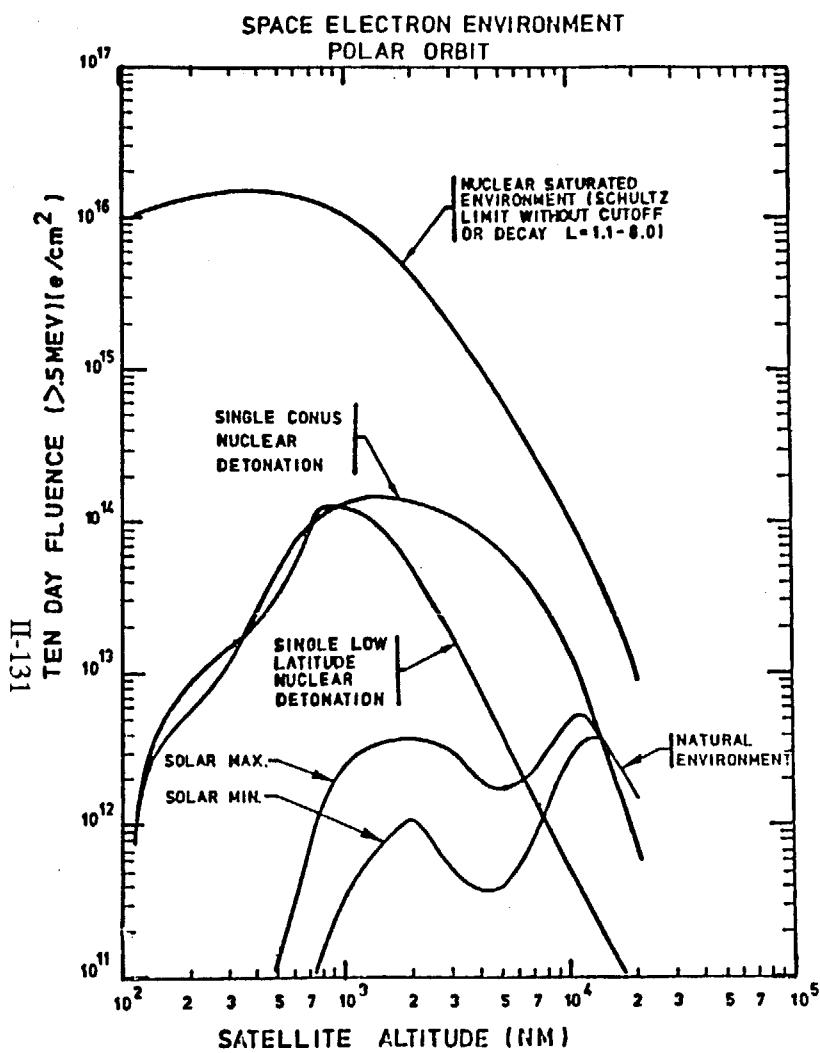


Fig. 59. The natural and manmade electron radiation dose environments for a.) polar and b.) equatorial orbits with altitudes between 10^2 and 10^5 nautical miles. The plots are in units of a.) e^-/cm^2 -10 days for $E>.5$ MeV and b.) units of e^-/cm^2 -10 days for $E>1$ MeV or rads(Si) for 0 shielding. Curves are presented for nuclear saturated environments (see text), a single continental US detonation, a single low latitude nuclear detonation, and the natural environment (for solar maximum and solar minimum). The shading for the equatorial plot shows the range of expected variations. (Adapted from [68] and [65].)

COMPARISON OF NATURAL AND ARTIFICIAL EQUATORIAL ELECTRON SPECTRA

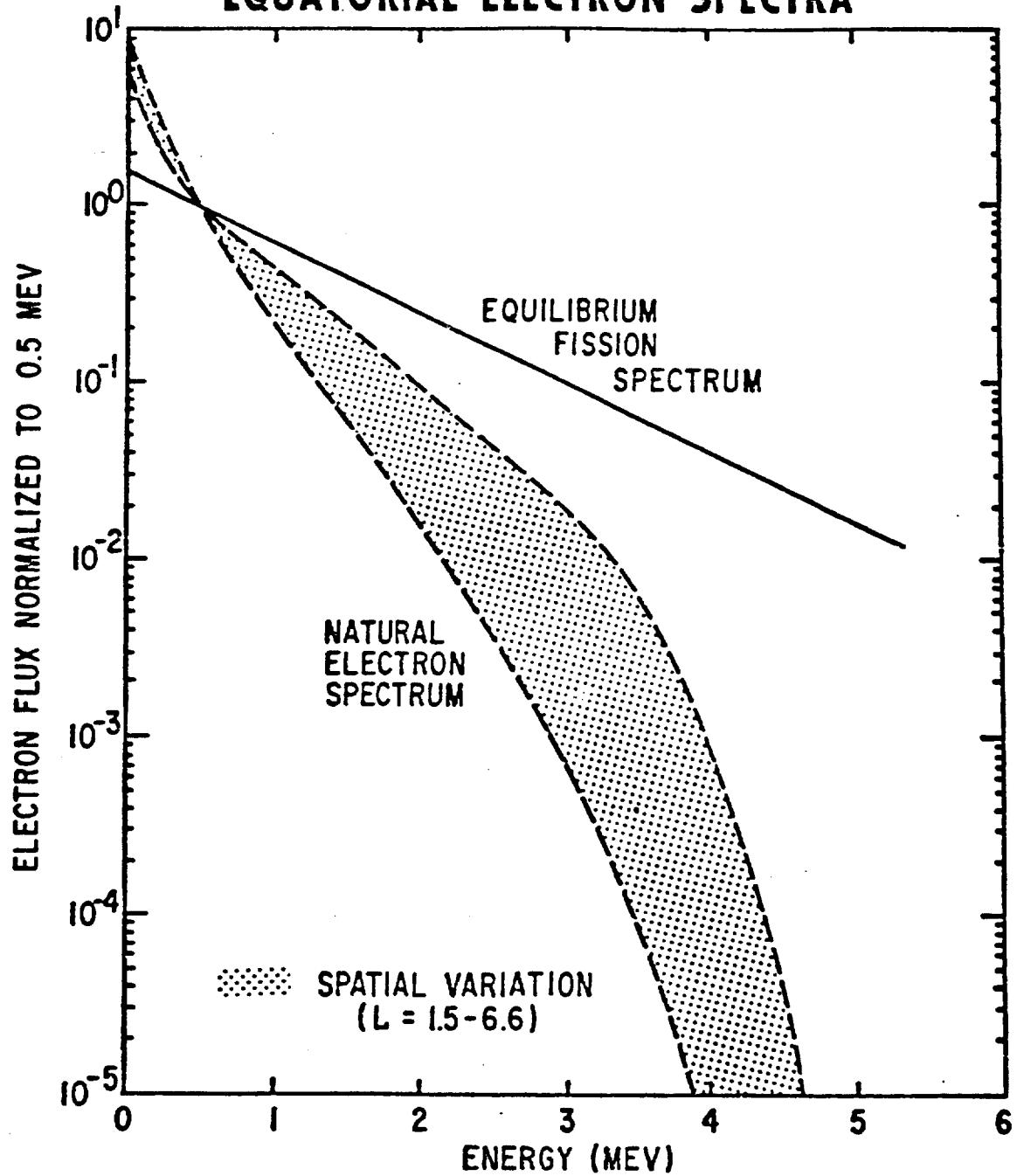


Fig. 60. Comparison of the natural and fission-produced equatorial electron flux spectra as a function of energy. The variation in the natural environment spectrum is with altitude (in terms of L-shell). The spectra are normalized to 1 at $E=0.5$ Mev.[68]

Fig. 61. Illustration of the construction of a Galileo General Purpose Heat Source (GPHS) Radioisotope Thermoelectric Generator (RTG).[69, 70]

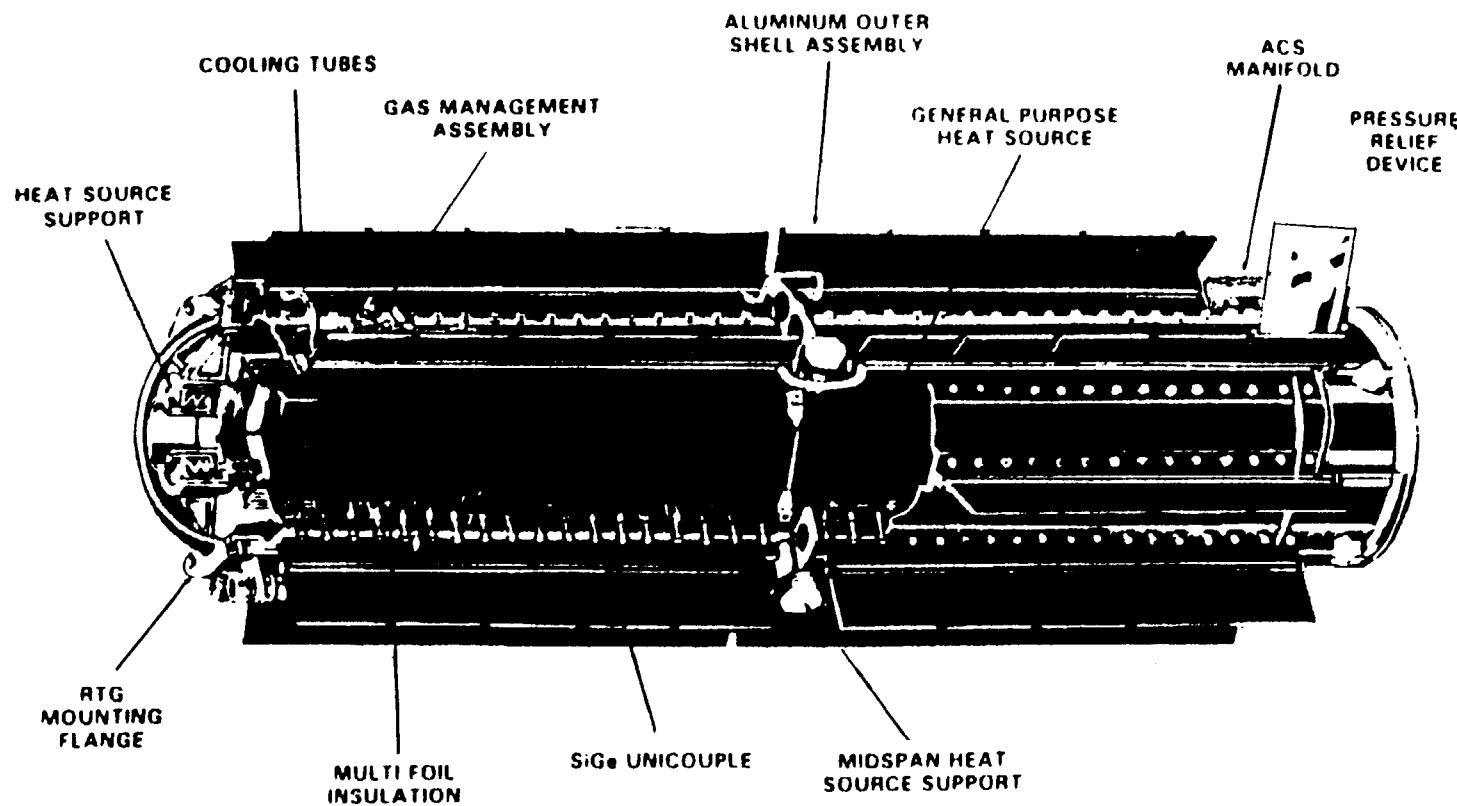


FIGURE 61.

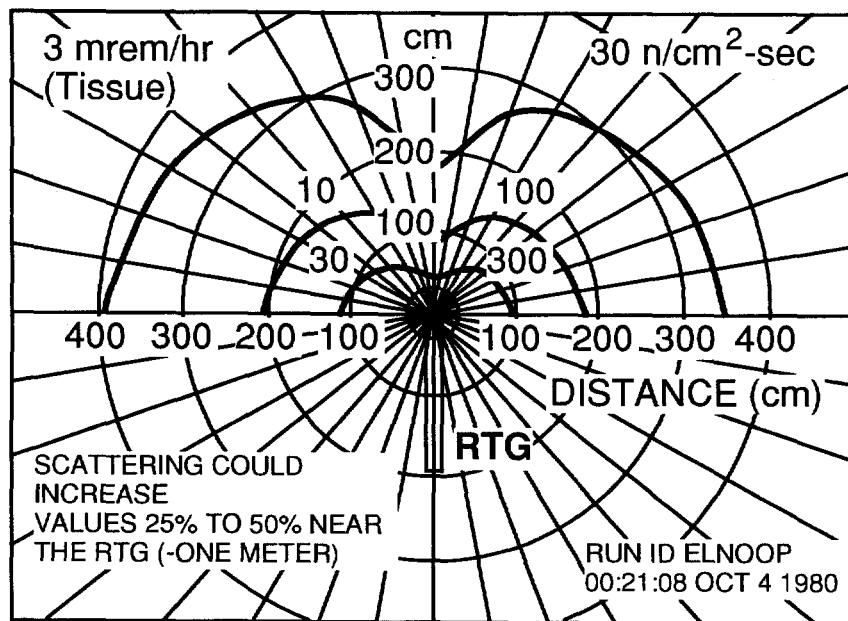


Fig. 62. Neutron isodose isoflux plots for a single 4500 Watts(th) GPHS RTG (point kernel calculation). [71]

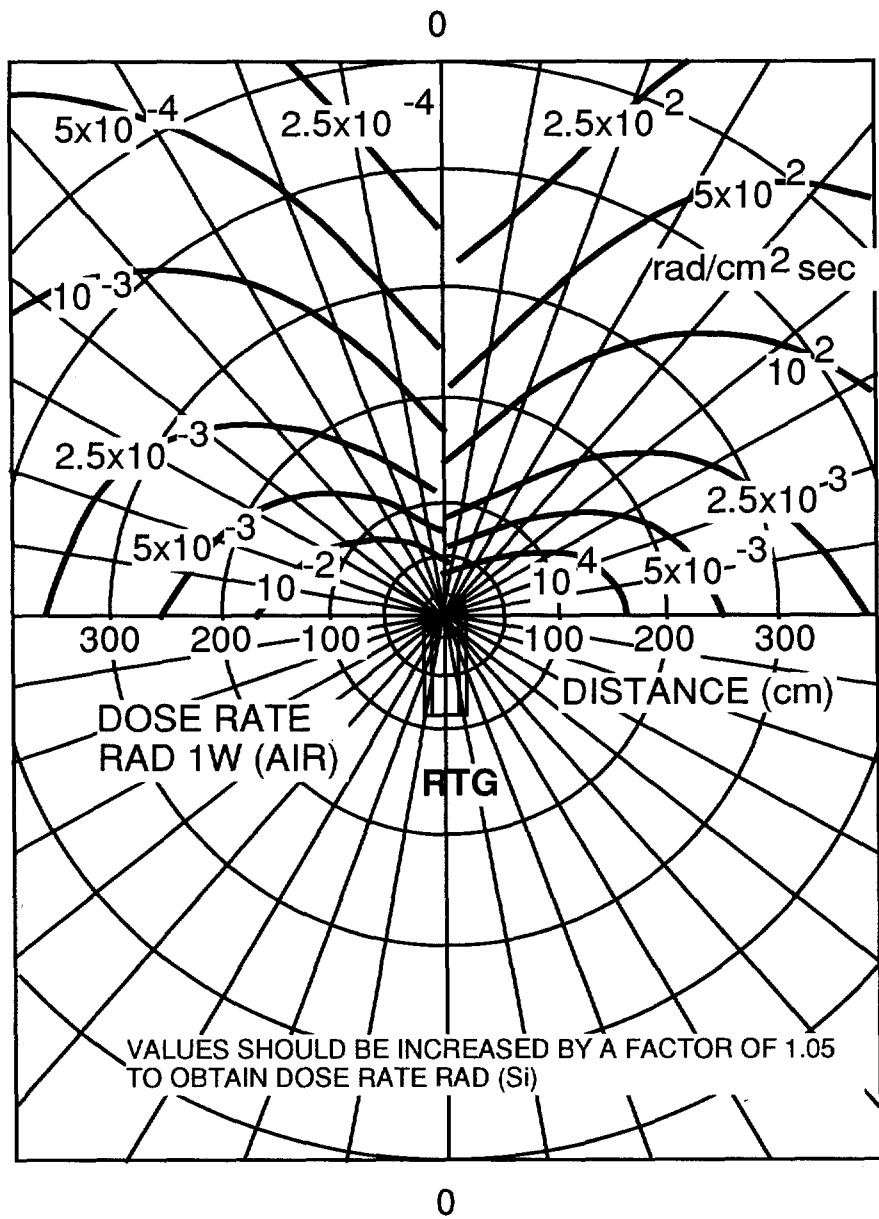


Fig. 63. Gamma isodose-isoflux contours for a single 4500 Watt(th) GPHS RTG (5-year-old fuel). [71] Values should be increased by a factor of 2.0 for the worst case; to determine refined estimates, use factors taken from table.

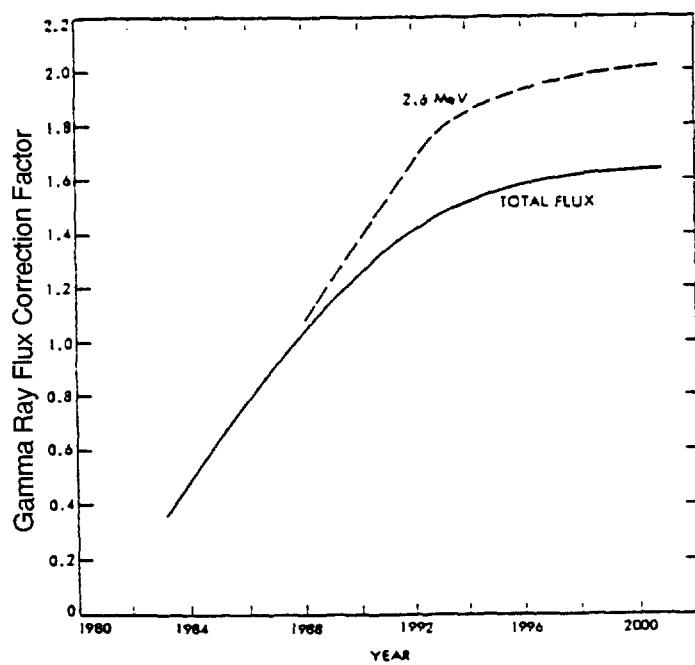


Fig. 64. RTG γ -ray flux correction factor for fuel aging. [71]

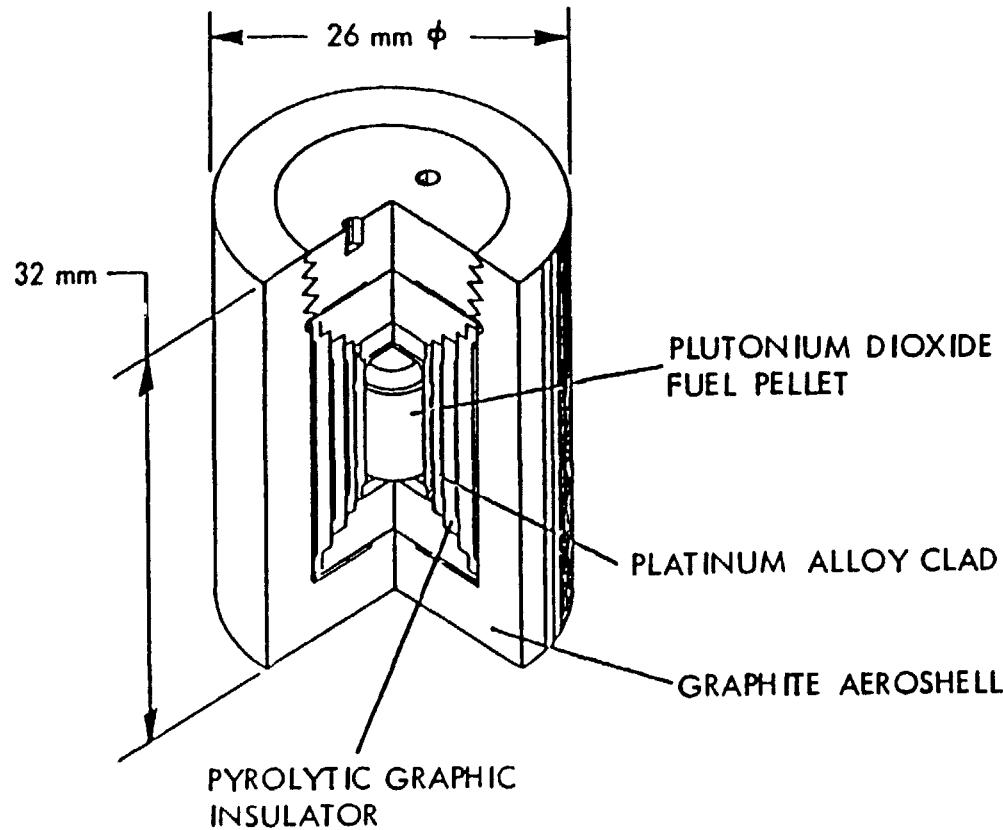


Fig. 65. Schematic of a Galileo Lightweight Radioisotope Heater Unit (LHRHU).^[69]

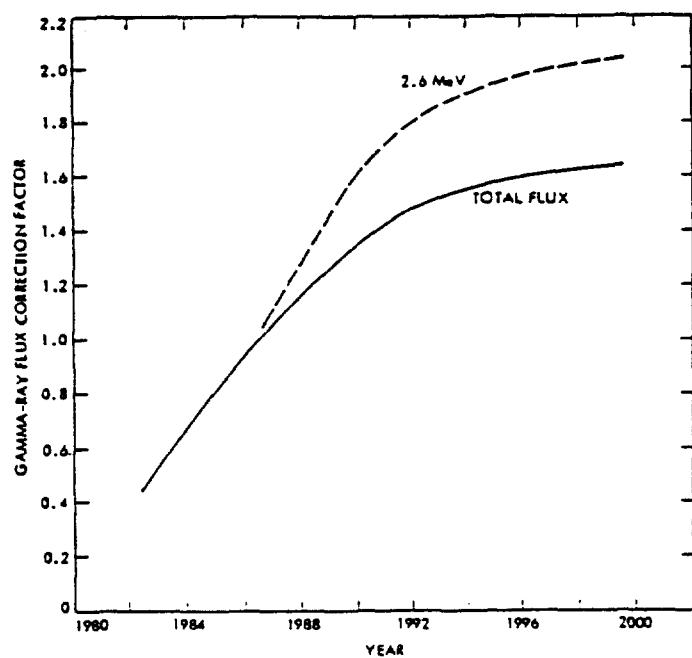


Fig. 66. LHRHU γ -ray flux correction factor for fuel aging. [71]

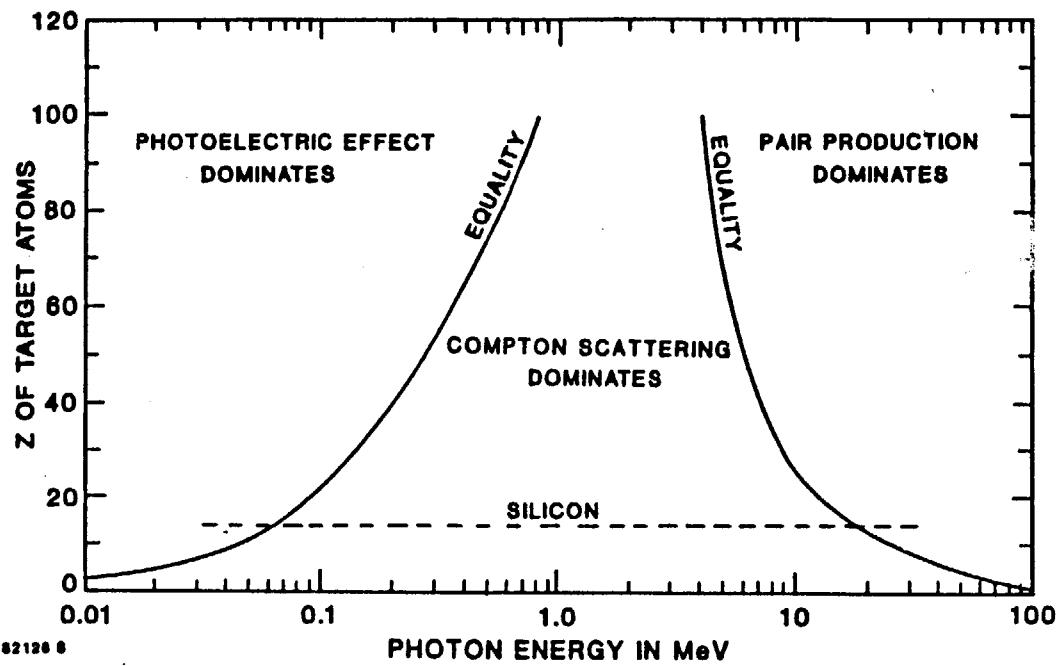


Fig. 67. Illustration of the relative importance of the three photon interactions as a function of Z and photon energy. The solid lines correspond to equal interaction cross sections for the neighboring effects. The dashed line illustrates the situation for photon interactions with silicon.^[73]

STOPPING POWERS IN SILICON

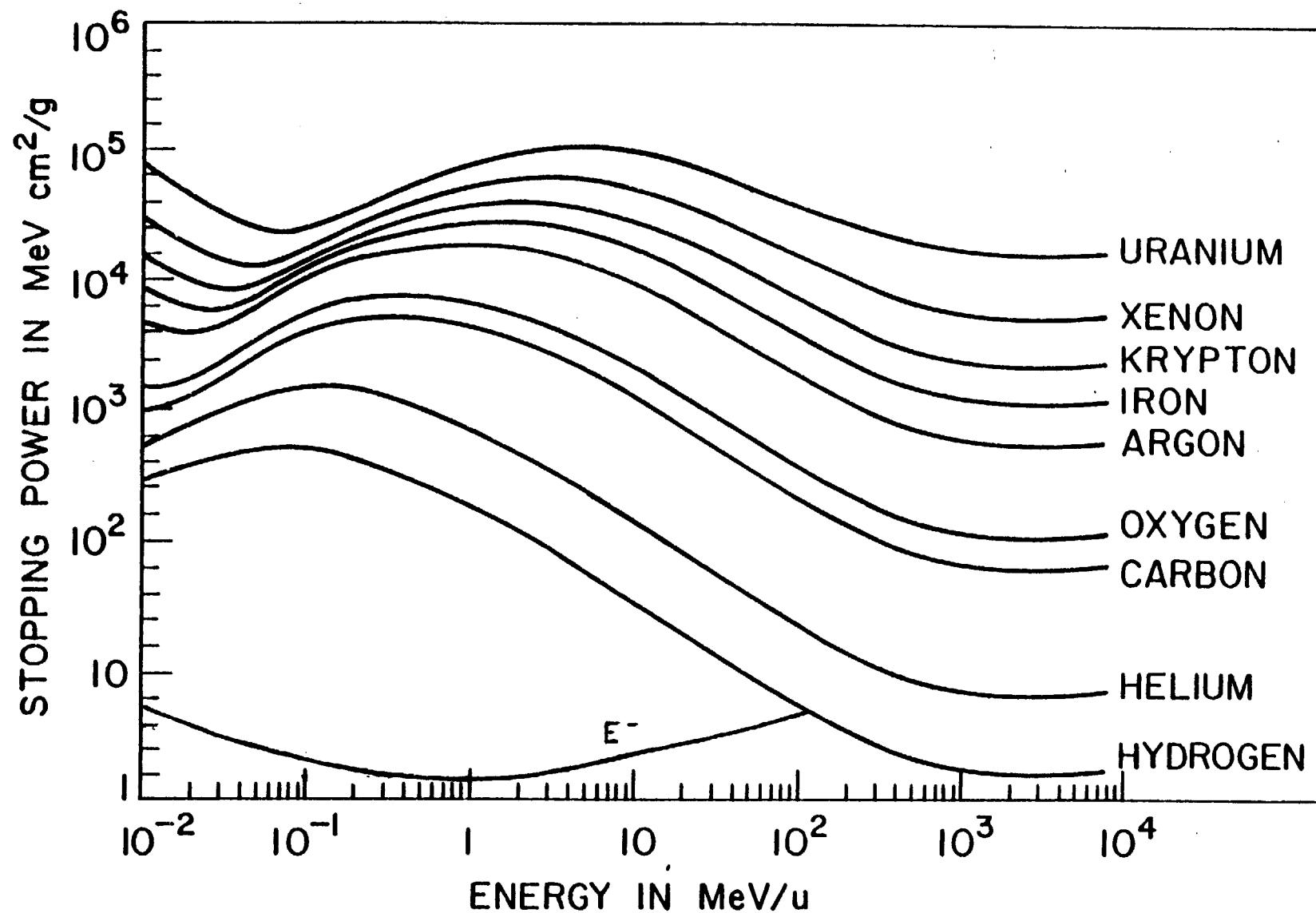


Fig. 68. The stopping power (or LET) in MeV-cm²/g versus energy per atomic mass unit for a variety of ions (MeV/ μ) and electrons (MeV) in silicon.^[3]

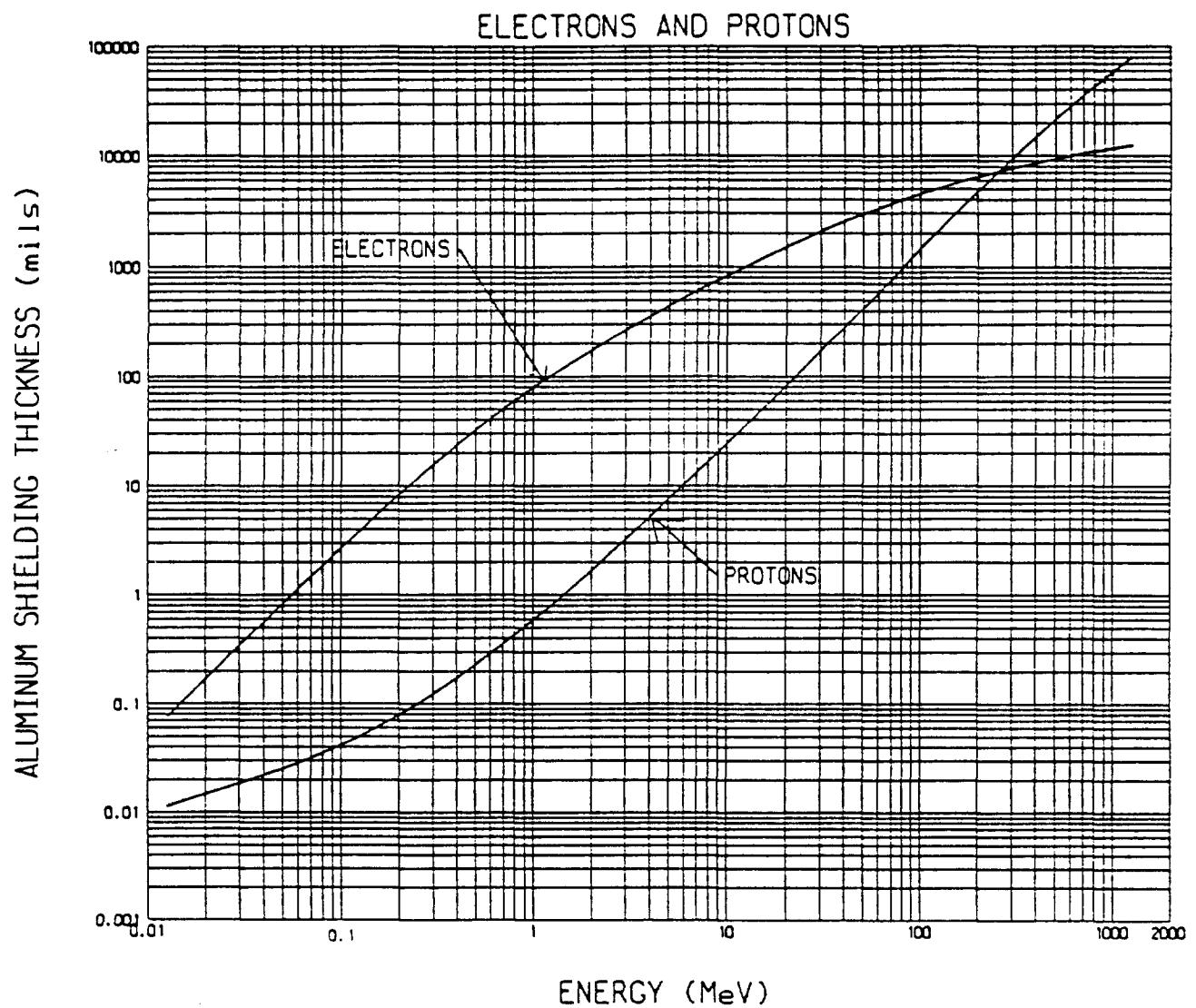


Fig. 69. Minimum penetration energy for electrons and protons versus shield thickness.

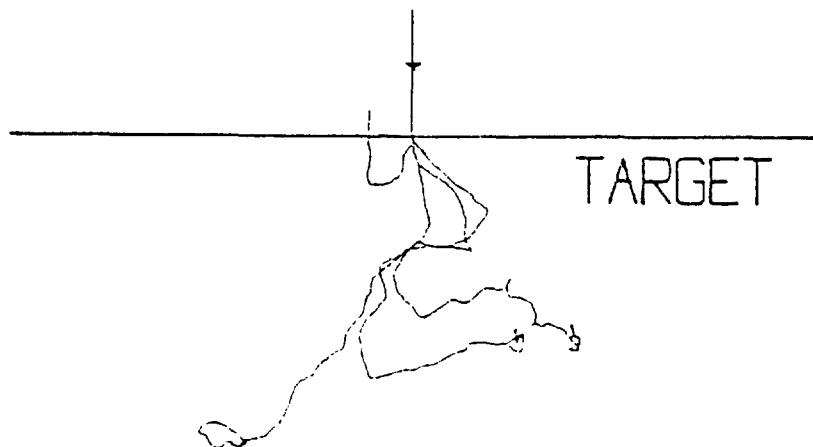


Fig. 70a. Computer plot showing five electron trajectories in a Cu target with a 20 keV incident beam normal to the surface.[77, 78]

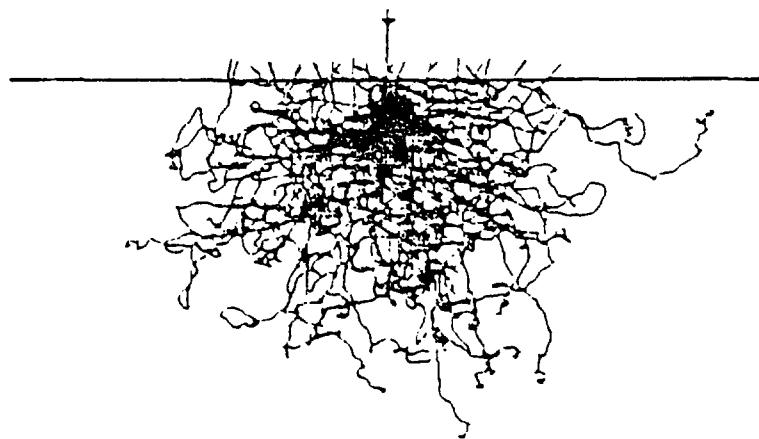


Fig. 70b. Monte Carlo results for 100 electron trajectories in a Cu target with 20 keV incident electron energy.[77, 78]

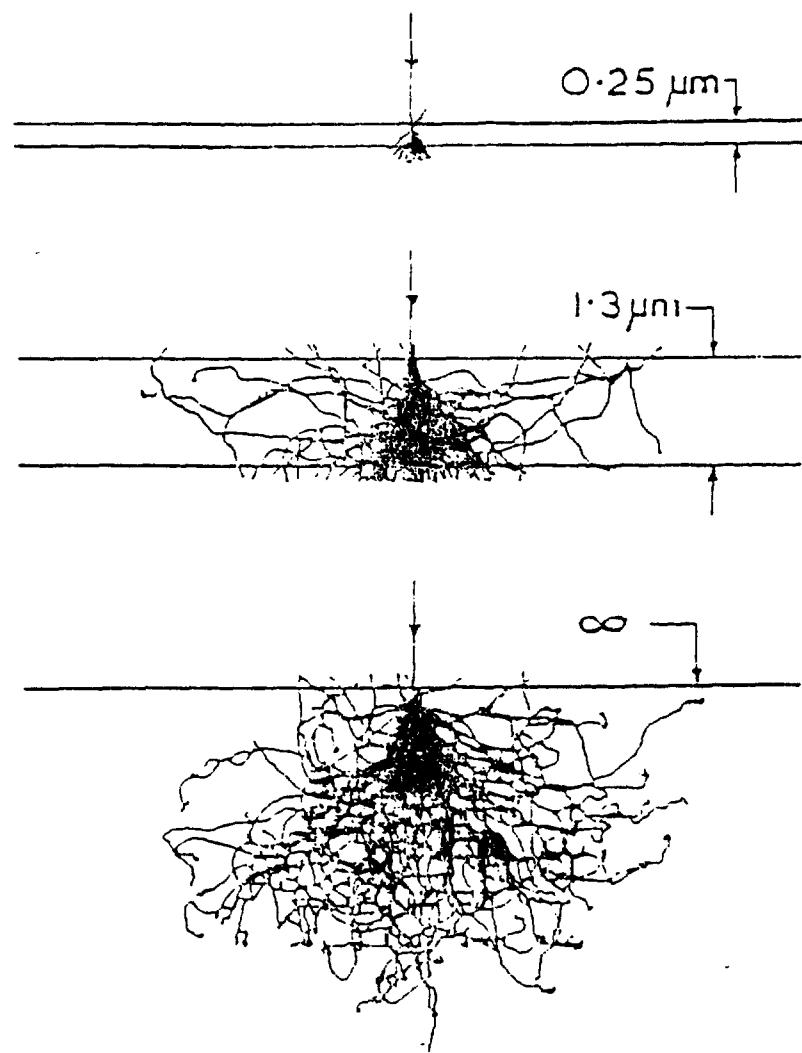


Fig. 71. Effect of varying target thickness; electron trajectories in an Al target of thickness $0.25 \mu\text{m}$, $1.3 \mu\text{m}$, and infinity, for an incident electron energy of 20 keV.[77, 78]

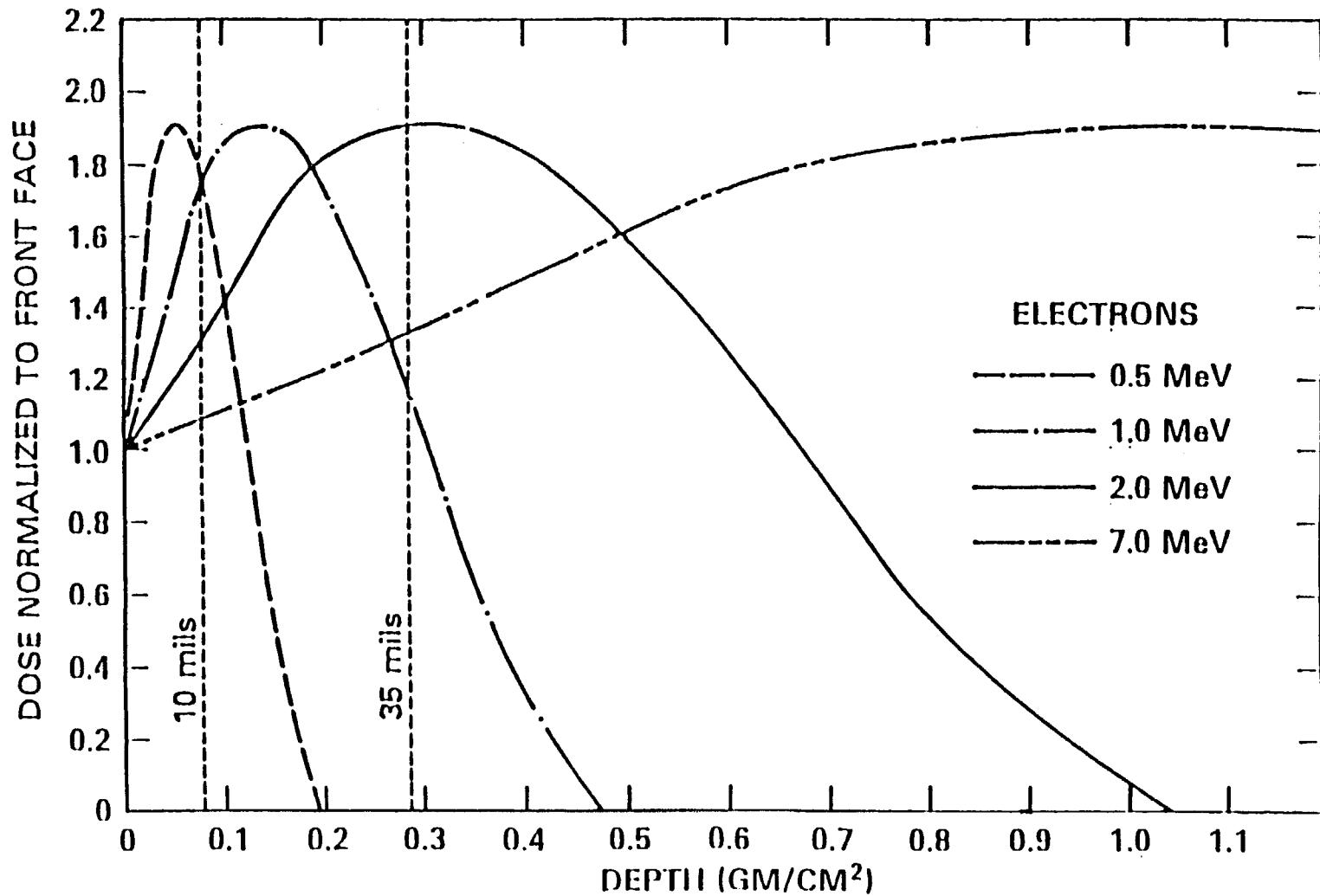


Fig. 72. Electron dose as a function of depth for CaFMn TLDs.^[80] The dosage is normalized to 1 at the front face.

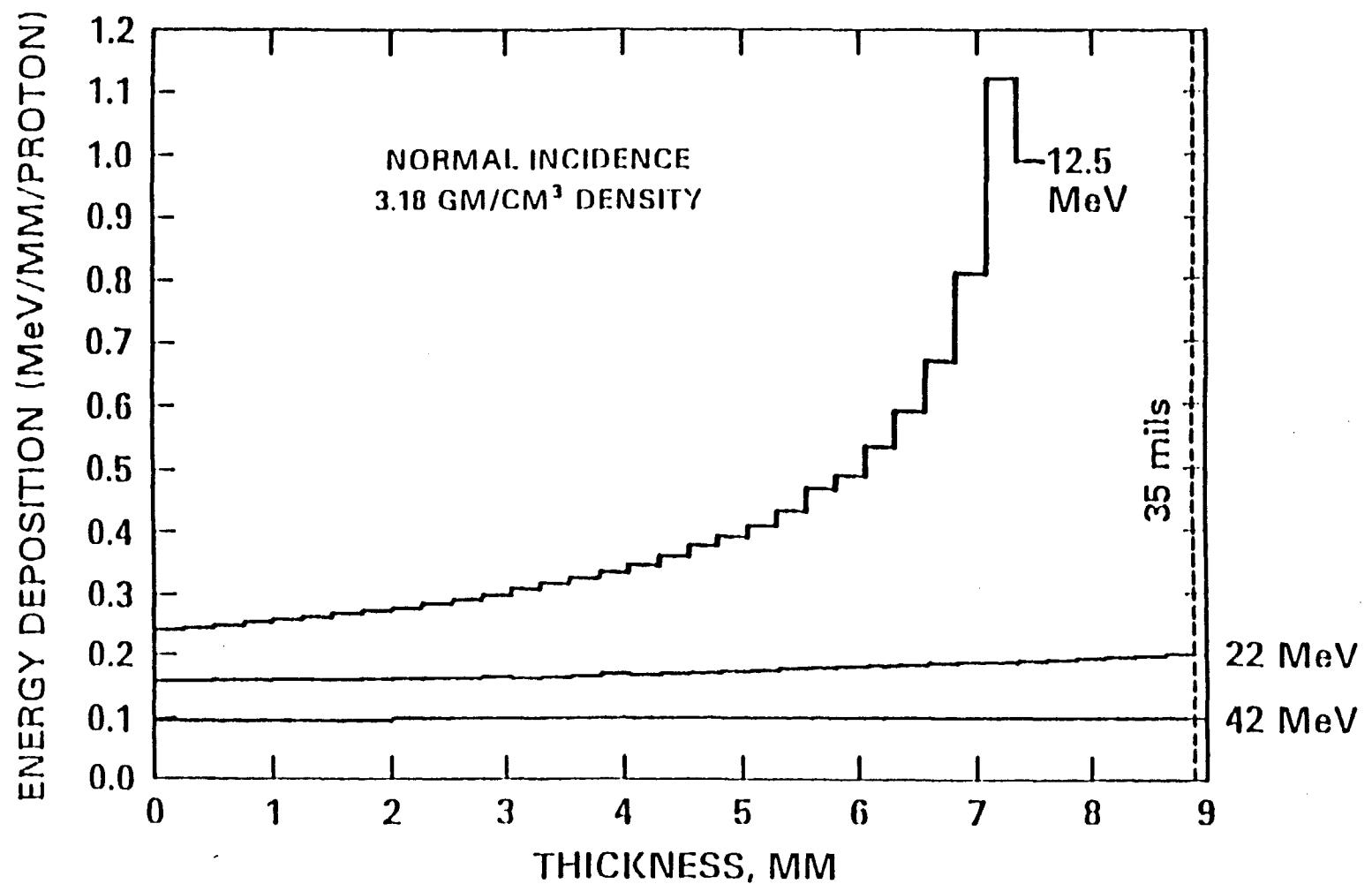


Fig. 73. Proton energy deposition (MeV/mm-proton) as a function of depth for CaF_2 (density of 3.18 g/cm^3).[81]

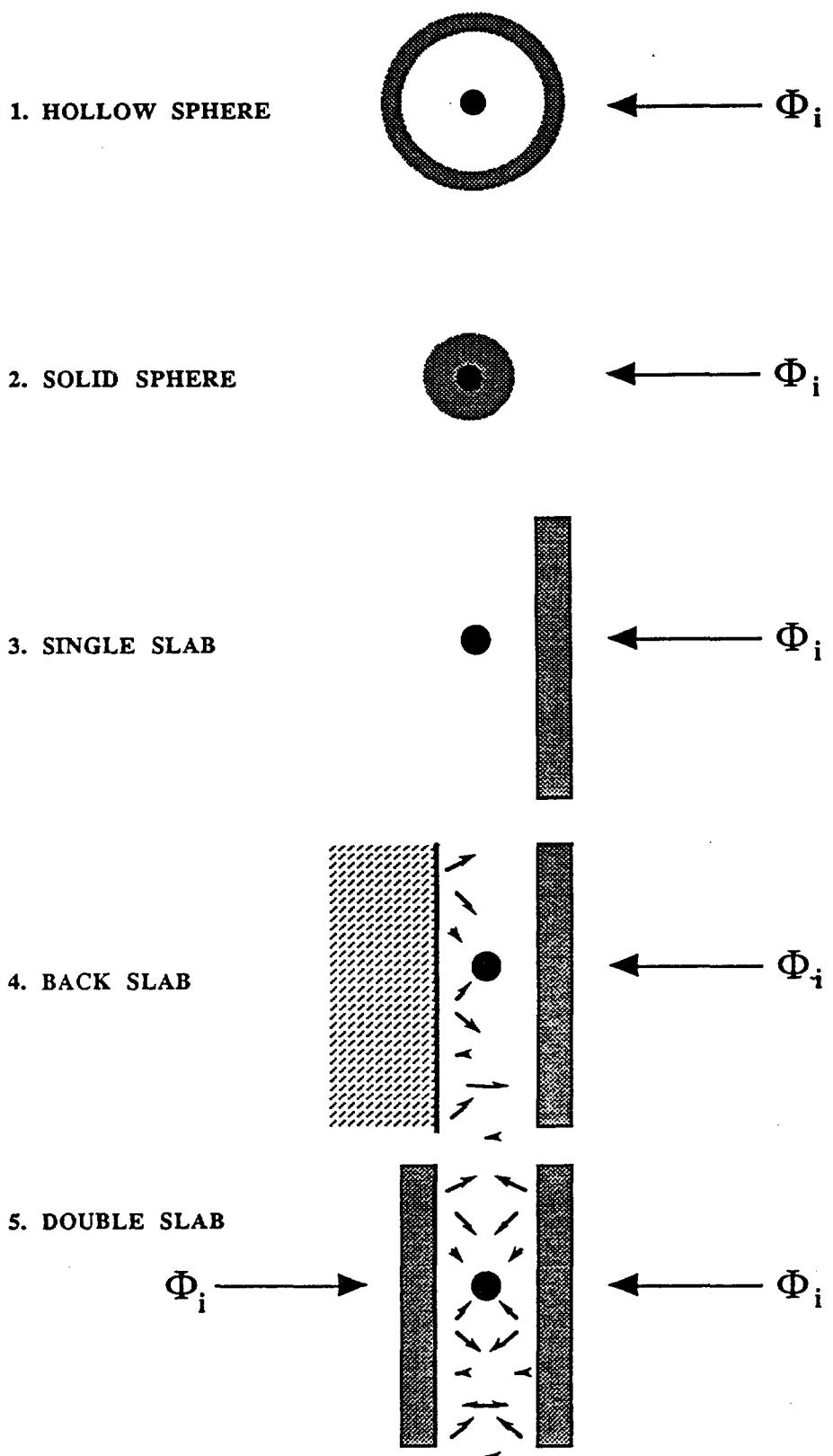


Fig. 74. 5 shielding configurations considered in the NOVICE code for calculating dosage. Note that examples 3 and 4 are doubled by the code.

RANGE OF IONS IN ALUMINUM

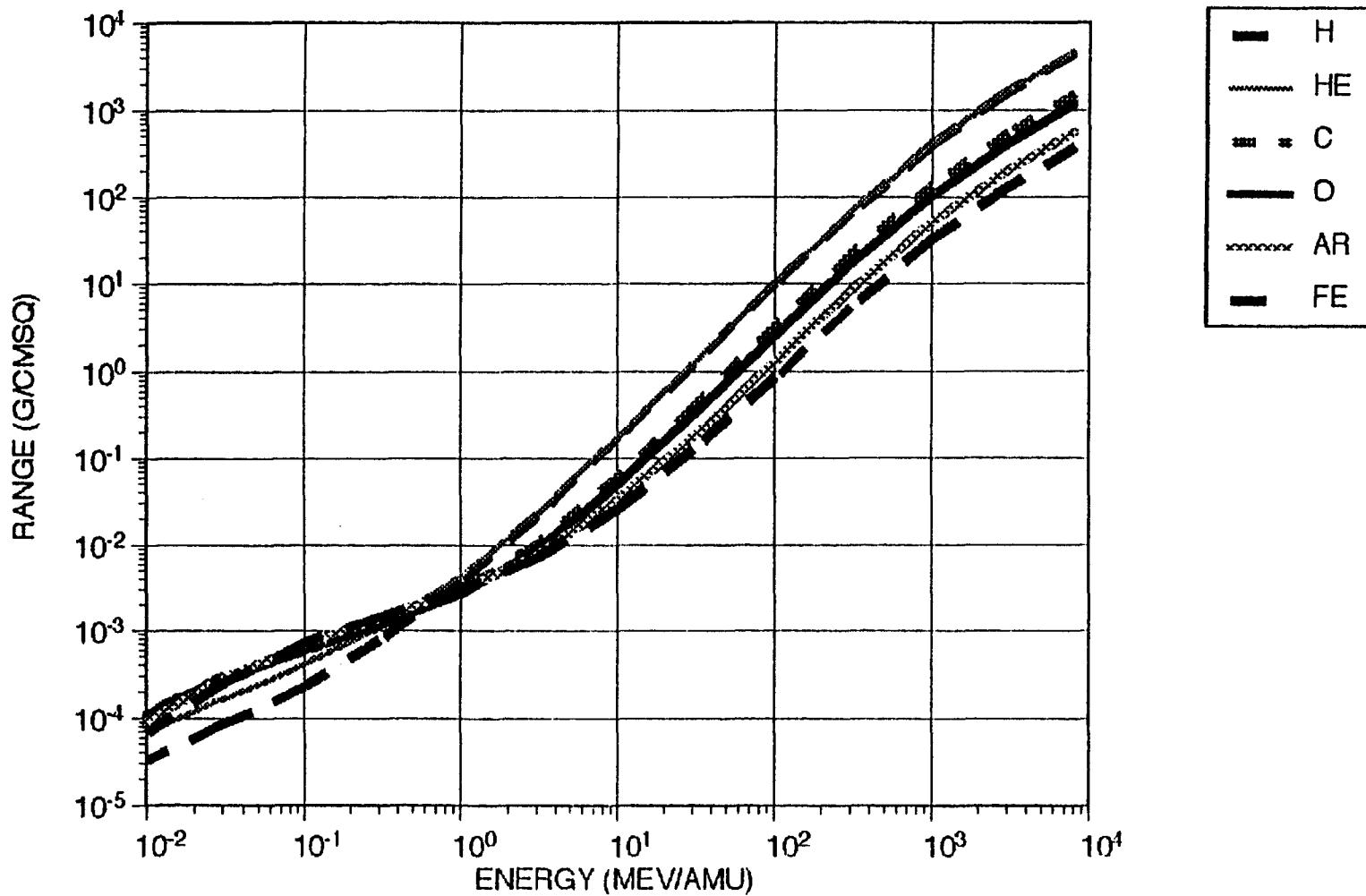


Fig. 75. Ion range versus energy in Al^[32] for H, He, C, O, Ar, and Fe. The range is in units of g-cm² and the energy in MeV/ μ .

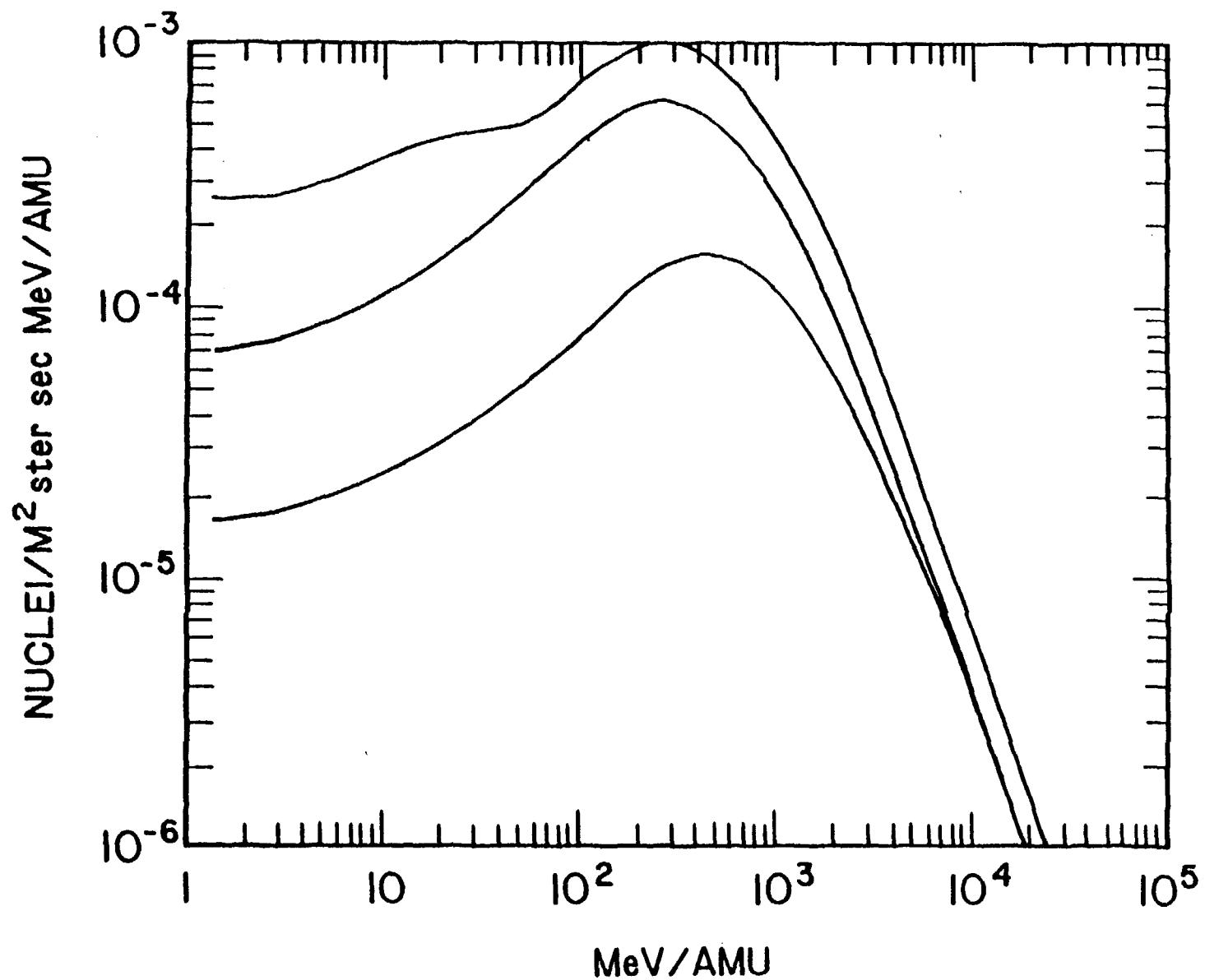


Fig. 76. Shielding attenuated cosmic ray differential iron spectra for the 3 cases in Figure 2: 90% worst case (upper), solar minimum (middle) and solar maximum (lower). [32] These spectra are for 1 AU (no magnetospheric shielding) and behind 0.025 inches of aluminum shielding. [32]

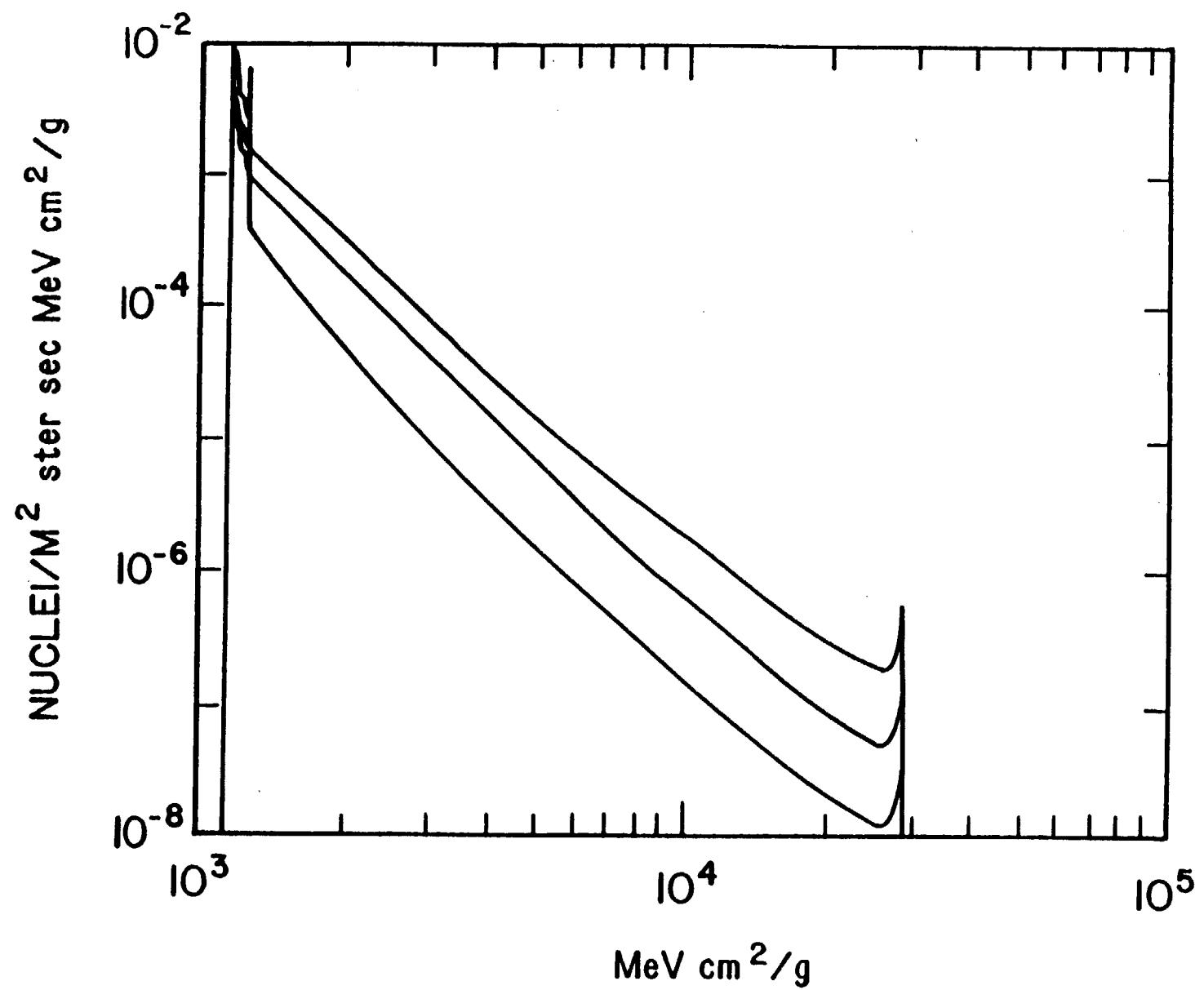


Fig. 77. The same three cases as in Figures 2 and 76, but now the spectra have been transformed into differential LET spectra.^[32]

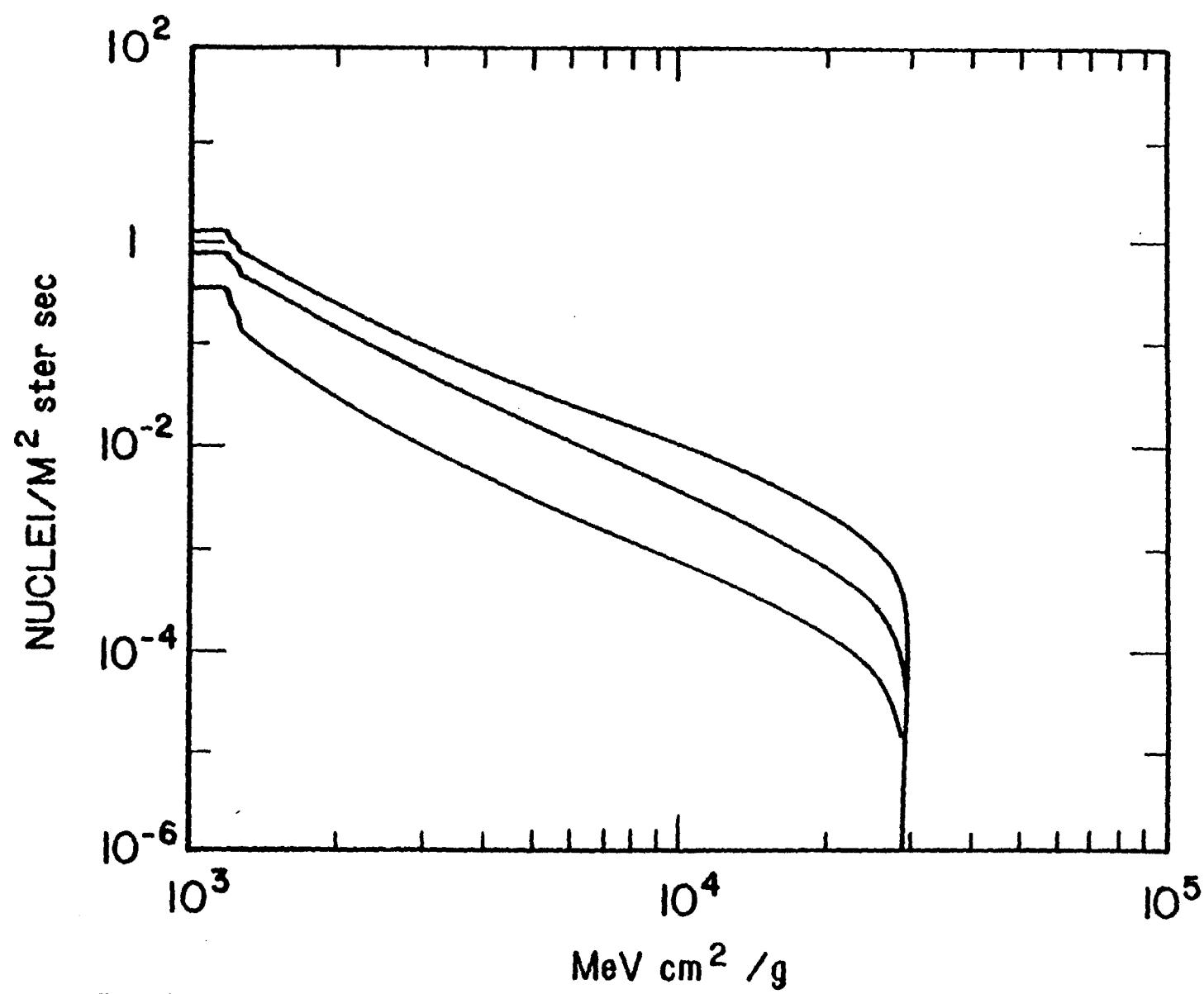


Fig. 78. The same three cases as in Figures 2 and 76, but now the spectra of Figure 77 have been integrated to give the integral LET spectra (Heinrich curves) for Fe.[32]

Fig. 79. Radiation dosage from the trapped proton environment (AP8) for solar-quiet conditions and for the first Earth-Moon Transfer Orbit {1994/01/30(07:00)--1994/02/09(12:00)}.

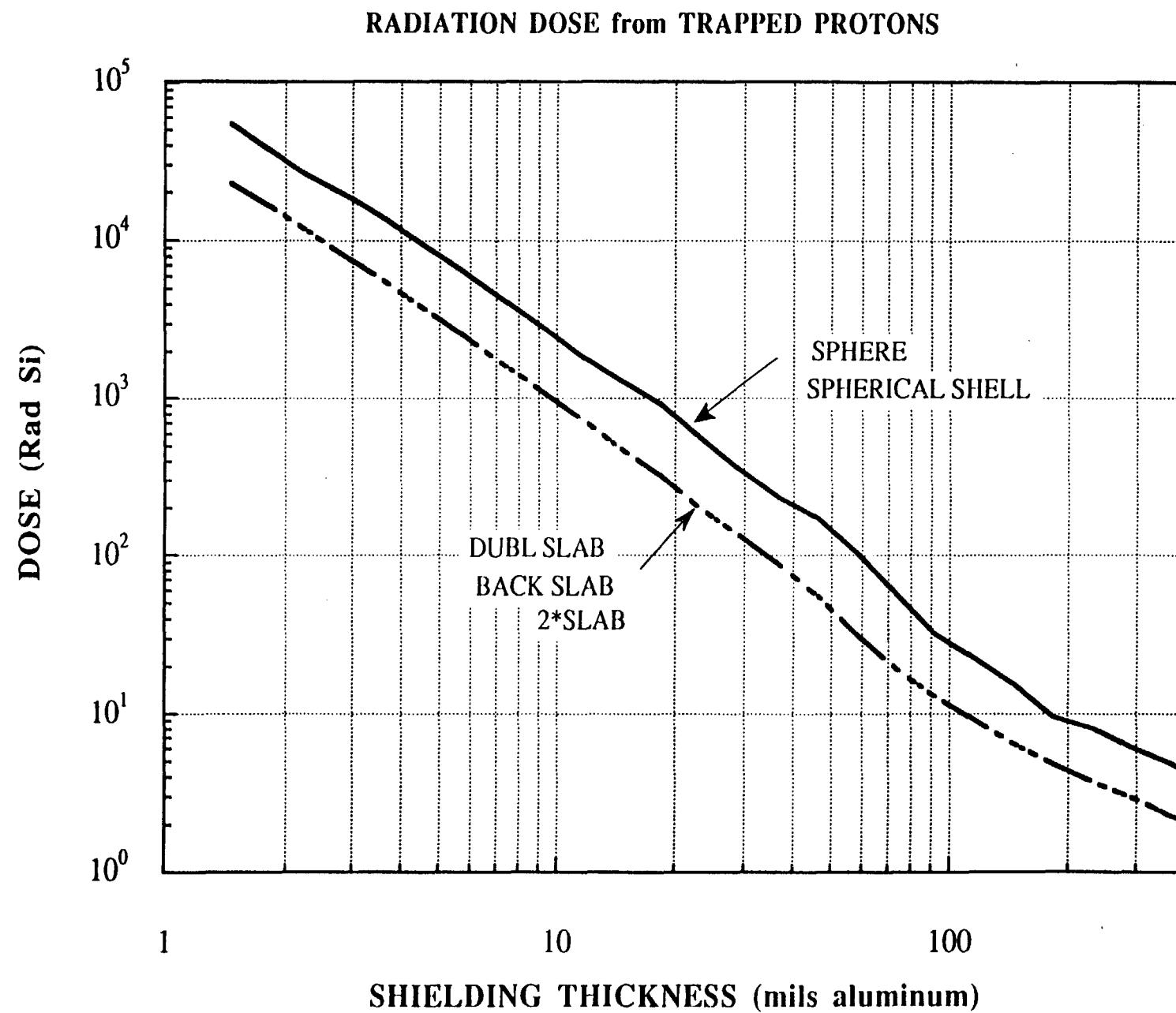


Fig. 80. Radiation dosage from the trapped electron environment (AE8) for solar-active conditions and for the first Earth-Moon Transfer Orbit {1994/01/30(07:00)--1994/02/09(12:00)}.

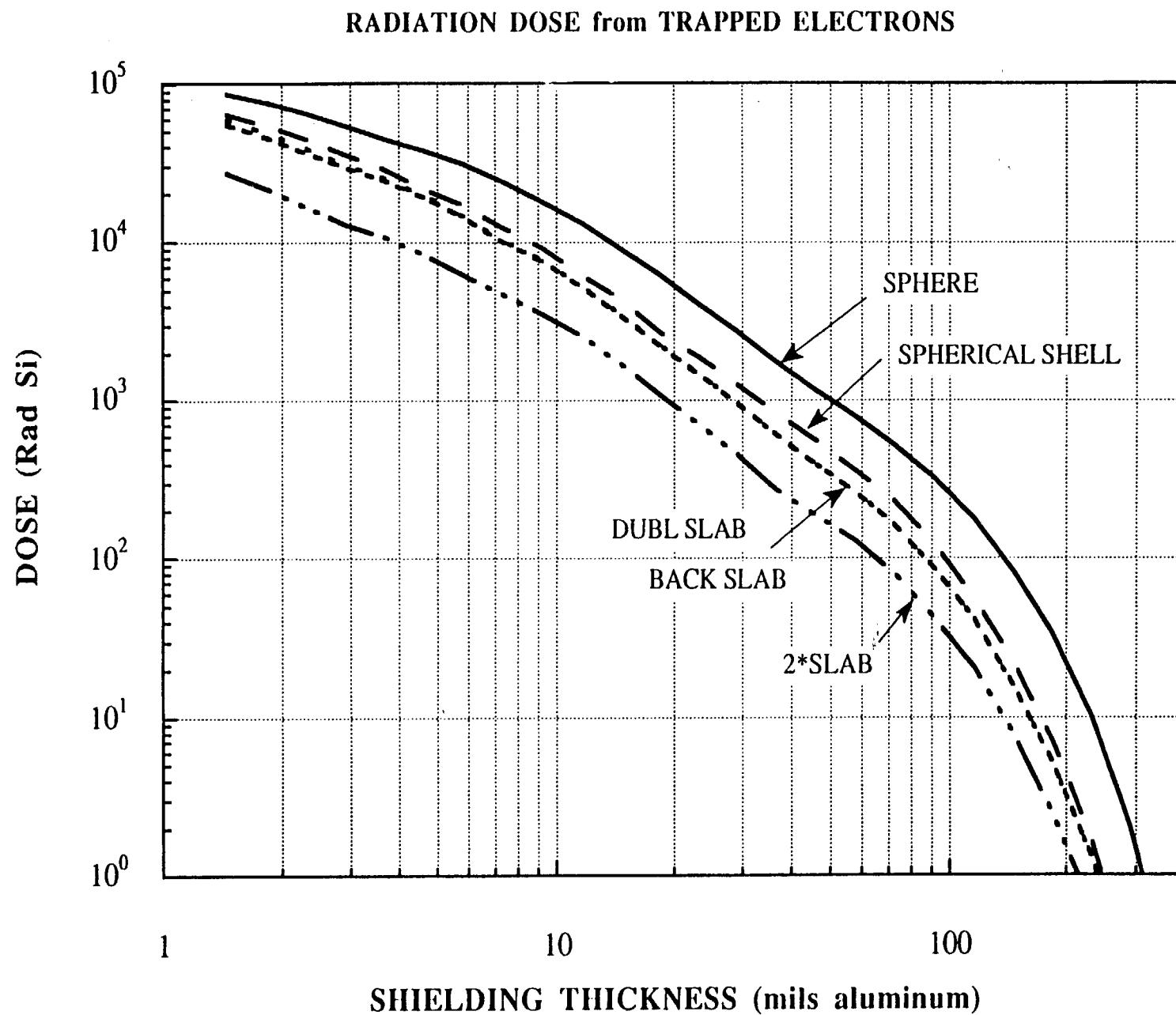


Fig. 81. Mission summary plot for the total radiation dose expected for the Clementine Interstage. This assumes a 450 day mission and a 95% confidence flare environment. Dosage is plotted as a function of aluminum shield thickness for a spherical shell geometric configuration.

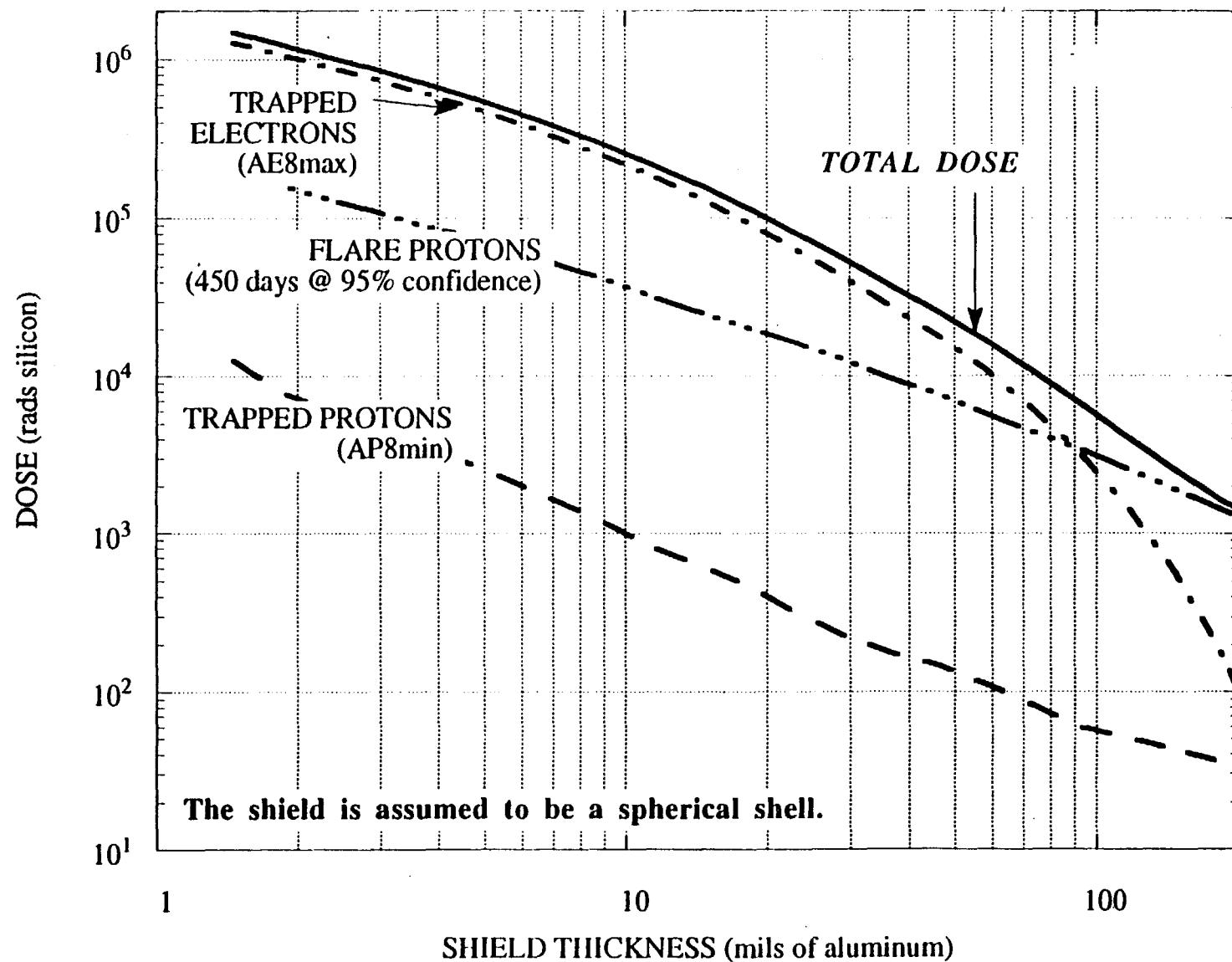


Fig. 82. Solar flare proton dosage (rad(Si)) for a 50% probable maximum dose after 1,2,3,5, and 7 years into the active portion of the solar cycle beginning on May 26, 1999. Results are for 1 AU (no magnetospheric shielding).

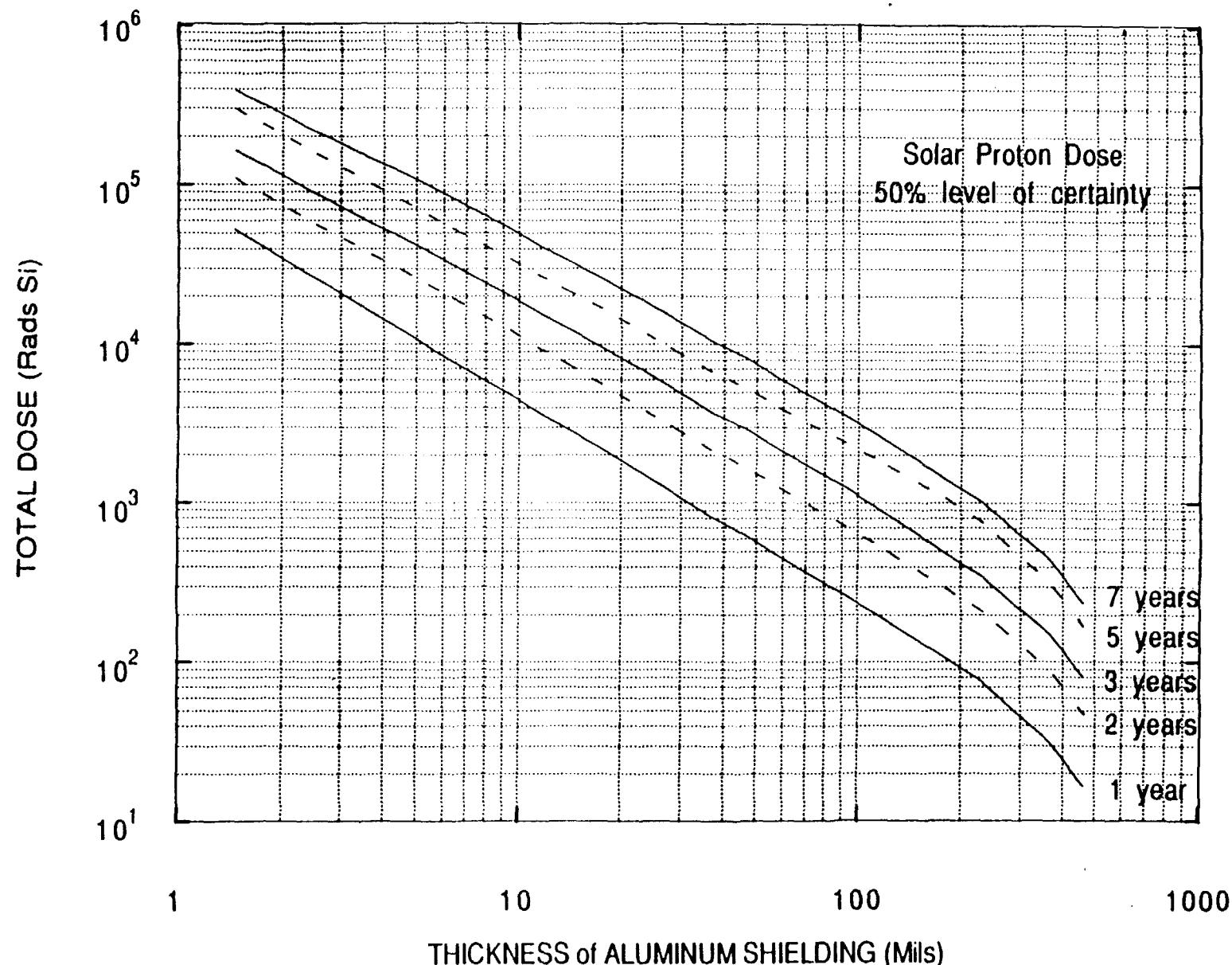
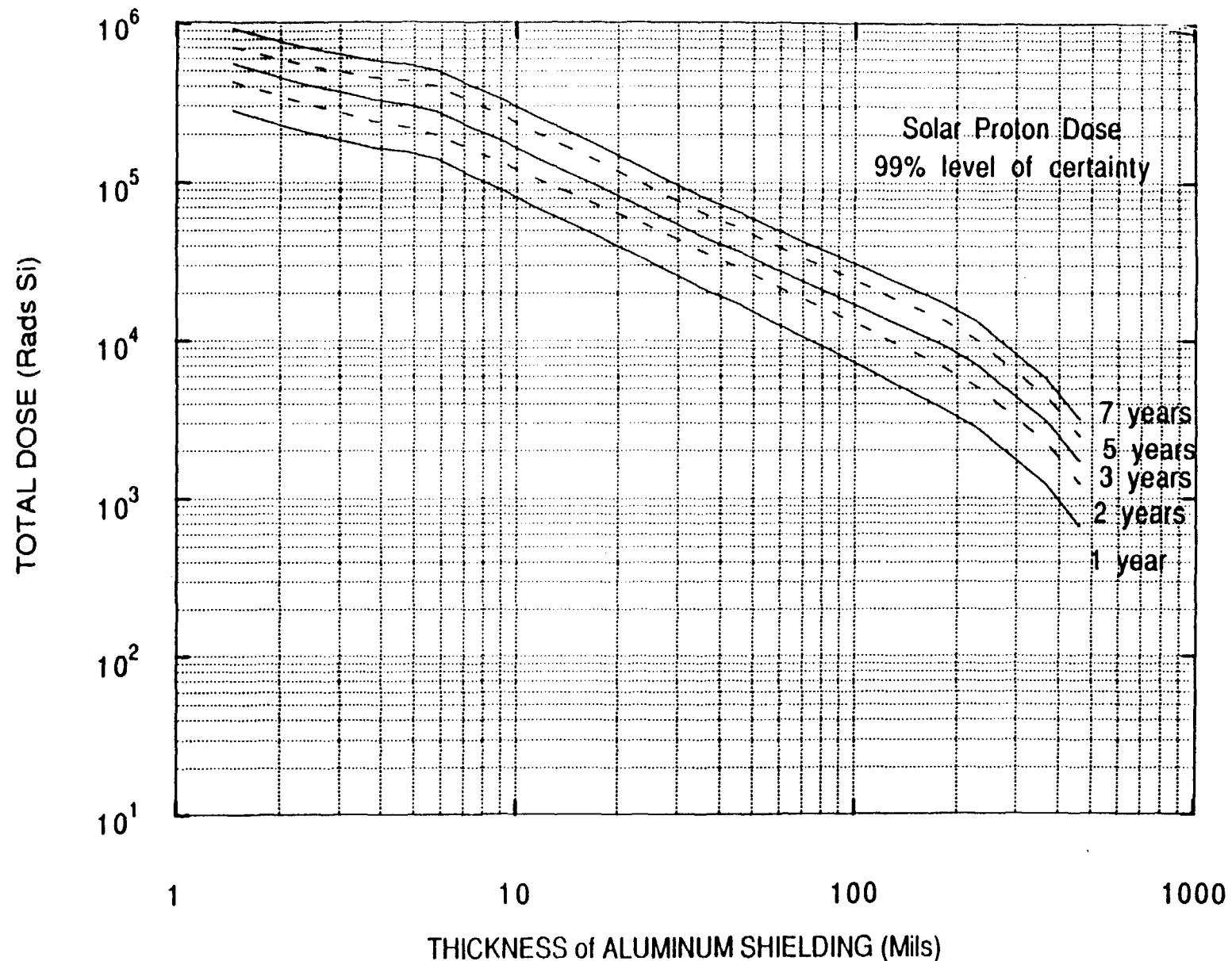


Fig. 83. Solar flare proton dosage (rad(Si)) for a 99% probable maximum dose after 1,2,3,5, and 7 years into the active portion of the solar cycle beginning on May 26, 1999. Results are for 1 AU (no magnetospheric shielding).



RANGE OF STABLE TRAPPING
LIMIT ENVIRONMENT FOR
SELECTED ORBITS

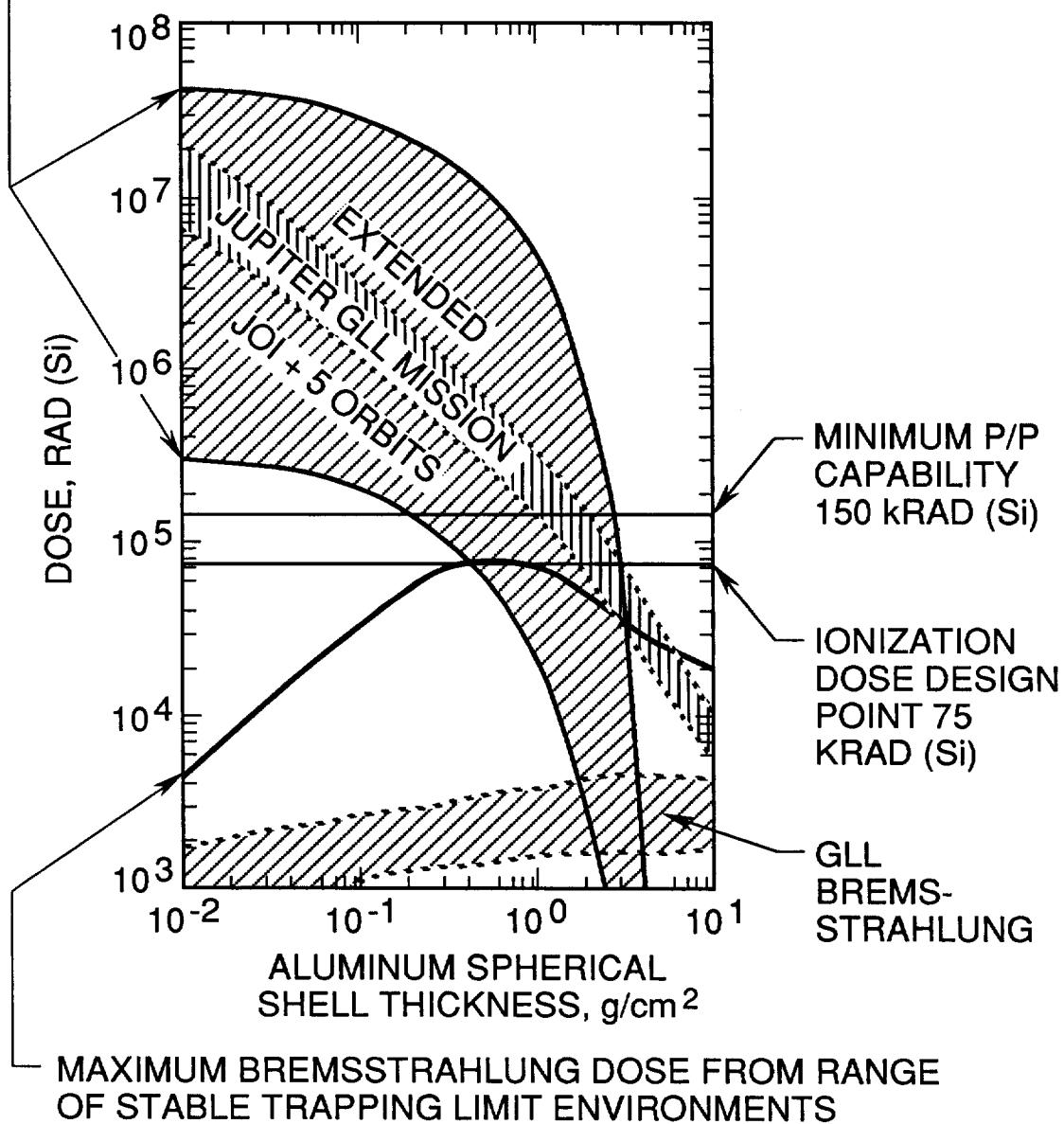


Fig. 84. Comparison between the 1 year dose from the saturated nuclear radiation environment for a range of Earth orbits and the total dose expected for the Galileo Jupiter mission after 5 orbits and after a 2 year extended mission.

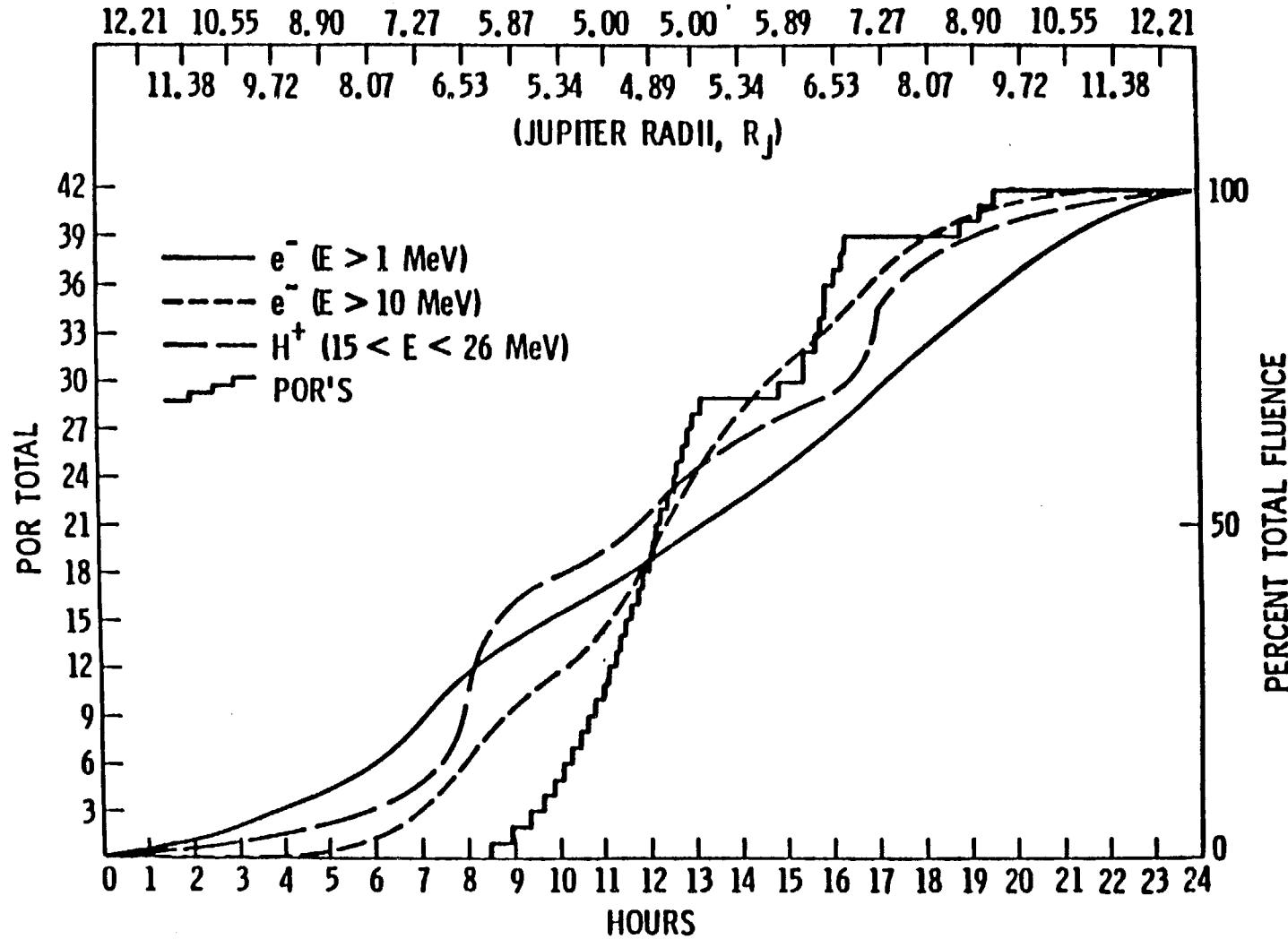


Fig. 85. Correlation of Voyager Power On Reset (POR's) Anomalies with the High Energy ($E > 1$ MeV and $E > 10$ MeV) Electron Environment.^[86] Also shown for comparison is the high energy proton environment ($15 \text{ MeV} < E < 26 \text{ MeV}$).

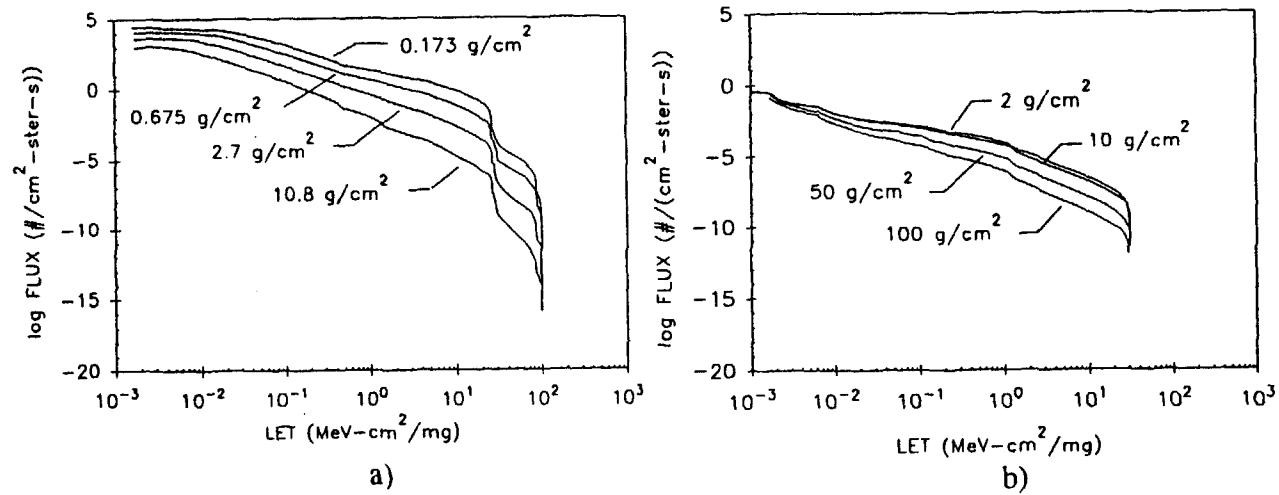


Fig. 86 Integral LET spectra as functions of spacecraft wall thickness for a) the Adam's worst case composite solar flare^[87, 93], and b) the GCR environment (elements 1 to 26)^[87, 93].

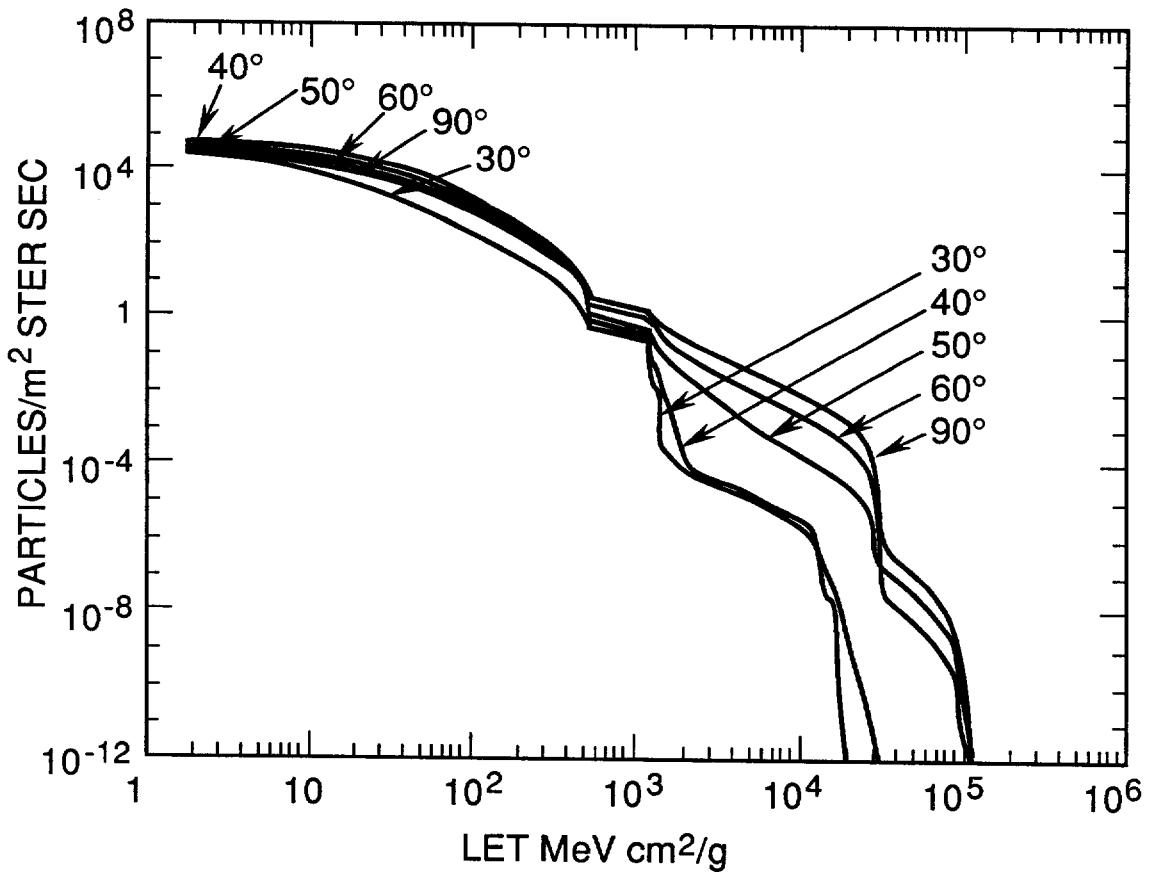


Fig. 87. Integral LET spectra inside a spacecraft (with 0.064 cm aluminum walls) in a 400 km circular orbit.[87] The 90% worst case environment is assumed in the interplanetary medium and the GSFC AP trapped proton environment at the Earth. The LET spectra are for the various orbital inclination indicated.

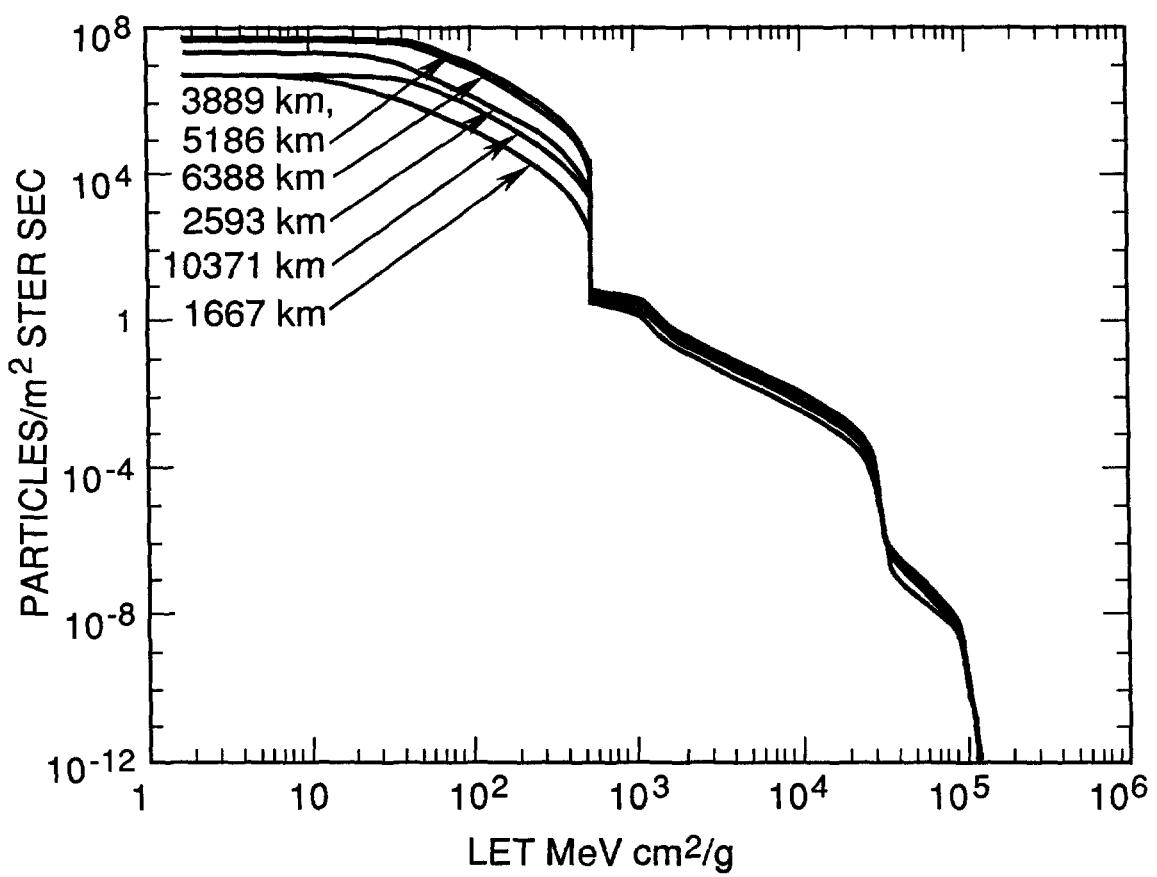


Fig. 88. Integral LET spectra inside a spacecraft with 0.064 cm aluminum walls that is in a circular orbit at a 60° inclination. [87] As in Fig. 87, the 90% worst case environment is assumed in the interplanetary medium and the GSFC AP trapped proton environment at the Earth. The LET spectra are for the various altitude as shown

Chapter 3

SEU Modeling and Prediction Techniques

Lloyd Massengill
Vanderbilt University
Nashville, TN 37235

Table of Contents

1. Introduction	III-3
1.1 Unique Payoffs of Modeling and Simulation.....	III-4
1.2 Lurking Pitfalls of Modeling and Simulation	III-5
1.3 Overview of this Document	III-6
1.4 Scope of this Document	III-7
2. Background.....	III-9
2.1 Overview of Single Particle Effects.....	III-9
2.2 Background on Space Environments.....	III-10
3. Micro-Modeling : Single-Particle Effects in IC Devices	III-15
3.1 Ionic Interaction with Semiconductor Materials	III-15
3.2 Charge Deposition/Generation	III-15
3.3 1-D/Analytical Charge Collection Models.....	III-17
3.3.1 Depletion region collection.....	III-17
3.3.2 Field-assisted funneling.....	III-18
3.3.3 Diffusion collection.....	III-22
3.3.4 Ion shunt collection	III-23
3.4 2-D/3-D Collection Models.....	III-26
3.5 Recap of Physical Modeling	III-32
4. Macro-Modeling : Circuit Errors Due to Single-Events	III-33
4.1 Circuit Model for Charge Collection.....	III-33
4.2 Circuit Response to the Single-Event Current	III-35
4.3 Destructive Effects.....	III-38
4.3.1 Latchup and Snapback.....	III-39
4.3.2 Burnout.....	III-41
4.3.3 Gate Rupture	III-41
4.4 Non-Destructive Soft Errors Defined	III-42
4.5 Upsets Models for Passive Storage Devices	III-44
4.5.1 DRAMs.....	III-44
4.5.2 RMOS SRAMs	III-47
4.6 Upset Models for Active Storage Devices	III-49
4.6.1 CMOS SRAMs	III-49
4.6.2 SOI SRAMs	III-54
4.6.3 Bipolar SRAMs.....	III-57
4.6.4 GaAs SRAMs	III-58
4.7 Transients in Combinational Logic	III-60
4.8 Other Circuit Types	III-63
4.9 Recap of Circuit Modeling.....	III-63
5. Ground-Based Experimental SEUs.....	III-65
5.1 Measured LET_C versus Simulated Q_C	III-65
5.2 Measured Cross-section	III-66
5.3 Limitations on the Interpretation of Experimental Results.....	III-66

6. Predicting Space Error Rates	III-69
6.1 Models for the Orbital/Space Environment -- $F(\lambda)$ function	III-72
6.2 Sensitive Area, Sensitive Volume, and Critical Charge	III-74
6.3 Chord-length distributions in the sensitive volume -- $P(\lambda)$ function.....	III-75
6.4 Single-Threshold Error-rate Calculation.....	III-76
6.5 Integral Error-Rate Prediction Technique	III-76
6.6 Recap of Space Error-Rate Predictions.....	III-77
7. Comparisons with Satellite Data	III-79
8. Future of Modeling	III-81
9. Summary and Conclusions.....	III-83
10. Acknowledgments.....	III-85
11. References.....	III-87

1. INTRODUCTION

Integrated circuits (ICs) destined for space deployment will certainly be exposed to several types of solar and galactic radiation. One of these radiation effects is the random bombardment of the circuit by ionizing particles of very high energy (cosmic rays), the so called **single events**. Single events can lead to seemingly randomly-appearing glitches in electronic systems -- frustrating errors which may cause anything from annoying (hopefully) system responses to catastrophic (hopefully not) system failures.

Since the postulation of electrical errors due to ionizing particles in 1962 by Wallmark and Marcus [Wal62], and the discovery by Binder et al of actual ionizing-particle-induced errors in the flip-flop circuits of communications satellites in 1975 [Bin75], interest in these single-event effects has compounded. Much of this interest has been driven by developments such as:

- the observation in 1978 by May and Woods of errors in ground-based systems due to naturally occurring alpha-particles [May79],
- the critical errors caused by cosmic ions in the Voyager and Pioneer probes and numerous satellites,
- the necessary retrofits, at great expense, of the Landsat D and Galileo systems due to heightened concern over single event upsets, and
- the recent errors in the guidance system of the Hubble space telescope as its orbit carries it through the earth's radiation belts, requiring frequent scrub and reload of the guidance system.

Because of these and other real world problems in space systems due to cosmic ions, an understanding of single-particle errors in integrated circuit (IC) electronics has become an important part of the design and qualification of IC parts for space-based use.

The issue becomes even more important as device dimensions scale, and denser, more powerful integrated systems are placed in space/satellite applications. Electronics are reaching integration levels where a single-bit of information is represented by an extremely small value of charge, and noise margins are very tight. For example, if a typical dynamic random-access memory (DRAM) cell can tolerate approximately 100mV of noise on the bit storage node with 100fF of storage capacitance, then this value of noise corresponds to only 62,500 electrons. Any perturbation of this delicate balance by

an impinging cosmic ion is intolerable. So, a recognition of, and familiarity with, the effects of space radiation on the electronics to be placed in that hostile environment is essential. Single event modeling plays a key role in the understanding of the observed-error mechanisms in existing systems, as well as the prediction of errors in newly-designed systems.

1.1 Unique Payoffs of Modeling and Simulation

Integrated circuit vendors have three avenues to address the single event susceptibility of a particular part. The part can be developed using the most advanced available design procedures and fabrication processes for mitigating single event effects, then placed in service in the hostile environment with the hopes that the conservative design and rad-hard technology effectively resists the single particle bombardment and that failure will not occur. This is probably the least desirable avenue, since it allows for no miscalculations or oversights. Once the part is in its service environment, in an orbiting system for example, redesign and replacement of failing circuits is prohibitive.

Secondly, the part can be built and subsequently extensively tested in ground-based particle accelerators which mimic the type and energy of particles expected in the service environment. Based on the survivability and upsets seen in these tests, that particular design and process technology can be verified as tolerant of the expected particles in its flight/service regime. This avenue effectively qualifies the part for service or screens specific parts to be used; however, accelerator beam time is quite expensive, and the design and implementations of the tests can be quite time consuming.

Thirdly, the circuit manufacturer can use prior experience and physical theory to develop accurate models of the response of existing circuits to the single particle environment of interest. Then, if and only if, the models are inclusive and robust enough to be predictive, the designer can apply these models during the design phase to optimize the circuit for its particular flight environment. Once the design is finished, a rough prediction of the effects of single events on the circuit can be derived.

Of course, the most common approach is a combination of the three procedures above. A designer can use a radiation-hardened process technology and rad-hard circuit design techniques, use predictive modeling to optimize his design for single-event tolerance and predict its failure level, use ground-based testing to verify radiation tolerance, and confidently place the part into service. Accurate and reliable modeling is an integral component of this design procedure.

Properly executed modeling of single-event radiation has several unique attributes. It can:

- provide insight into physical mechanisms leading to experimentally observed effects, especially if these observed effects are intertwined with others, or if the underlying mechanisms are impossible to experimentally measure
- uncover the relationship between physical parameters of the material and/or circuit and the observed response to the radiation
- provide a designer with 'what-if' results on design or processing changes without the time and expense of fabrication and experimental verification; that is, a tight feedback loop between design changes and predicted radiation vulnerability
- provide for the prediction of a circuit's response to the conditions in the field of operation before actual deployment
- identify design flaws and 'bottlenecks' in the system leading to upset vulnerability
- provide insight into the effectiveness of hardening schemes.

1.2 Lurking Pitfalls of Modeling and Simulation

Of course, the positive aspects of modeling outlined above do not come without serious constraints and unexpected pitfalls. One of the most important, yet illusive, qualities of effective modeling is the proactive recognition of limits and deficiencies of the model. All too often, models are extrapolated to an application where they no longer apply. Erroneous results then reflect badly on the original model, while the real problem is incorrect application of the model outside of its valid constraints. It is one of the intents of this document to define these bounds when discussing the subsequent modeling techniques.

A second pitfall is the completeness of the model. The model will, obviously, not account for effects which are not included in the model (the Modeler's Law). Even if the model developer believes an effect is negligible, ignoring it will only make it go away in the model, not in the actual response. One of Murphy's axioms states that this one 'negligible' effect will come back to haunt the modeler when least expected. The completeness of a model is a constant tradeoff with its useability and efficiency.

A third pitfall of modeling, which is complementary to the second, is the 'forest and trees' problem. A model which places too much emphasis on the quantitative accuracy of unimportant effects can mask the qualitative results important to the modeler. Not only is this a philosophical statement, it describes very real, practical limits also. In many cases, a model which overemphasizes secondary details can be totally prohibitive in complexity when applied to a complete system. The appropriate level of complexity and accuracy of a model is entirely dependent on the application, and the choice requires cleverness from both the model developer and the model user. This document attempts to introduce and build on a hierarchical modeling approach; using each modeling technique as an input to a higher-level model applied to a more complex system.

Finally, the development of new models, or even the application of existing models to a particular part or system, can be a time-consuming, expensive task. The rewards lie in the list of Section 1.1, not the least of which is the close link between design actions and predicted radiation consequences.

1.3 Overview of this Document

This paper presents an overview of the methods and procedures involved in computer modeling of single-event phenomena which have been proposed and utilized in recent years. The goal of these procedures is to model the interaction of a radiation environment with microelectronic circuits, and to predict the resulting influences on proper IC operation.

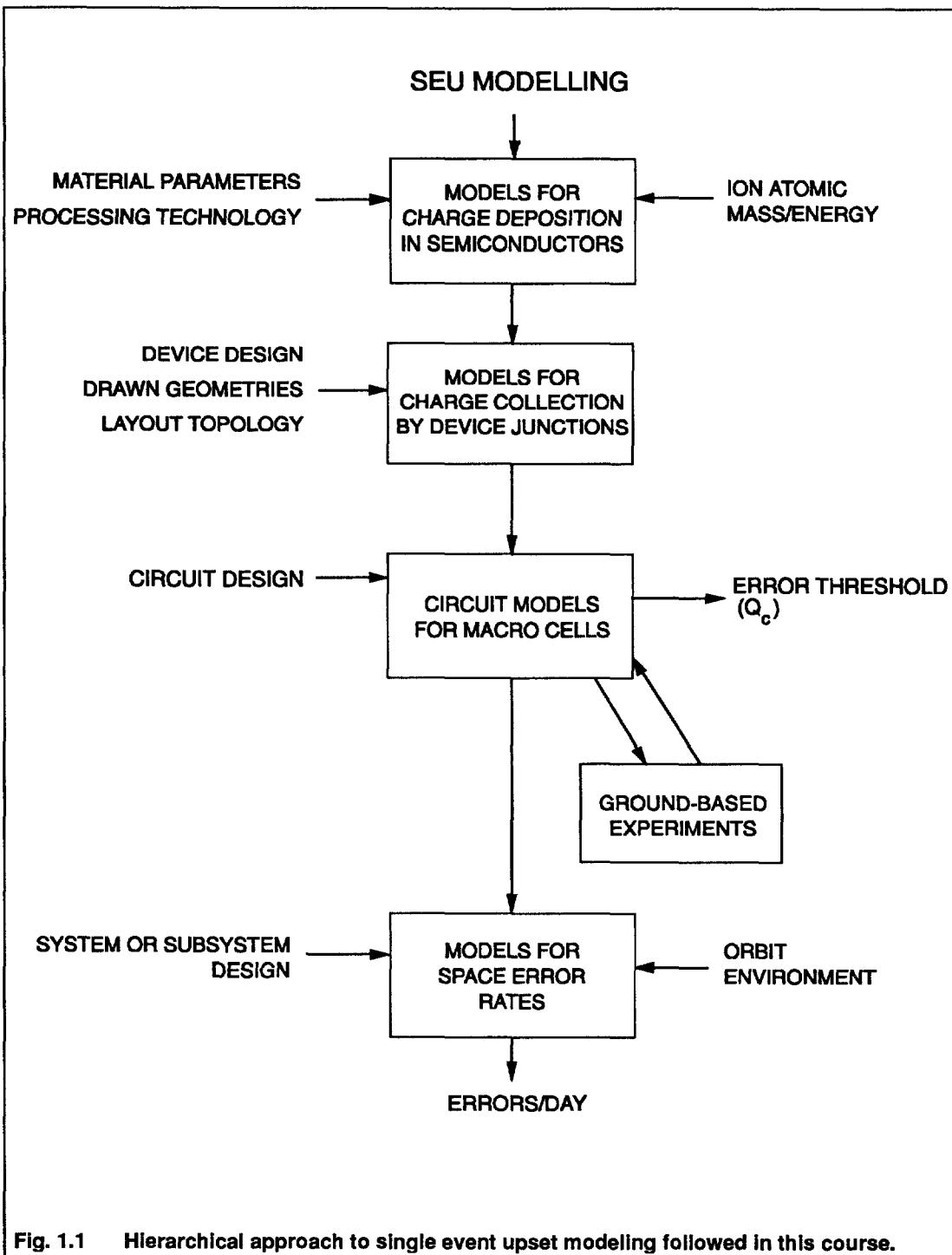
The material here is organized in a rather unique pedagogy; not from the simplest concepts to the more complex, or on a topically flat basis, but is instead organized from a microscopic view of radiation interactions with the semiconductor material up to the macroscopic view of the overall result of this interaction with the system of interest. It follows the steps the designer might follow when attacking a new problem. It begins with a description of the flux and energy of the particles expected to bombard a circuit of interest. It presents a description of the fundamental interaction of the particle with the semiconducting material of an IC and discusses physical models used to describe the charge carrier generation and motion within the devices. Next, it presents effects these carriers have on local circuit nodes, then discusses the response of the particular circuit macrocell (or subcell) to these node perturbations. From these results, analysis of the response of the total circuit to perturbations of the macrocells is presented. Finally, the IC response is used to define and quantify a usable system-level parameter which describes the vulnerability of the system to the radiation environment. **Figure 1.1**

shows this hierarchical approach followed in the course. It is hoped that this approach presents the reader with an effective view, both quantitative and qualitative, of the various levels of single-event modeling.

1.4 Scope of this Document

This document is intended to be a tutorial, covering the basic terminology and concepts of single-event upset modeling or simulation. It is hoped that it presents enough information for the reader to find the relevant literature of a specific topic of interest and study in depth. It is also hoped that it provides an overview of the basic available models and computer tools so that the reader can undertake, though not necessarily rigorously exhaust, the evaluation of an IC to single-event effects.

This work is not intended to be a review of all the past and ongoing work in SE modeling, as fascinating as that would be, because the sheer volume of material would mask the tutorial intent of this work. Therefore, it is certain that this document does not adequately recognize much of the enlightening work performed by so many groups in this area. Many interesting techniques and results, regrettably, could not be included, and the author apologizes for these omissions



2. BACKGROUND

Before tackling the details of single event modeling, we will introduce and discuss some of the terminology used throughout this text.

2.1 Overview of Single Particle Effects

When an energetic nuclear particle penetrates any semiconducting material, it loses energy through Rutherford scattering (Coulombic interactions) with the semiconductor lattice structure. Through predominately Compton interactions with the nuclei of the crystalline structure, the slowing of the particle as it transfers energy to the lattice leaves an ionization trail of free electron-hole pairs; mobile charge carriers which were electrically nonexistent before the radiation event. Within an integrated circuit structure, these excess carriers can deposit charge in unexpected and unwanted places, often leading to voltage transients on the nodes of the circuit and current transients across device junctions.

Unlike **total dose** radiation which causes gradual global degradation of device parameters and **dose-rate** radiation which causes photocurrents in every junction of a circuit, a **single event** interaction is a very localized effect, and can lead to a seemingly spontaneous transient within a region of the circuit. If this transient influences a node which is storing information, it may lead to an **upset**; that is, the corruption of the information to an unrecognizable, unreadable, or unstable state. This upset can, in turn, lead to a **circuit error** if this corrupted state alters legitimate information stored in or propagating through the circuit. That is, an upset becomes an error when it is either latched or is misinterpreted as valid data by other circuitry. The working definition of **upset** in this work is a corrupted electrical state, and an **error** is the finalized effect of that state.

Localized information errors due to single event upsets (SEUs) can be either (1) **transient**, (2) **permanent**, or (3) **static**. **Transient errors** are spurious signals which can propagate through the circuit paths during one clock cycle. These asynchronous signals can either propagate to a latch and become static, or be overwhelmed by the legitimate synchronous signals of the circuit. Timing of the radiation-induced signals relative to the synchronous signals plays a key role in the possibility of errors. These types of errors are most important in combinational (non-sequential) circuitry and analog subsystems.

Permanent errors are often called **hard errors** because of their destructive, non-correctable origins. In this case, the single event causes physical damage to the circuit, leading to a non-correctable fault. Single-event(SE)-induced burnout (SEB) and gate rupture (SEGR) in power transistors are examples of hard errors. These errors are most often analyzed and modeled at the individual device level.

Single-event soft errors (due to single-event upsets, SEUs) and **multiple-bit soft errors** (due to multiple-bit upsets, MBUs) belong to a class of errors which are **static** (latched by the circuitry) but can be corrected by outside control. These soft errors overwrite information stored by the circuit, but a rewrite or power cycle corrects or resets the part to proper operation with no permanent damage.

A special class of single-particle effects can lead to either errors (2) or (3) above, depending on the severity of the circuit response. **SE-induced snapback** (SES) in n-channel MOS output devices and **SE-induced latchup** (SEL) in CMOS structures are regenerative current conditions which, if the current levels are benign, can be reset. However, if the regenerative current energy exceed the thermal dissipation of the affected region, these effects can cause melting and permanent physical damage to the circuit.

2.2 Background on Space Environments

Integrated circuits destined for space applications may be exposed to several environments of radiation particles. For the analysis of single-event effects, we are most interested in the ionizing particles: **heavy ions** and/or **protons**. Heavy ions arise from solar flares or from deep space, thus the terms 'cosmic' or 'galactic' ions. Protons also can be generated by solar flares, and can be trapped within the geomagnetic field lines, the **van Allen radiation belts**.

A satellite in **low earth orbit** (LEO) is effectively shielded from galactic heavy ions by the magnetic field lines of the earth, the geomagnetic cutoff. Most particles do not have the energy needed to overcome the magnetic barrier, which increases in energy with lower altitudes to 2900MeV at an altitude of approximately 6×10^3 km ($L=2$) at the equator. However, this same satellite could pass through the heart of the trapped proton belt, which extends to an altitude of approximately 18×10^3 km ($L=1$ to $L=3.8$), as shown in **Figure 2.1**. Protons, because of their low mass, do not penetrate deeply through shielding and into semiconductor materials; however, at high energies, protons can produce secondary recoil reactions. The energy deposited by the proton can dislodge an ion, often silicon or

oxygen, from the lattice with a finite momentum. This ion, then, can penetrate the material as a single event. Bendel and others have developed empirical models for satellite upsets due to these proton reactions [Ben83][Ben84][Sta90] and McNulty has developed computer models for proton effects [McN85]. Here, we will concentrate on heavy ions, not overly concerned with the origins (outside the spacecraft or from a secondary proton reaction).

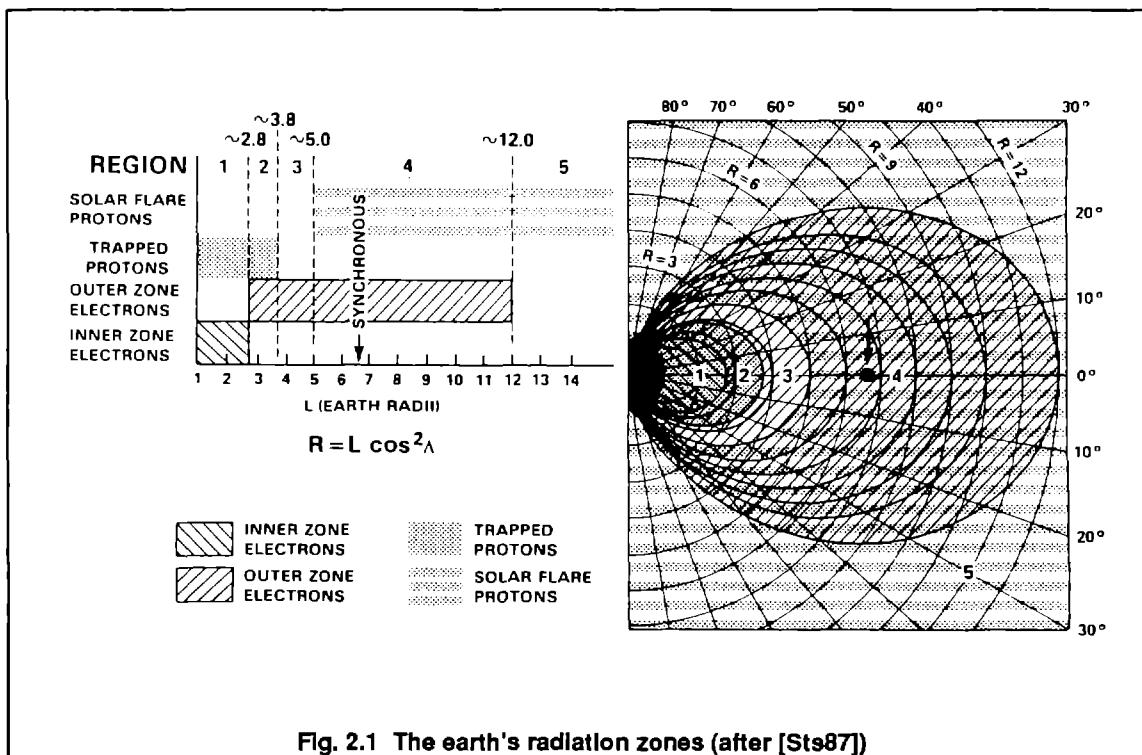


Fig. 2.1 The earth's radiation zones (after [Sts87])

Satellites in geosynchronous orbits (altitude of approximately 36×10^3 km, $L=6.6$) or polar orbits, as well as probes outside the earth's geomagnetic influences, are not exposed to the proton belts but are exposed to galactic ions and solar protons. The relative flux density of these ions range from 92.9% for protons, to 0.4% for medium nuclei with charge 6-9 (including carbon and oxygen), to 0.18% for heavy nuclei with charge greater than 10 (such as silicon, chlorine, and iron) as shown in Figure 2.2. Even though the abundance of these heavy ions is much lower than the lighter particles, the heavy ions pose a significant threat to space systems because of their penetration power through the spacecraft exterior and any shielding, as shown in Figure 2.3 [Sil87]. This work will concentrate most heavily on the circuit effects of these energetic particles.

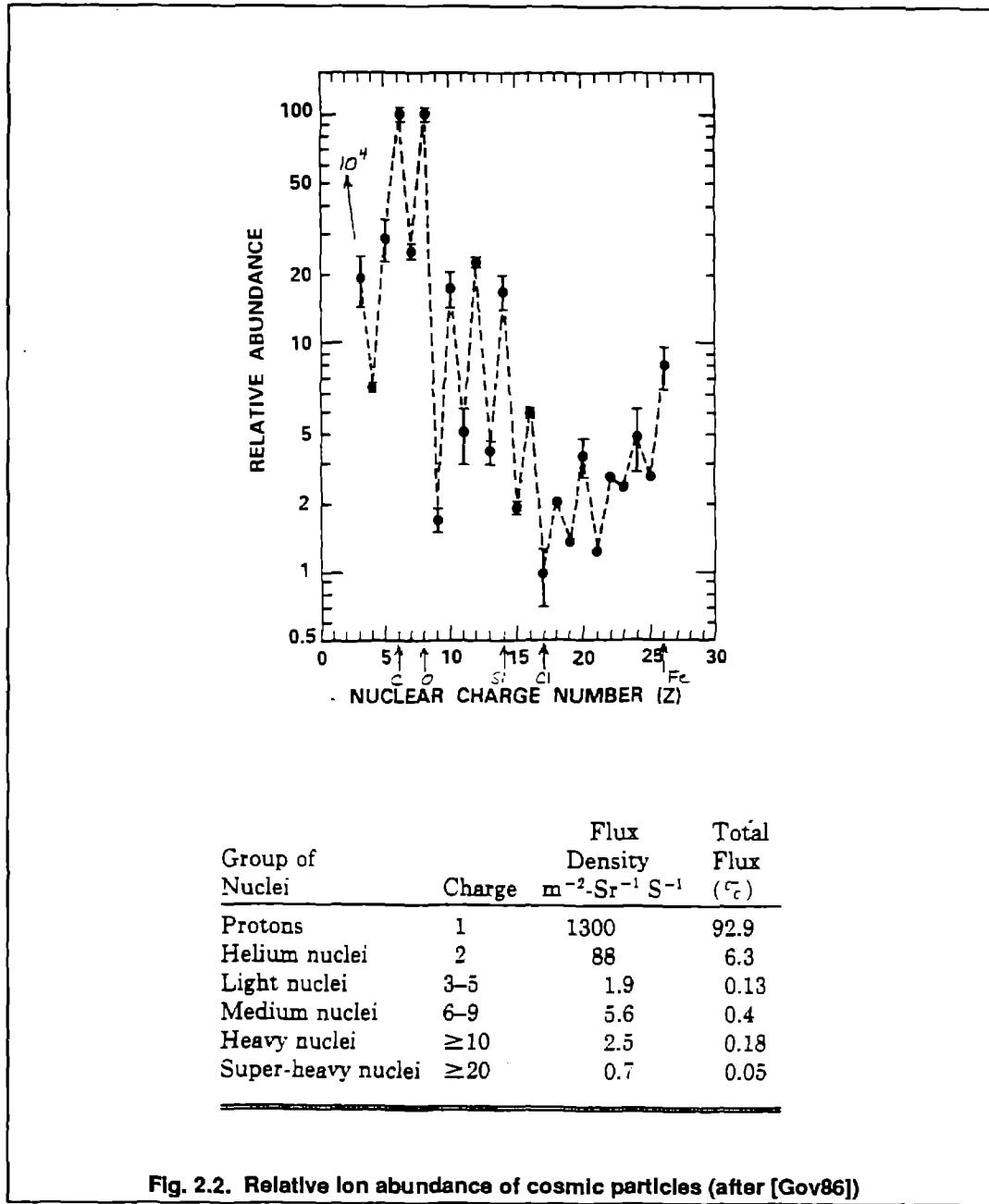


Figure 2.4 shows the regions of interest in this work and the single-event threat in those regimes. The previous short course section has described the properties and dynamics of the radiation environment, and those could not be repeated here. For the remainder of this tutorial, it is assumed that the environment for our space system is known, and the particular particles of interest can be classified. Subsequent sections will concentrate on the interaction of these particles with particular devices and circuits.

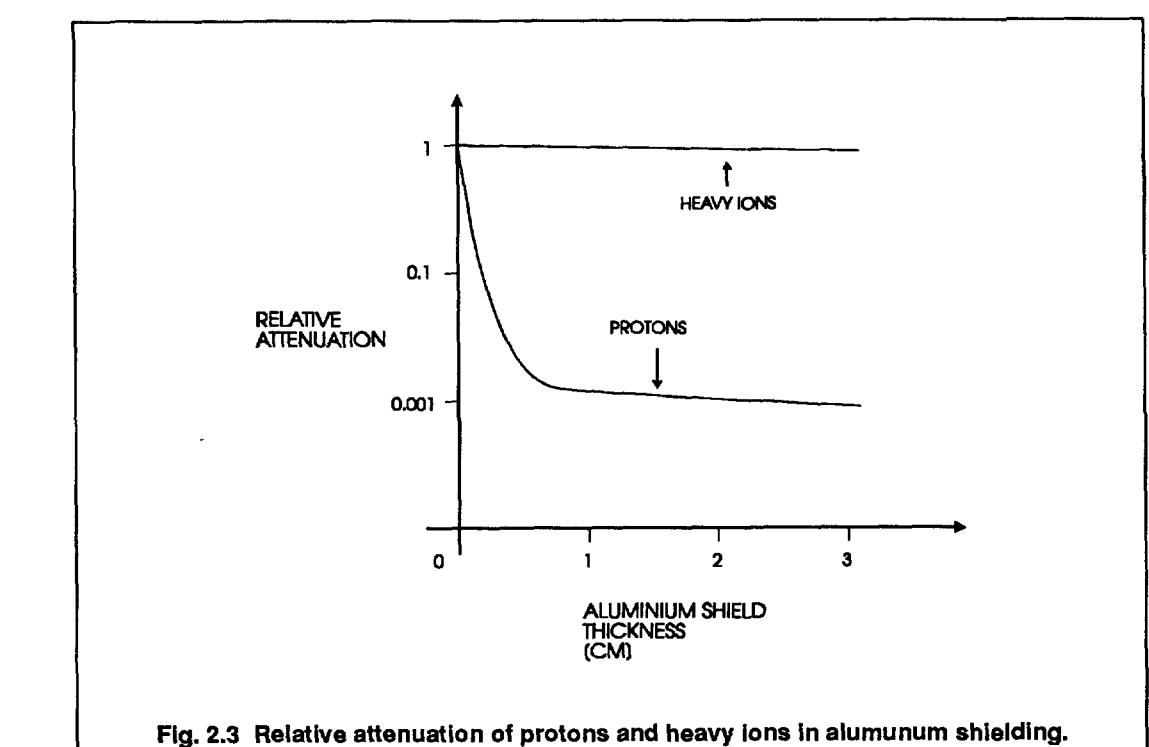


Fig. 2.3 Relative attenuation of protons and heavy ions in aluminum shielding.

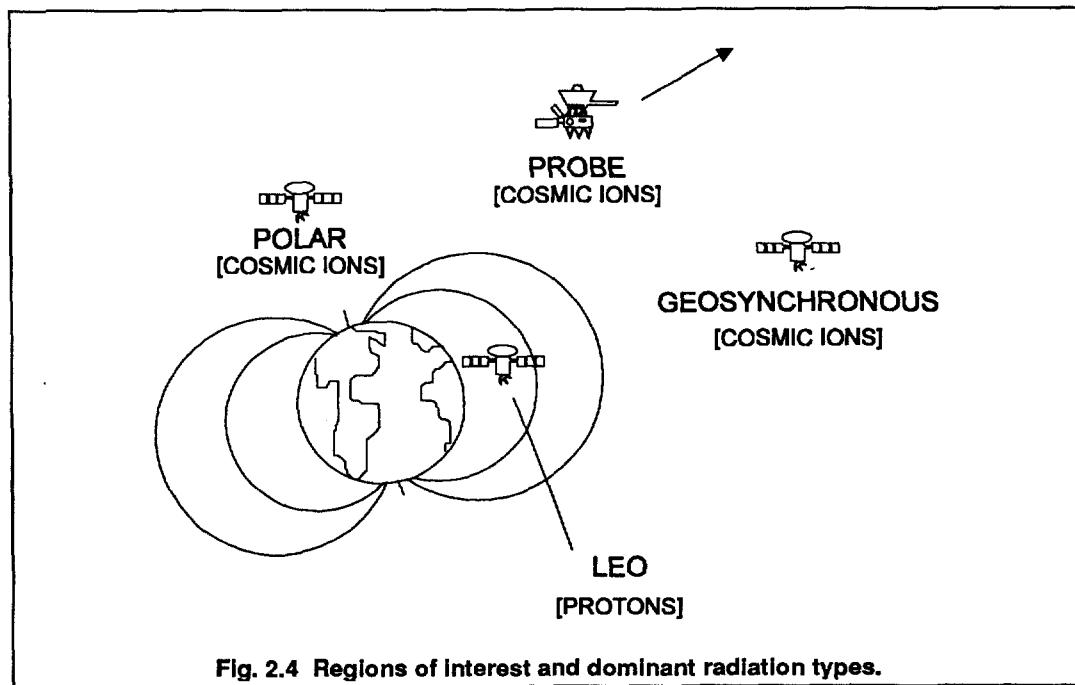


Fig. 2.4 Regions of Interest and dominant radiation types.

3. MICRO-MODELING : SINGLE-PARTICLE EFFECTS IN IC DEVICES

The first step in the simulation of single-particle effects in circuits is to adequately model the complex interaction of an ion with the semiconductor material of the chip, and the resulting charge collection by the circuit. This section is called 'micro-modeling' because it is concerned with charge carriers in the proximity of a single semiconducting junction or device.

3.1 Ionic Interaction with Semiconductor Materials

As a heavy, charged particle (ion) passes through a semiconductor material, it loses energy by Rutherford scattering with the nuclei of the lattice. Energy is transferred from the particle to bound electrons which are ionized into the conduction band, leaving a dense plasma track of electron-hole pairs (EHPs). The rate of this energy loss to EHP creation, often expressed as **stopping power** or **linear energy transfer (LET)**, has the dimensions of energy per unit length along the path of the particle. The LET of any particular particle depends on the mass and energy of the particle, as well as the density of the target material. Thus, units of LET are usually MeV/gm/cm², or converted to MeV/ μ m in a specific target material. Stopping powers and ranges for various ion species and energies have been tabulated, and are available in [Zie85]; or can be calculated with Ziegler's TRIM computer code.

3.2 Charge Deposition/Generation

3.6 electron Volts (eV) are needed to create one electron-hole pair in silicon; 4.6 eV are needed in gallium-arsenide. If one has a constant LET along the path and the density of the target material is known, the amount of charge created by ionization (i.e. the number of carriers \times the electronic charge) can be easily calculated. However, as a particle loses energy on its flight through the target material, it slows and the LET changes; LET is not a constant along the path, but is a function of distance into the material. Along an incremental length, the charge creation by the ion is given by:

$$dQ[pC] = L(x)[pC/\mu m] \cdot dX[\mu m] = \frac{LET(x)[MeV/g/cm^2] \cdot \rho[g/cm^2] \cdot 1.6 \times 10^{-5}}{G[eV]} \cdot dX[\mu m] \quad (1)$$

where the units of each variable is given in the brackets and:

dQ is the differential charge ionized in the incremental length,
 $L(x)$ is the linear charge deposition,
 dx is the differential (incremental) length,
LET is the linear energy transfer of the particle at the incremental point along the path,
 ρ is the density of the target material, and
 G is the electron-hole generation energy for the target material (3.6eV for Si and 4.6eV for GaAs).

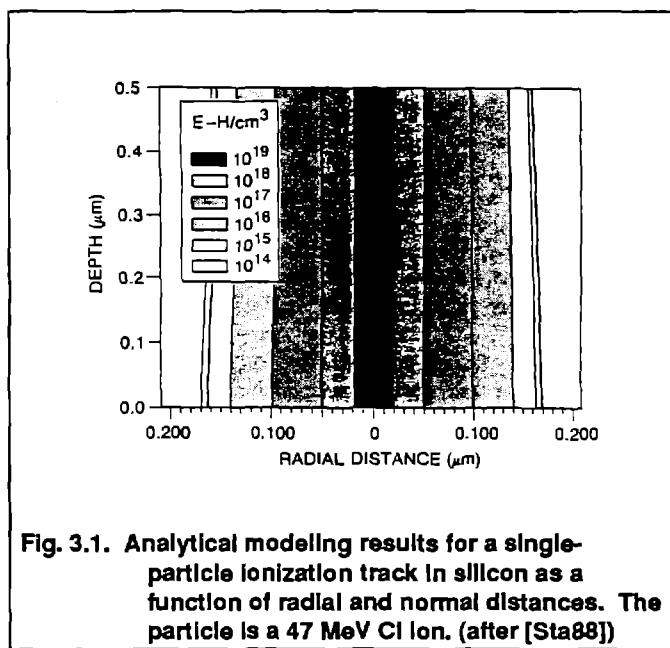
For Si and GaAs, these conversions are, respectively:

$$L(\text{Si}) [\text{pC}/\mu\text{m}] = 1.035 \times 10^{-5} * \text{LET} [\text{MeV/g/cm}^2] \quad (2)$$

$$L(\text{GaAs}) [\text{pC}/\mu\text{m}] = 1.850 \times 10^{-5} * \text{LET} [\text{MeV/g/cm}^2] \quad (3)$$

To robustly model the carrier generation in a circuit due to an impinging ion, this differential must be evaluated iteratively along the ion path. Codes have been developed to perform these tasks; these include the public-domain SEFEC (N.C. State University) and several commercial packages.

Recently, Stapor et al have presented an analytical method of applying delta rays in cylindrical coordinates to distribute the energy lost by an incident particle [Sta88]. This technique simulates the radial distribution of generated charge as a function of track position for any particle of any energy. A typical result of this modeling is shown in Figure 3.1. This figure gives a good example of the plasma track structure in relation to typical IC device dimensions. The radius of the high-density portion of the track is on the order of a few hundred angstroms, much smaller than typical device topologies. Also, the variation of charge deposition along the length of the track in the 0.5 μm depth shown is minimal. In very thin regions, which are typical of modern integrated circuits, it is often acceptable to assume a constant LET along the region of interest.

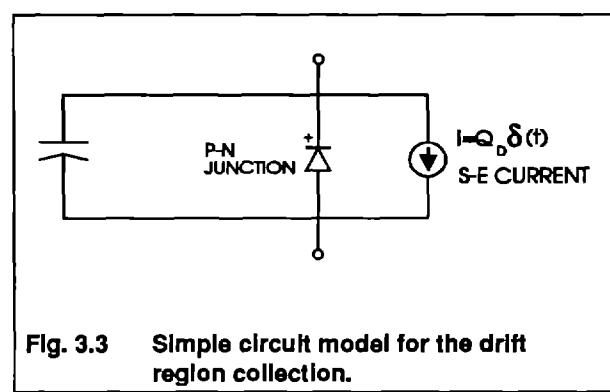
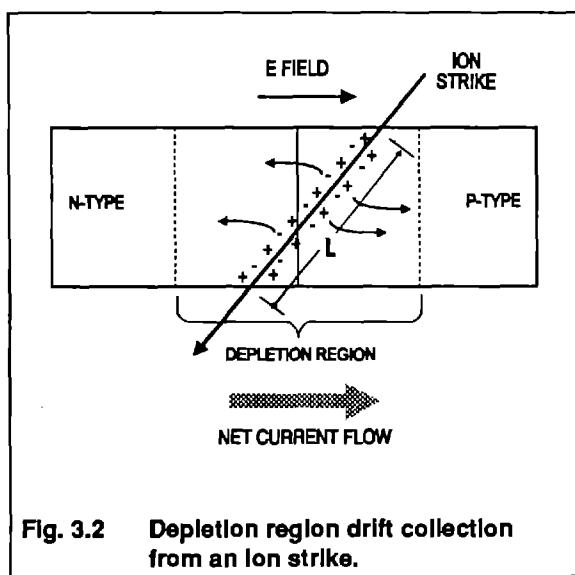


3.3 1-D/Analytical Charge Collection Models

Charge deposition, as described in the previous section, in a bulk semiconductor region is of no consequence; it will eventually recombine. If, however, that charge is deposited in or near a p-n junction, the electrons and holes will be separated, leading to charge flow and a photocurrent generation. This charge collection by the IC junctions leads to the circuit response to the single event penetration. There are several mechanisms for this collection and conduction process including depletion-region drift collection, field-assisted funneling collection, diffusion collection, and ion-shunt collection. It should be noted that charge collection and charge generation are distinct processes, and they must be modeled as such.

3.3.1 Depletion region collection

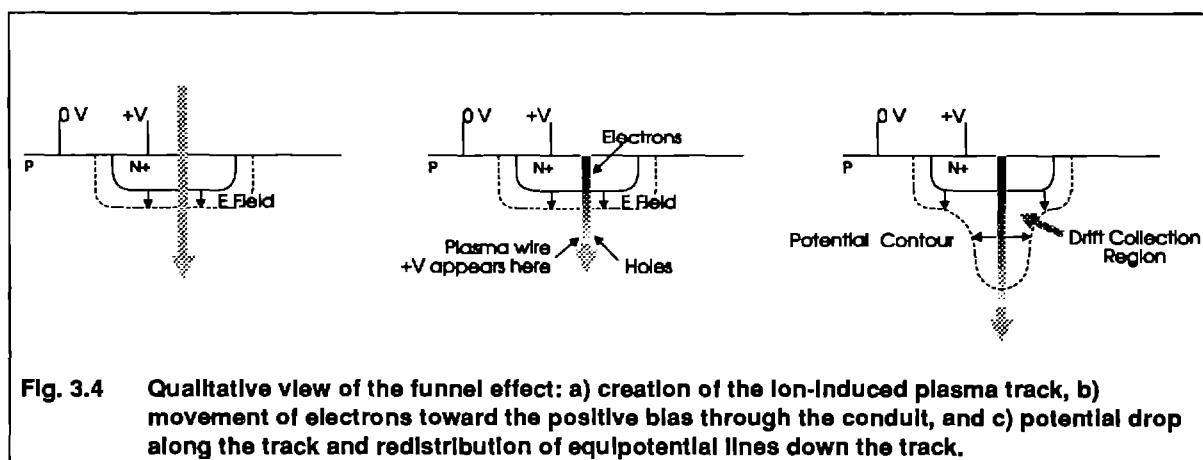
Figure 3.2 shows an ion penetrating the depletion region of a p-n junction. The built-in electric field causes electrons to be swept to the n-region and holes to the p-region. This drift motion is limited by the saturation velocity of the carriers, which for electrons in silicon is approximately 1×10^7 cm/sec, so the time period of this transient is very short. For example, the speed of electron drift across a $0.5\mu\text{m}$ depletion region which is heavily reverse biased can be approximated by 0.5×10^{-4} cm / 1×10^7 cm/sec = 5 ps. The simplest model for this extremely fast transient would be a current impulse across the junction with area equal to the total amount of charge deposited in the depletion region, i.e. $I_G = Q_D \delta$. Figure 3.3 shows the equivalent circuit for this model.



3.3.2 Field-assisted funneling

Early on in the investigations of ionizing particle effects on junctions, it was recognized by Hsieh et. al. [Hsi81] that the creation of a highly-concentrated free carrier track (a plasma track) within a junction depletion region perturbs the region itself, so that the simplified depletion current calculation presented above does not adequately describe the actual charge collection. In fact, this discovery was accomplished by insightful computer simulation of microscopic carrier motions based on the semiconductor carrier equations [Hsi83]. This discovery is an example of one of the unique advantages of modeling discussed in the introduction; that is, the investigation of processes at a level which is not experimentally plausible, leading to a new fundamental understanding of the experimentally observed characteristics.

Hsieh showed that the generated carrier track from an alpha-particle penetration of a junction severely distorts the potential gradients along the track length, creating a field funnel. Figure 3.4 shows a qualitative schematic of the mechanism leading to the creation of the field funnel. Upon the creation of the plasma track, a path of free carriers appears between the n and p regions (Fig. 3.4(a)). The depletion region is effectively hidden from the carriers, which are free to move toward (electrons) or away from (holes) the positively-biased n region (Fig. 3.4(b)). The spreading resistance along this plasma 'wire' leads to a voltage drop along the length of the track. Thus, the potential which initially appeared across the depletion region is distributed down the track (Fig. 3.4(c)). Carriers outside the original depletion region will be accelerated by this field and be collected promptly at the junction by a drift mechanism, the funneling effect.



The charge flow across a junction due to funneling is a complex, three-dimensional, microscopic process affected by material parameters of the junction, the geometry and location of the hit, the time profile of the thermalization of the free carriers and subsequent ambipolar diffusion out from the track center, and the creation/collapse processes of the field funnel. However, a straightforward, one-dimensional, analytical model was developed by Hu to describe the charge flow (current) by field funneling [Hu82]. This model can be used as a simple approximation of the pulse generated by a SE hit on a junction (Figure 3.5) for macroscopic circuit simulations.

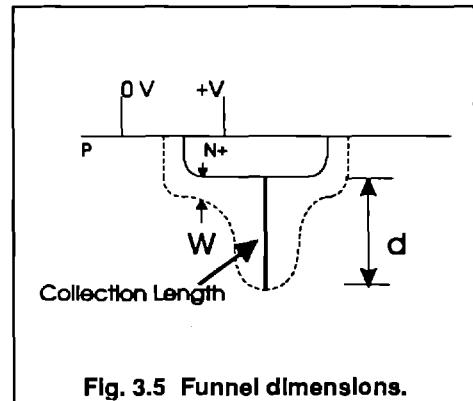


Fig. 3.5 Funnel dimensions.

Hu predicted that the depth of funnel collection in an n+/p junction below the junction edge (see Figure 3.5) is given by:

$$d = \left(1 + \frac{\mu_n}{\mu_p}\right)W \quad (4)$$

where μ_n and μ_p [$\text{cm}^2/\text{V}\cdot\text{sec}$] are the n and p type carrier mobilities, respectively, and W is the depletion region width *after* the funnel collapses. The initial current flow across the junction predicted by the Hu model is given by

$$I_O = (V + V_{BI})G \quad (5)$$

where V [Volts] is the junction bias, V_{BI} [V] is the built-in junction potential, and G [mhos] is the conductance of the plasma track created by the single event. The conductance G is given by

$$G = \left[\frac{d}{q(\mu_n + \mu_p)L} + \frac{1}{2q\mu_p \sqrt{2\pi N_a L}} \right] \quad (6)$$

where d [cm] is the collection depth below the junction (Eq. 4), q [C] is the electronic charge, L [cm^{-1}] is the linear charge density along the track (defined in Sec. 3.2), and N_a [cm^{-3}] is the p-type doping level.

Hu's time dependent current pulse is given by:

$$I(t) = \frac{I_O}{\cosh^2 \left[\left(\sqrt{\frac{(V + V_{BI})N_a}{2q}} + \frac{\mu_p G \cos \theta}{(\mu_n + \mu_p)L} \right) \cdot t \right]} \quad (7)$$

where θ is the angle of particle incidence from the normal and t is time. The Hu model can also be applied to p+/n junctions with the appropriate switching of hole and electron mobilities and doping levels.

The integral over time of Eq 7 yields the total SE charge collection by the junction (Q_{coll}). This total charge is a useful parameter used in later circuit-level macro-modeling.

As with any analytical model, Hu's model makes several assumptions in order to keep the solution tractable. These assumption include a very thin, highly dense column-shaped track structure, neglecting carrier currents outside the column, a one-sided collection dominated by the lower-doped side of the junction, and a constant G (conductance) of the track independent of time.

The model of Eq. 7 produces a time dependent current across the struck junction with an instantaneous rise time and a fall time shape governed by the \cosh^2 function. Messenger developed a model for the SE current pulse as a double exponential given by [Mes82]

$$I(t) = I_0 [e^{-\alpha t} - e^{-\beta t}] \quad (8)$$

where α [sec^{-1}] is the time constant of charge collection from the funnel (similar to Hu's model above) and β [sec^{-1}] is the time constant for the initial formation of the funnel region. This type of SE current pulse is shown in Figure 3.6. The double exponential form of the SE charge collection is the most common form used in circuit-level simulations. It is relatively easy to implement the current source of (8) in a circuit simulation code such as SPICE; with the values of α and β empirically derived.

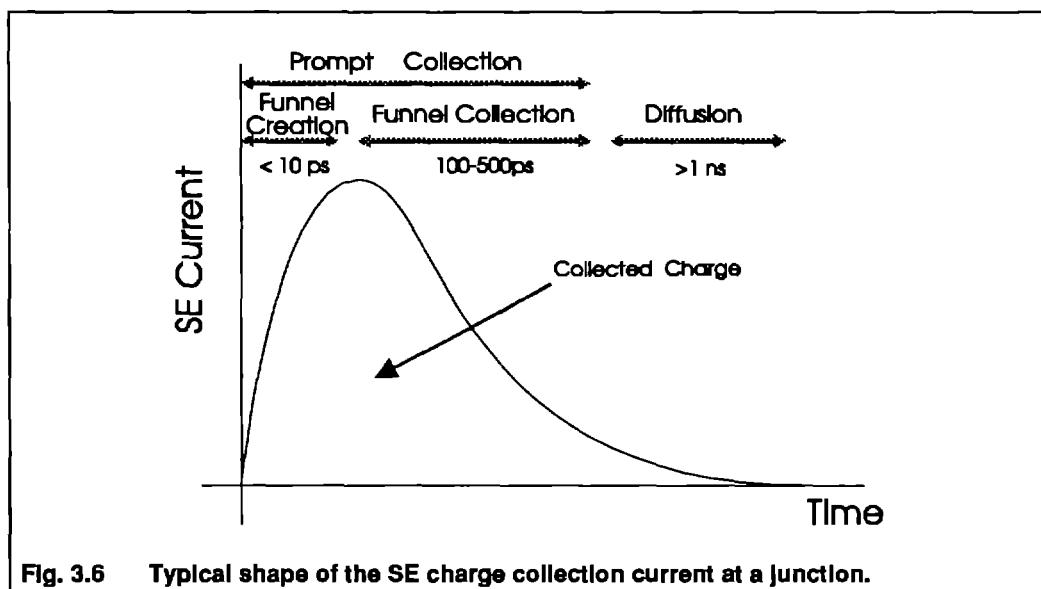


Fig. 3.6 Typical shape of the SE charge collection current at a junction.

McLean and Oldham developed an analytical model of the funnel effect [McL82], approaching the problem on a total charge basis rather than current motion. The model assumes one can represent the temporal and spatial history of the funnel field by an effective field related to the relaxed depletion region field after the event, and leads to the following equation for collected charge:

$$Q = q\bar{N}_o \sqrt{\mu_n V} \left[\frac{3N_o}{8\pi N_a v_p D^{1/2}} \right]^{1/3} \quad (9a)$$

and the collection time is:

$$\tau_c = \left[\frac{3N_o}{8\pi N_a v_p D^{1/2}} \right]^{2/3} \quad (9b)$$

where D [cm^2/sec] is the ambipolar diffusion constant, v_p [cm/sec] is the escape velocity for holes, \bar{N}_o [$1/\text{cm}$] is the averaged charge line density along the track, and N_0 [$1/\text{cm}$] is the line density near the surface.

Assumptions contained in the model of Eq. 9 include: the radius of the charge column is given by the ambipolar diffusion length; charge separation occurs along the outer edges of the column due to hole movement radially out to the substrate by the field lines created by the funnel potential gradient from the column to the substrate; the field near the surface is the same as the initial depletion layer field; the electrons trapped in the center of the plasma column by the radial field lines are swept out to the electrode due to the longitudinal bias; and most importantly, the collection length (funnel length) is shorter than the total track length.

This model overpredicts the average funnel field (and thus the collected charge) for very short, dense particle tracks, that is ions with high LET where the track length is less than the collection (funnel) length [Old83]. Oldham et al have since revised the funnel model to account for this discrepancy at high LET [Old86]. The modification makes the assumption that most of the potential is dropped across the end of the track, so the actual effective funnel field is lower than predicted by the previous model. This correction is in the form of an empirical modification of the applied potential given by:

$$V_{eff} = V e^{-KN_o} \quad (9c)$$

where V_{eff} is the new bias used in Eq. 9a and K [cm] is a screening constant determined from experimental measurements. For most particles, the value of K has been found to be $K=1.3 \times 10^{-10}$ cm in silicon and $K=4.0 \times 10^{-10}$ cm in GaAs.

Eq. 9 provides a modeler with the total charge expected to be collected at a junction due to a particular ion hit. This charge can be delivered to the circuit as an impulse current or a more realistic exponential model, as in Eq. 8.

3.3.3 Diffusion collection

Charge generated outside the funnel region, but within a diffusion length of a junction, diffuses to the junction and can be swept across the depletion region, leading to another current mechanism [Kir79]. This collecting junction may be the hit junction, or an innocent neighboring region, as shown in Figure 3.7. Diffusion is a much slower process, so this current component is delayed with respect to the field-assisted collection current. Typical time domains are nanoseconds for diffusion collection.

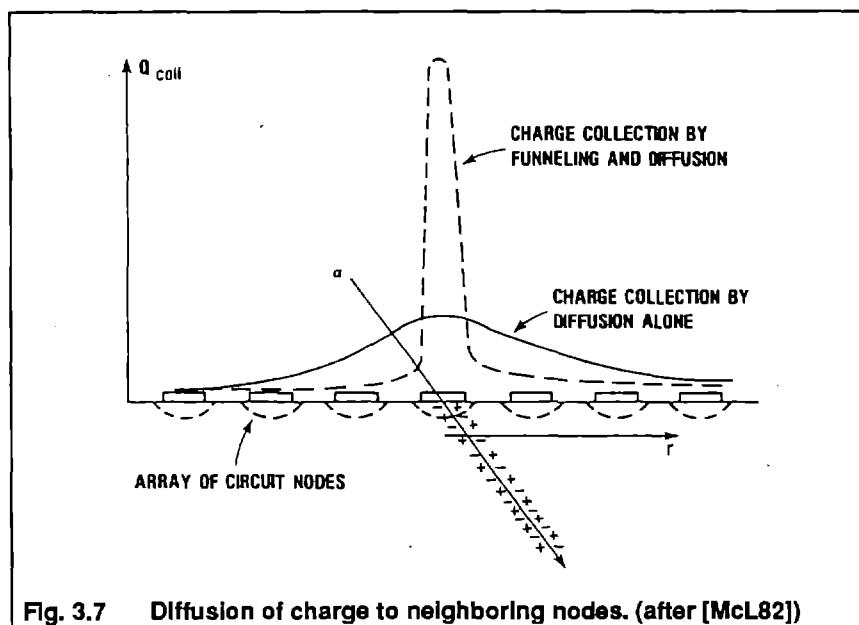


Fig. 3.7 Diffusion of charge to neighboring nodes. (after [McL82])

Because of the complex, three-dimensional nature of the diffusion charge transport, charge collection by this method is very dependent on the geometry circuit layout and the distances from the hit location to nodes in the vicinity. Most modeling of this effect is performed using 2-d or 3-d finite-element charge transport codes, as described in Section 3.4.

An example of the results of this type of modeling is shown in Figure 3.8 [Zou88a]. In this figure, a 2-dimensional finite-element simulation clearly shows the lateral diffusion of charge (electrons)

away from the track center following the ion strike. If another junction were in very close proximity to the simulated junction, this diffusion charge could be collected (at a much slower rate) by the neighboring junction. Diffusion is most important in dynamic random-access memories (DRAMs), where sensitive nodes are packed very densely and the circuit can integrate the collected charge over long periods of time.

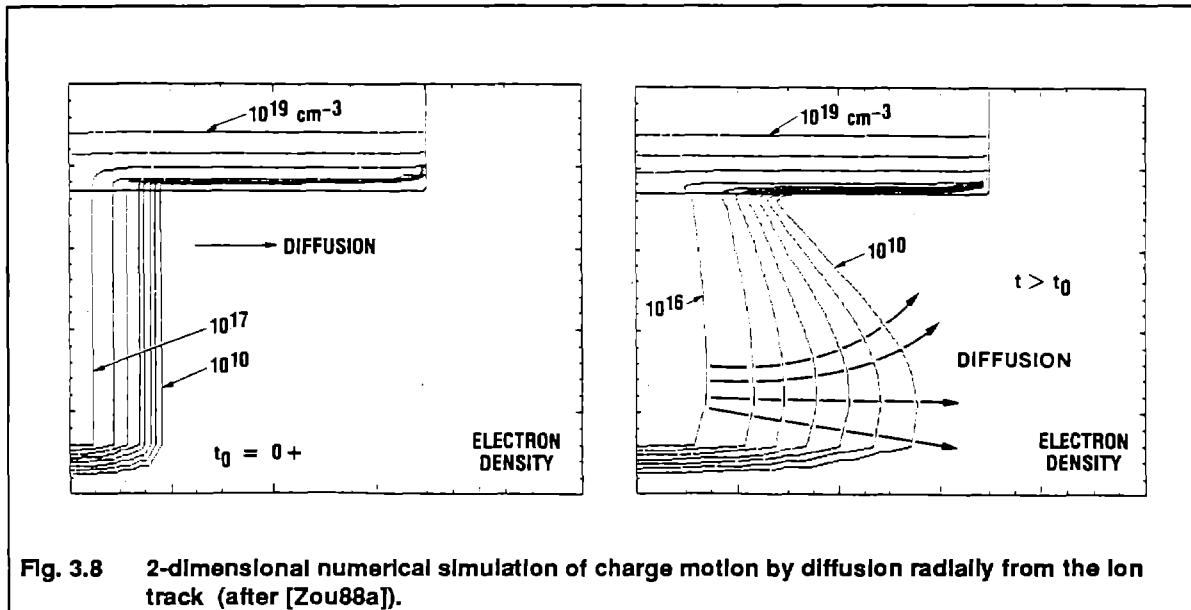


Fig. 3.8 2-dimensional numerical simulation of charge motion by diffusion radially from the ion track (after [Zou88a]).

3.3.4 Ion shunt collection

Figure 3.9 shows an unique and interesting phenomenon which has become increasingly important in modern, dense integrated circuits where geometries are scaled to small dimensions. In this figure, an ion track of free carriers has penetrated two proximal junctions. Since the plasma track, while it

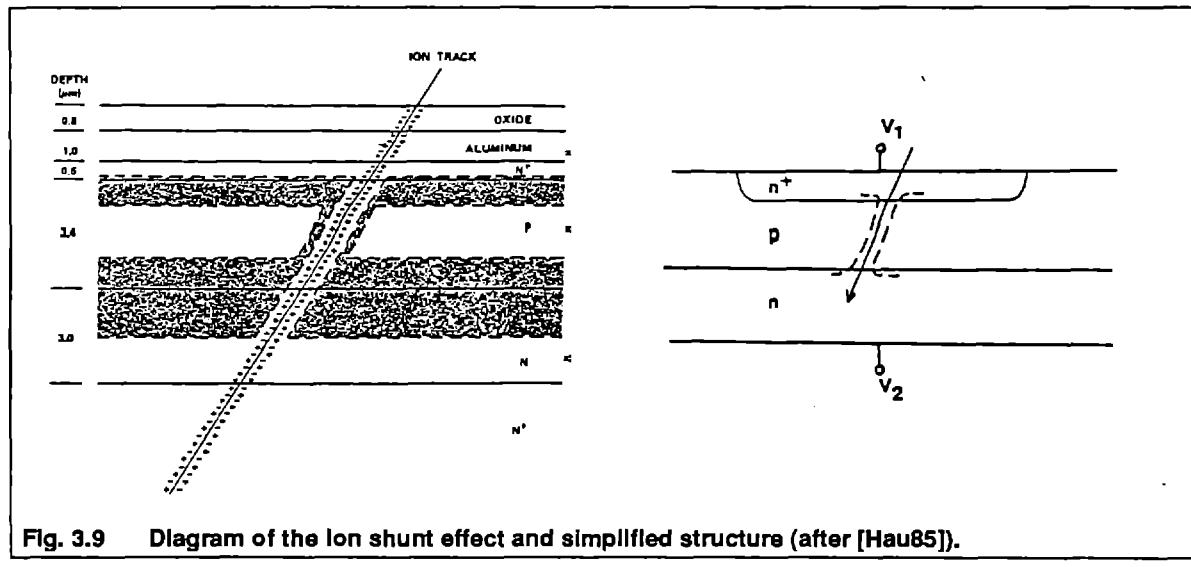


Fig. 3.9 Diagram of the ion shunt effect and simplified structure (after [Hau85]).

exists with a high carrier concentration, acts as a conductive path; the path between the two n regions in the figure can act as a current conduit, or resistive connection, between the regions. Charge which was not even generated by the ion hit can move through this conduit just as current through a wire.

Knudson et al experimentally observed more current at collecting nodes in multilayer structures than could be accounted for by the total ionized charge [Knu84]. The reason for this extra current is conduction through the conduit from the different biases applied to the top and bottom n regions. This connection of two regions by the ion's plasma track is called an **ion shunt**. Of course, in modeling dense integrated circuits for SE effects, the inclusion of any possible ion shunts is essential. This is especially important in dense CMOS on thin epitaxial layers and bipolar circuit technologies. Bipolar transistors are fabricated with very thin base regions, making them highly susceptible to the shunt effect. Zoutendyk and coworkers have published extensive experiment and modeling work on bipolar charge collection effects, including the shunt effect [Zou83][Zou88b], and the reader is referred to these citations.

Hauser et al have developed an analytical model for the ion shunt based on physical first principles [Hau85]. The Hauser model describes the total charge delivered to a particular circuit node as the sum of the normal charge collected by funneling (Sec. 3.3.2) and charge conducted through the shunt conduit. The shunt charge is given by:

$$Q_R = \frac{T_R \bar{\mu}}{L_T L_R} (V_2 - V_1) Q_T \cos \theta \quad (10)$$

where:

T_R [sec] is the time the shunt exists

L_T [cm] is the length of the track (including funneling)

L_R [cm] is the distance between like regions spanned by the shunt

θ is the angle of particle incidence from normal incidence

V_2 and V_1 [V] are the applied biases to the like regions

$\bar{\mu}$ [$\text{cm}^2/\text{V}\cdot\text{s}$] is the average mobilities of charge carriers in the shunt, and

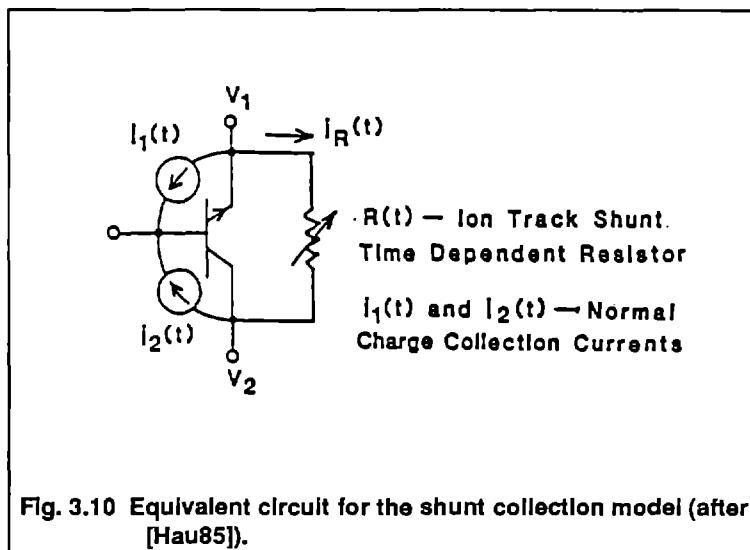
Q_T [C] is the total charge deposited in the track.

This analytical model is based on the assumptions of uniform doping along the plasma track and a uniform charge generation (LET) along the track. However, the model is very enlightening as to the major contributors to the shunt effect in ICs. It shows that shunt charge depends on (1) the carrier mobilities, (2) the mean time the track exists, (3) the length of the resistive track, and (4) the voltage difference between the two nodes connected by the track. Notice that the ion species and energy are not present explicitly in Eq. 10, but are implicit in T_R , the time the track exists.

Equation 10 is in the form of a resistance between the shunt nodes which appears just for time T_R . Figure 3.10 shows the equivalent circuit model for these currents, where the transient shunt resistance can be derived from Eq. 10 by making the assumption that $Q_T/L_T = N_o$, the linear charge density in the track, giving:

$$R = \frac{L_R}{\bar{\mu} N_o \cos \theta} \quad (11)$$

where N_o [cm⁻¹] is the linear charge density along the track. This resistance exists in the circuit for time T_R .



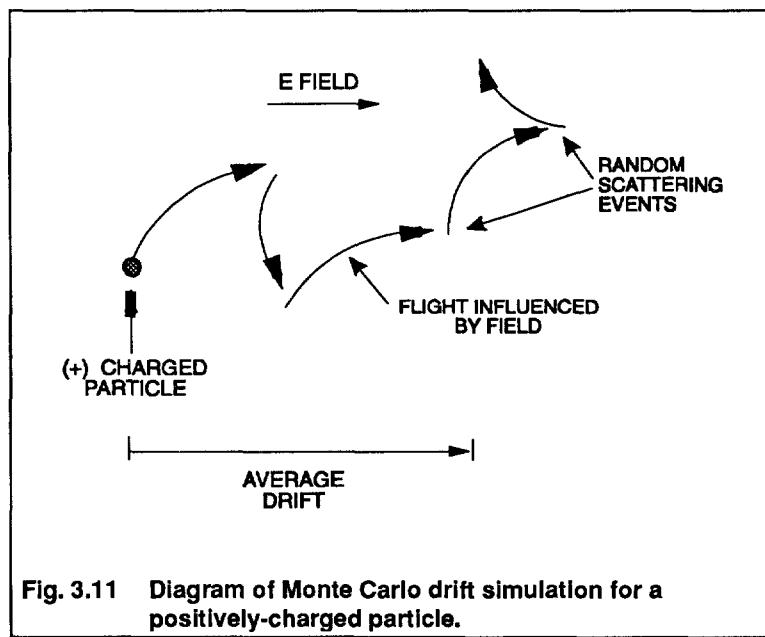
All of the parameters of this shunt model come from known material/device parameters except T_R . The time that the shunt exists is a complicated function of the voltages on the shunt nodes, and has not been analytically modeled by first principles. It can, however, be modeled by the 2-d and 3-d physical models discussed in Sec. 3.4.

Knudson et. al. later expanded the Hauser shunt model to include a spreading resistance at the ends of the shunt in series with the actual shunt resistance [Knu86]. Experimental results of charge collection in various geometries were used to empirically fit the parameters of this model.

3.4 2-D/3-D Collection Models

There exist a class of device modeling tools which can be used to effectively reveal the basic physical effects of SE interactions with IC devices. **Monte Carlo** and **Finite-Element Boltzmann Transport Theory** codes model the basic physics of carrier motion under the influences of the device geometries (boundary conditions) and applied electric fields. These types of simulations, at the very detailed microscopic level, can do much to expand the understanding of the underlying physics involved in the complex process of SE charge deposition and collection. These detailed simulations also aid in the development of analytical models, such as those already discussed.

A simplistic schematic of a one-carrier Monte Carlo simulation is shown in Figure 3.11. The movement of a single charge carrier under the influence of an electric field is shown as a sequence of accelerations due to the E-field and momentum shifts due to scattering events. Monte Carlo simulations perform a statistically large repetition of these detailed calculations with randomized scatterings in order to approximate the average expected movement of this one carrier. This sequence must then be repeated for many other carriers to describe the motion of a group of charge. Although each calculation is relatively simple, the sheer number of calculations which must be performed to arrive at a statistically viable result is tremendous. Monte Carlo simulations are useful for the study of basic material characteristics or the study of the charge deposition of one particular ion in a bulk semiconductor material [Sai80] [Mar87], but have limited usefulness in the scenarios of charge collection among the junctions and 3-d geometries of dense, complex IC devices.



A more practical tool for the analysis of charge movements in devices is a 2-d or 3-d simulation based on the semiconductor equations developed from Boltzmann transport theory applied to a finite-element grid. Given a structure such as shown in Figure 3.12, these models evaluate the semiconductor carrier equations at every point in the grid. The equations governing this process are

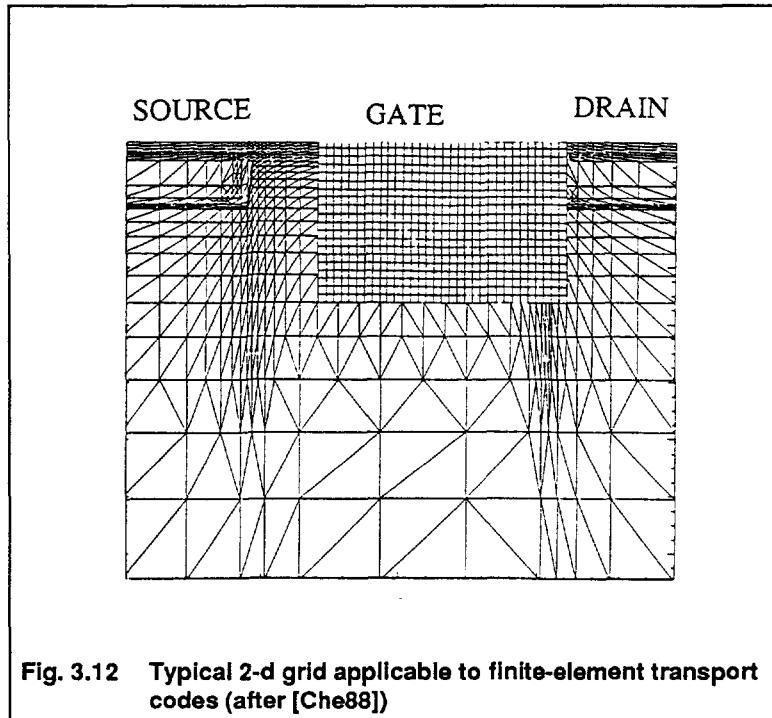


Fig. 3.12 Typical 2-d grid applicable to finite-element transport codes (after [Che88])

Poisson's equation, the current density equations, and the continuity equations. Thus, for each finite element (grid point), the following set of coupled partial differential equations are solved:

$$\begin{aligned}
 \hat{j}_p &= -\mu_p p \nabla w = -\mu_p [p \nabla u + \nabla p] \\
 \hat{j}_n &= -\mu_n n \nabla v = -\mu_n [n \nabla u - \nabla n] \\
 -\nabla \cdot \hat{j}_n &= \nabla \cdot \hat{j}_p = g - r \\
 -\nabla^2 u + n - p + k &= 0
 \end{aligned} \tag{12}$$

where:

- v, w are the normalized quasi-Fermi levels for electrons and holes, respectively
- u is the normalized potential,
- n, p are the normalized carrier concentrations for electrons and holes, respectively,
- \hat{j}_n, \hat{j}_p are the normalized current densities for electrons and holes, respectively,
- g=normalized net carrier generation rate,
- r=normalized net carrier recombination rate, and
- k=normalized net impurity density.

All of these quantities are normalized to aid in numerical overflow control. For example, doping levels in semiconductors can be of order greater than 10^{15} while quasi-Fermi levels in the forbidden band can be on the order of tens of electron-volts. This disparity of real numbers, if normalized out, greatly enhances computational efficiency and numerical accuracy. The usual normalization method is to define all spatial dimensions in units of Debye length, all densities in units of intrinsic concentration, n_i , and all potentials in units of the thermal voltage, kT/q [Che88].

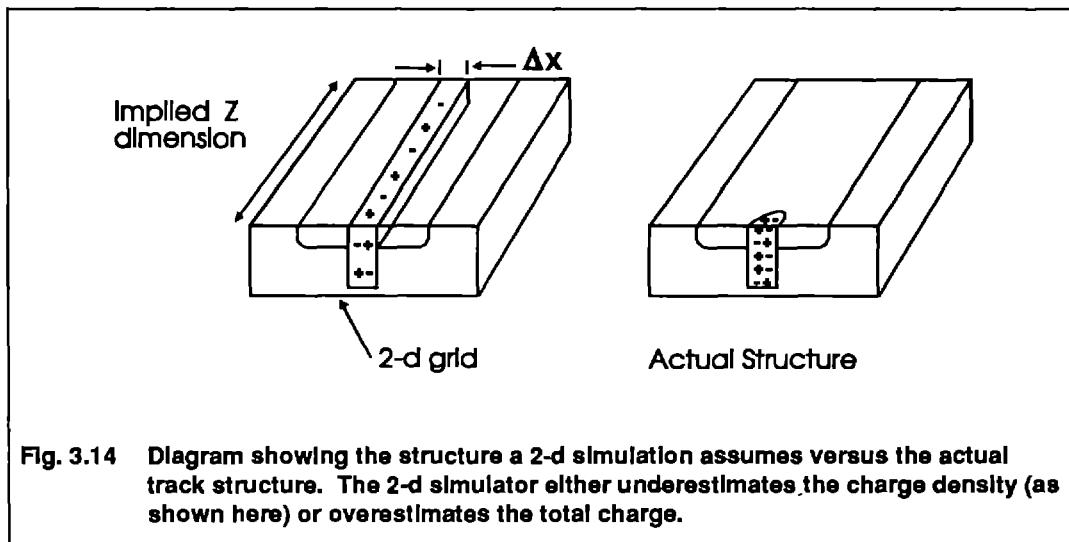
Boltzman finite element analyses have very few inherent assumptions associated with the computations, thus they can be very accurate simulations of real carrier movements in complex geometries. The key assumption in these codes is the quasi-static local field approximation, which reduces space and momentum coordinates to only spatial. It is assumed that the response of carriers to changes in the electric field is much faster than the rate of change of the field itself; or in device term, the device interior responses are slower than the fundamental responses of the charge carriers. This approximation is entirely applicable to all current mature technologies, and most which will emerge in the foreseeable future. The approximation will break down with super-sub-micron devices which operate at the quantum level.

When applying these Boltzman finite element analysis codes, the user must be especially conscious of and careful about the boundary conditions he defines in the simulation. Boundaries include Dirichlet conditions (ohmic, with carrier concentration = 0), Schottky, and Neumann conditions (reflecting). As an example, the simulation grid must be large enough that these boundary constraints, which are necessary around the edges of the grid, do not artificially influence the carrier physics in the region of interest.

It can be seen, due to the gradients in Eq. 12, that the potential and carrier concentrations at every point are dependent on the values of the surrounding eight grid points. Thus the need to solve all of these equations at all points simultaneously, using matrix iterative techniques -- a monumental computational task. While these finite-element codes are very accurate and can produce considerable insight into the microscopic interactions of the SE with the surrounding junctions, the computational complexity of these codes usually prohibit their use on all but the most simple device structures, and certainly not on complete circuit topologies.

Two popular 2-dimensional finite element semiconductor codes in the public domain are PISCESII (Stanford University) and MINIMOS (Technische Universitat Wien, Vienna). Application of these

simulation tools yields results similar to those of **Figure 3.13** (generated by Silvaco's THUNDER). Several groups have used such results to gain valuable insight into the complex processes that take place following a SE strike [Gru84] [Iwa92] [Knu91] [Kre85] [Rol88b]. However, the modeling of an inherently cylindrical plasma track and a parallelepiped device structure with a 2-d simulation leads to questioned validity [Kre86]. The actual 3-d structure modeled by this 2-d simulation is shown in **Figure 3.14**, where the 2-d simulator implicitly defines a z dimension of identical 2-d slices. The particle strike and the device's z dimension extend to infinity, leading to a sheet of charge along the ion track rather than a column. The most serious concern in this scenario is proper charge conservation. If one sets the total charge of the sheet equal to the ionic deposited charge, then the charge density of the simulation is too low. On the other hand, if one sets the charge density in the sheet equal to the real charge density of the ion track, then the total charge in the simulated sheet is too large.



Three-dimensional finite element device models apply the same techniques just presented, only with added grid points in the z dimension. These codes effectively address the charge conservation problem mentioned above, as well as providing a more realistic and interpretable view of the potential gradients and current flows in the device. An example SE hit in a MOSFET device (generated by the Silvaco THUNDER code) is shown in **Figure 3.15**. PADRE (Bell Laboratories), THUNDER (Silvaco Corporation), and DAVINCI (Technology Modeling Associates, Inc.) are examples of commercial 3-d finite element codes for device modeling at the physical level.

The drawback of full 3-d simulations is the computing resources required. The computational problem grows geometrically with the addition of the z dimension grid points. It is not uncommon for one 3-d finite element simulation to consume tens to hundreds of hours of CPU time on a typical engineering workstation.

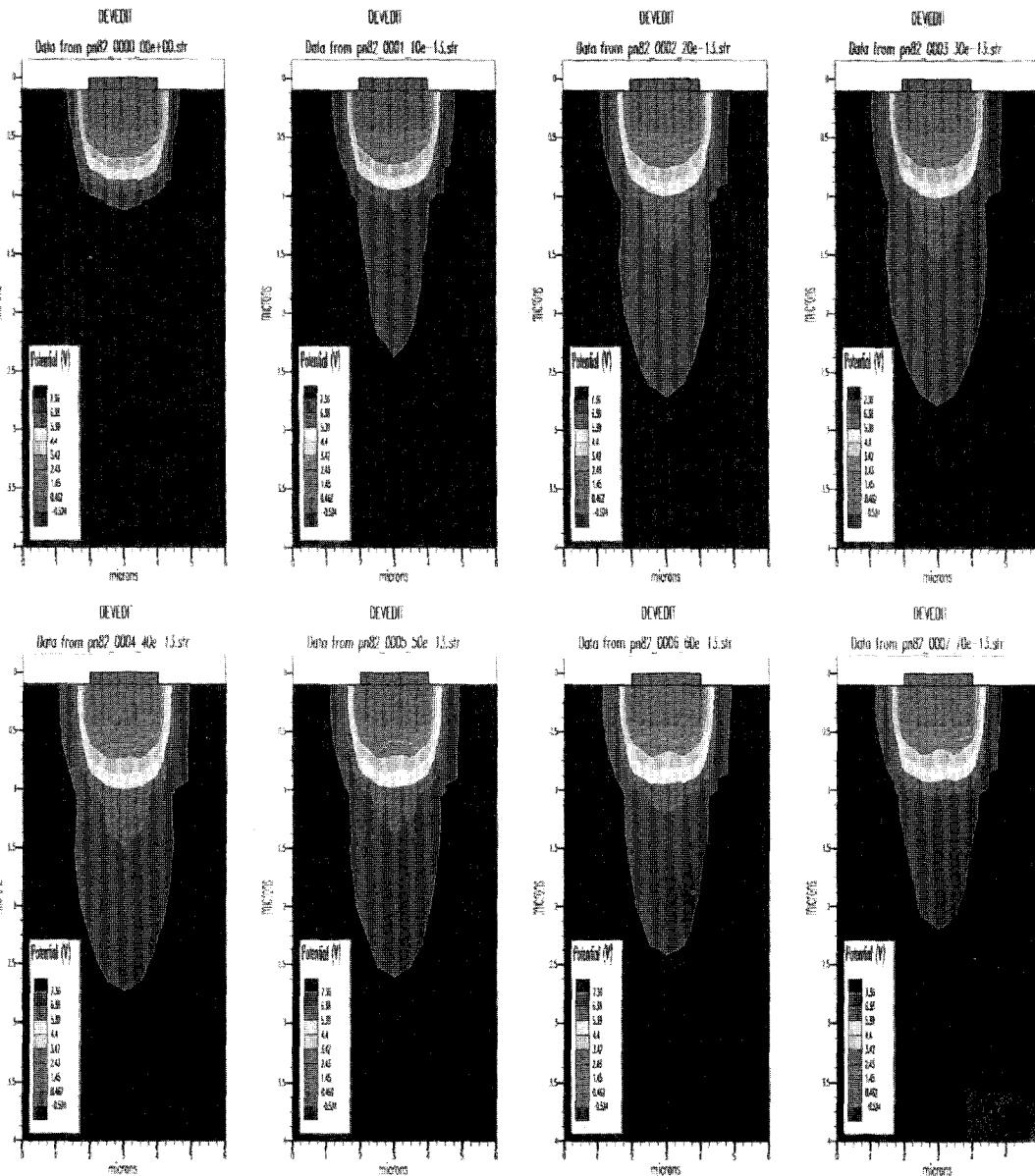


Fig. 3.13 Example 2-d finite-element analysis results. The plots show the creation of the field funnel under an n+/p junction due to the ion track. Shading indicates equipotential contours and the plots span 0 to 8 ps in 1ps time steps. (simulation results from Silvaco's THUNDER code in 2-d)

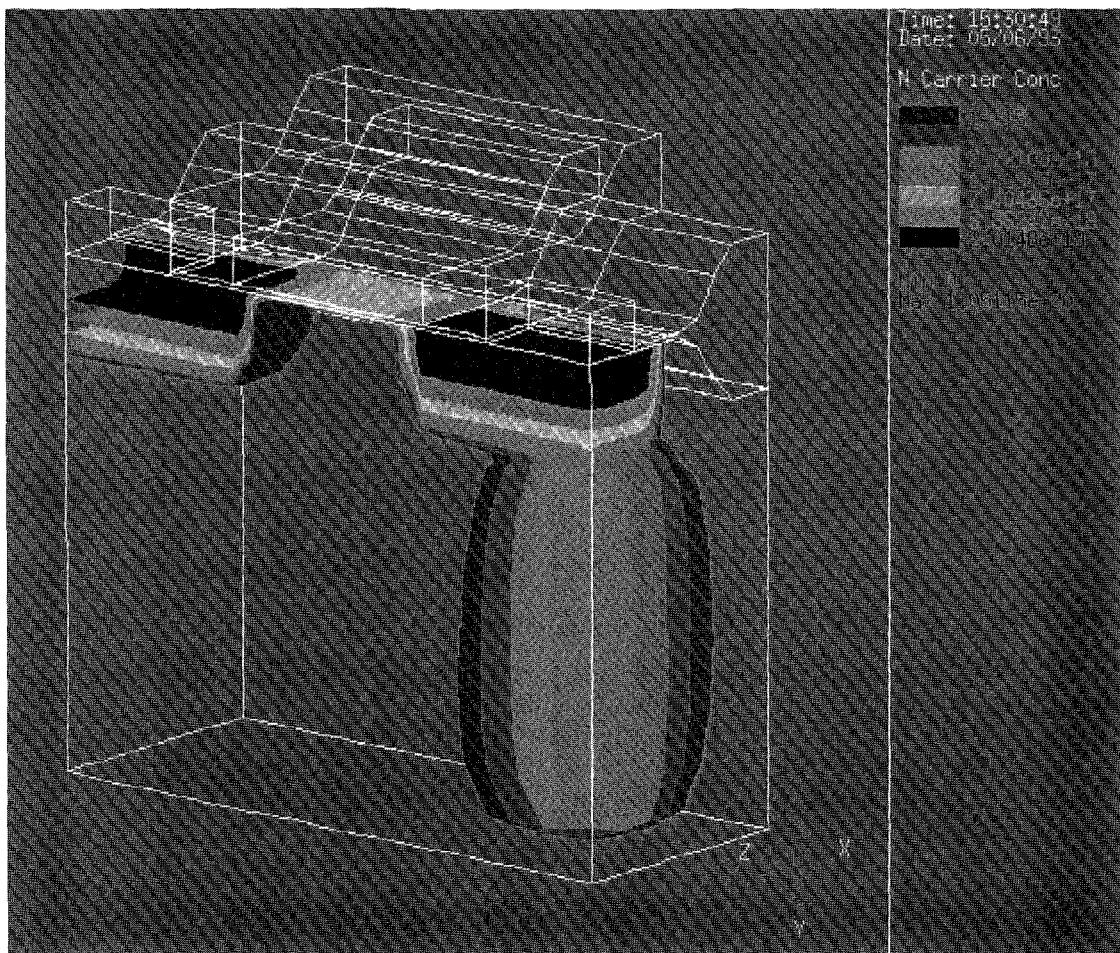


Fig. 3.15 Example simulation results for a 3-d finite element analysis. The plot shows contour surfaces for minority carrier charge density surrounding an ion hit to the drain of a MOSFET device. The shading indicates a cut-away view in 3-d (surfaces) of charge and the time interval is 15ps following the ion strike. The drift and diffusion of these carriers can be monitored at the top contacts of the device to accurately simulate the SE transient waveform and total charge. (Simulation results from Silvaco's THUNDER code).

3.5 Recap of Physical Modeling

In this section, we have studied the modeling of the physical interaction of an ion strike with an active semiconductor junction. Models for the charge generation along the SE track, which the engineer can use to approximate the charge deposited by the SE ion have been presented. Several analytical models for the subsequent charge collection by a p-n junction via drift, funneling, or the shunt effect were discussed. These analytical models have been manipulated into a form that the engineer can easily apply them to his particular problem with only the basic material and physical parameters for his technology. Assumptions and limitations of these analytical models were presented to guide the user in proper application of the tools. Also, more detailed 2-d and 3-d modeling tools have been introduced and can be utilized for more detailed charge collection calculations if the user so desires.

The result of this modeling effort is either a current model (I_{SE}) or a collected charge value (Q_{coll}), which can be used as the input to the next level of the simulation hierarchy. This current/charge input will form the basis of our modeling of the subcircuit response to the SE strike. Note that everything developed in this section relates to the ion, its charge deposition, and the charge collection by proximal junctions. Q_{coll} is a property of the ion, its LET, the semiconductor material and geometry, and possibly the applied bias of the collecting junction; it is not a property of the circuit configuration or the charge needed to upset the circuit. This concept is quite important in relation to the next section.

4. MACRO-MODELING : CIRCUIT ERRORS DUE TO SINGLE-EVENTS

In the previous section, we discussed models for the charge generation by a single penetrating ion and the charge collection by the surrounding junctions. Stepping up in a hierarchical view, these models can be incorporated into macro-models of the devices interconnected in a subcircuit. The macro view of the circuit will relate the collection of charge in individual device junctions to changes in the circuit currents and voltages. Our goal is to model the relationship between the SE-induced perturbations in circuit currents and voltages to upsets and errors. These circuit models can then be used to: (1) assess the vulnerability of a circuit to particular LETs, (2) compare the SE hardness (tolerance) of different designs and/or technologies, (3) develop and simulate the effectiveness of hardening schemes, and (4) predict the error response of a circuit in a particular environment.

In this section , we will introduce a circuit-level description of the SE photocurrent, based on the models developed in the previous section. This circuit model will be integrated into models of subcircuits and circuits at the macro level (transistors, resistors, capacitors, etc.) to study its effect on sensitive circuit nodes where information is stored.

4.1 Circuit Model for Charge Collection

A common circuit model for the charge collection at a junction due to direct funneling or diffusion is a double-exponential, time-dependent current pulse, as described in Eq. 8 and shown in Figure 4.1. Typical parameters of this pulse are a rise time on the order of tens of picoseconds and a fall time on the order of 200 to 300 picoseconds. The equation for this current pulse which matches the form in circuit simulation codes such as SPICE is

$$I(t) = \begin{cases} I_o \left(1 - e^{-t/\tau_r} \right) & ; \quad t < t_D \\ I_o \left(1 - e^{-t_D/\tau_r} \right) e^{-(t-t_D)/\tau_f} & ; \quad t > t_D \end{cases} \quad (13)$$

where the parameters are defined in Figure 4.1. The total charge delivered by the current pulse of Eq. 13 is the integral over time of $I(t)$. Performing this integration gives:

$$Q = I_o \left[t_D + \tau_F - \tau_R - (\tau_F - \tau_R) e^{-t_D/\tau_R} \right] \quad (14)$$

These two equations give the user the tools to relate the collected charge produced by the modeling of Sec. 3 with a proper waveform to be included in circuit simulations.

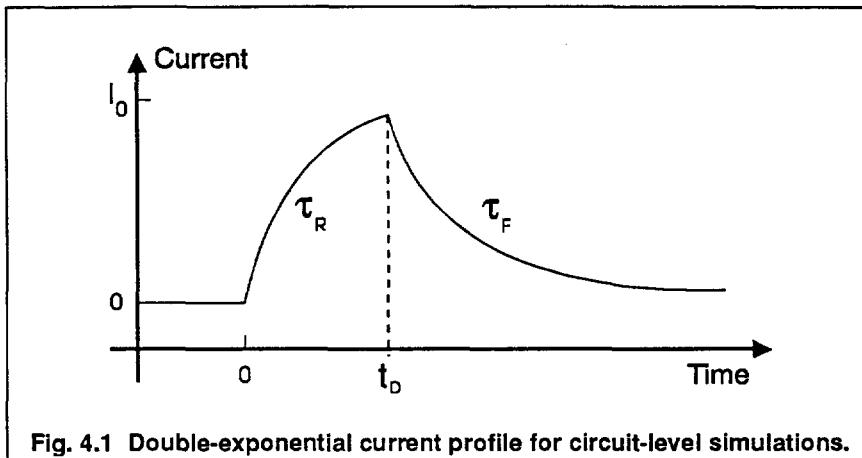


Fig. 4.1 Double-exponential current profile for circuit-level simulations.

The magnitude and time profile of the current model depend on material parameters, the ion species, the ion energy, and the hit location relative to the junction. As shown in the experimental pulse-shape results of Figure 4.2, both the magnitude and time profile of the actual collection current pulse can be highly variable [Wag88]. If the time profile of this collection current is not important to the circuit response to the hit, then the 1-d analytical models of Sec. 3.3 usually adequately describe the current (or total charge). If, however, the time profile is critical to the circuit response, more

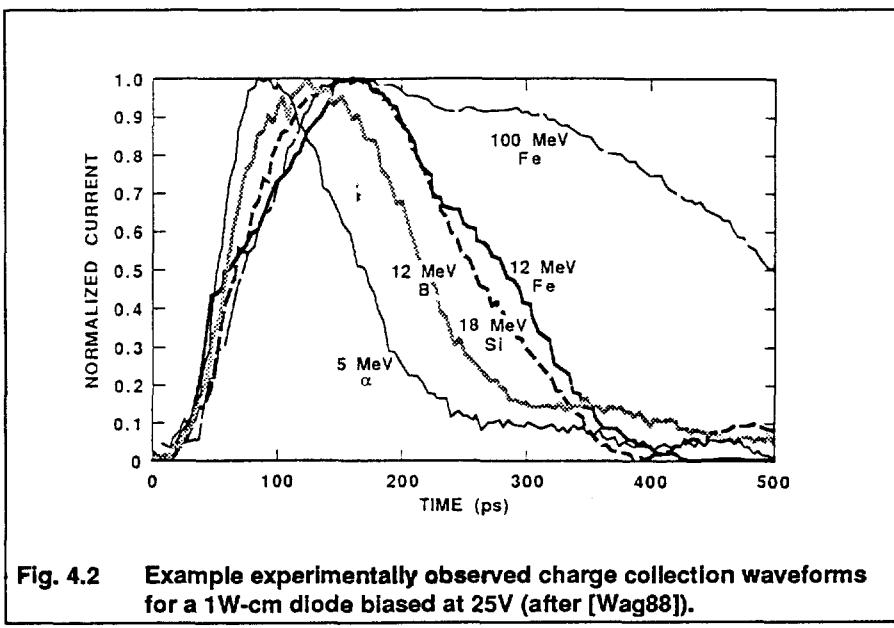


Fig. 4.2 Example experimentally observed charge collection waveforms for a 1W-cm diode biased at 25V (after [Wag88]).

detailed development of the current shape is necessary, as with the 2-d and 3-d finite-element models of Sec. 3.4. In either case, at this level of the hierarchy, a circuit-level model of the collection current at each junction of interest is included in the lumped-element circuit description, as shown in Figure 4.3.

Historically, it has become common practice to use the total charge delivered by the current waveform (as in Eq. 14) as a single descriptor of the SE effect on the affected circuit node. This can be an extremely dangerous simplification, as will be presented in subsequent sections, since it assumes the time-profile of the charge delivery to the sensitive node is unimportant to the response. From this simplification comes the concept of **critical charge**, Q_C . Q_C is a property of the particular circuit (not the ion or environment) and is defined as the minimum charge delivered by the SE current waveform of Figure 4.1 (given by Equation 13) which causes that circuit to lose information and create an error condition. In rough terms, it is the amount of charge needed on a sensitive node to cause an upset. Of course, Q_C , tells nothing about the time profile of the delivery of that charge.

Critical charge is commonly used as a figure of merit in the comparison of circuit design types and technologies. The quantity describes the vulnerability of a circuit to single events without the complications of ion species, ion energies, LET, or type of charge collection.

4.2 Circuit Response to the Single-Event Current

In the previous subsection, we studied a simple circuit model for the SE-induced photocurrent. When this model is included in a circuit-level, lumped-parameter simulation, the effect of the SE strike on a circuit's critical nodes can be modeled. The importance (or lack thereof) of the precise time profile of this current model depends entirely on the dominant high-frequency pole associated with the circuit's hit node, given by the open-circuit time constant (total nodal capacitance and equivalent ac resistance) seen by the current source. That is, the hit node acts as a low-pass filter, integrating the charge delivered to the node by the SE current pulse.

If the response of the circuit at the collecting node is much slower (the time constant is much longer) than the characteristic time constant of the SE pulse ($\tau_R + \tau_F$ in Figure 4.1), then the pulse is

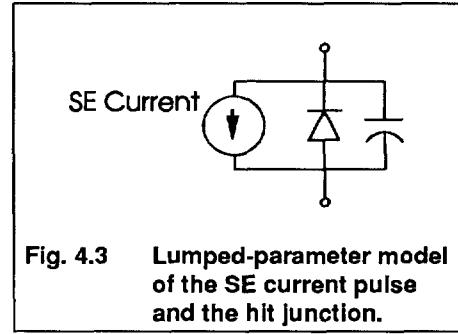


Fig. 4.3 Lumped-parameter model of the SE current pulse and the hit junction.

effectively integrated by the nodal capacitance and only the total charge delivered by the pulse is important to the circuit response. If, however, the time constant at the node is much shorter than the time constant of the SE pulse, then the circuit responds to the delivered charge faster than the pulse can deliver it; so the pulse shape is critically important to the circuit response. These concepts are essential to the accurate modeling of SEUs at the circuit level, since they define the boundary between valid modeling using only the collected charge and modeling requiring a more accurate description of the time profile of the charge collection.

To illustrate this concept, consider the circuit node shown in Figure 4.4, with a current model for a SE strike. The open-circuit time constant associated with the node is given by

$$\tau_N = R_{EQ} C_N \quad (15)$$

where C_N [F] is the total nodal capacitance, and R_{EQ} [Ω] is the equivalent resistance seen by C_N .

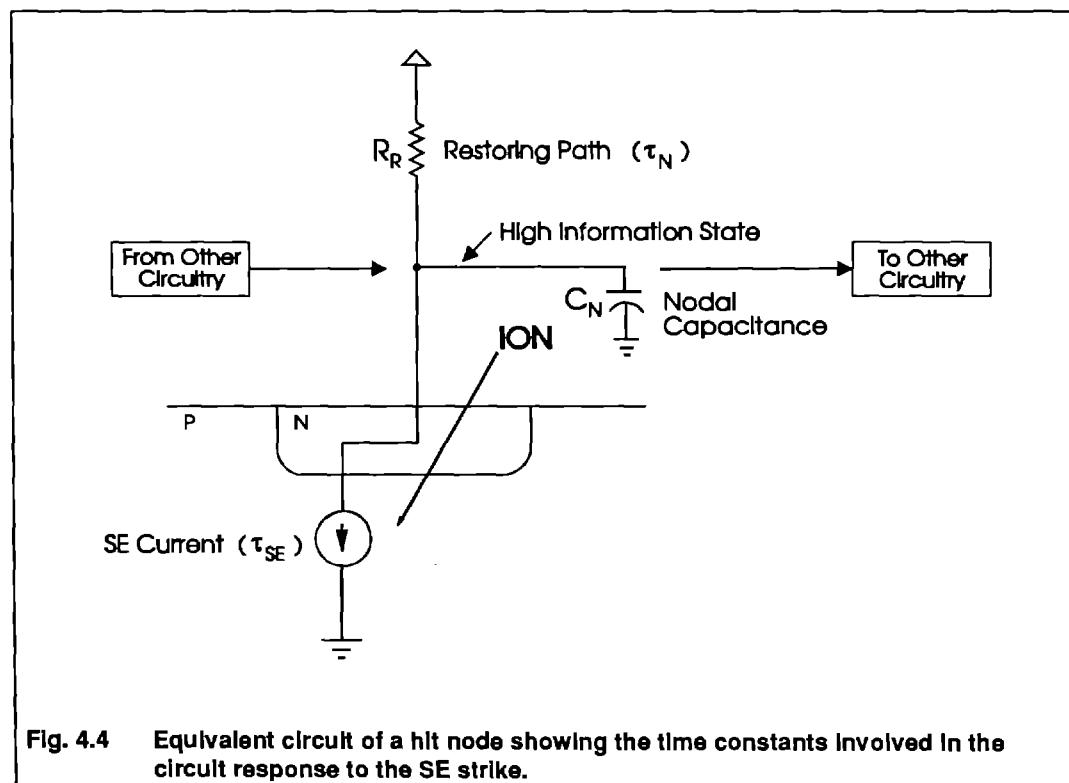


Figure 4.5 shows circuit simulations of the response of the node when $\tau_N = R_{EQ} C_N$ is much longer than the time constant of the SE current pulse, τ_{SE} . In both parts (a) and (b), the total charge delivered by the SE pulses is the same, however the delivery time profile is different. As can be seen, the final voltage state on the node is identical following the SE strike in both cases. When $\tau_N \gg \tau_{SE}$, the concept of critical charge holds, as the circuit response to the simulated hit is independent of the time profile of the SE current model. In fact, the magnitude of the voltage response at the node is given by:

$$\Delta V_N = \frac{Q_{collected}}{C_N} \quad (16)$$

and

$$Q_C = C_N \cdot \Delta V_{NM} \quad (17)$$

where Q_C is the critical charge and ΔV_{NM} is the noise margin of the circuit.

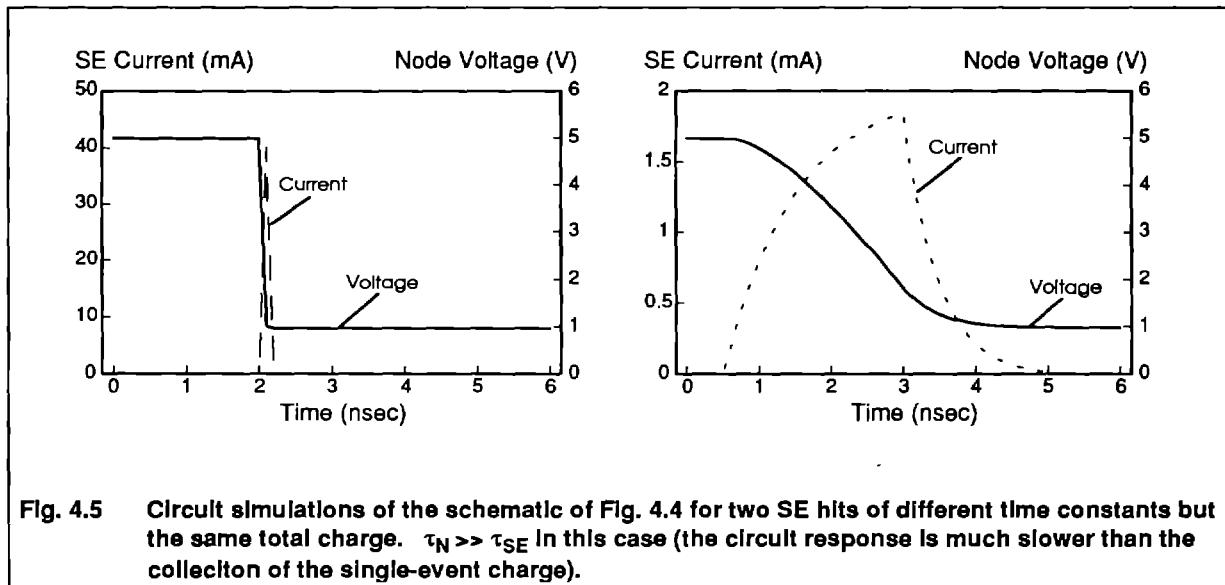


Fig. 4.5 Circuit simulations of the schematic of Fig. 4.4 for two SE hits of different time constants but the same total charge. $\tau_N \gg \tau_{SE}$ In this case (the circuit response is much slower than the collection of the single-event charge).

It is from this type of response that the concept of critical charge as a measure of circuit vulnerability has developed. This response is typical of DRAMs, resistive-load SRAMs, most CMOS, and older bipolar technologies. It is usually not, however, a valid model for very fast bipolar, VHSIC, microwave, high-speed analog, and some GaAs technologies. The modeler must be aware of these limitations.

In contrast, **Figure 4.6** shows the same SE current models (both with the same charge Q) applied to a circuit node with τ_N much shorter than τ_{SE} . The voltage perturbation on the hit node can be seen to be very different in parts (a) and (b), even though the total charge delivered is identical in both

cases. The response of the node is dominated by the low pass filter characteristics of the node, i.e. the impulse response at the node. Thus, an accurate model for the time profile of the SE charge collection is essential in this case in order to hope for an accurate simulation of the circuit response. This type of response is typical of high-speed circuit technologies, such as those mentioned above. In these cases, the concept of critical charge is a misnomer, since it is obvious that it is not possible to adequately describe (or model) upsets with just a single charge parameter.

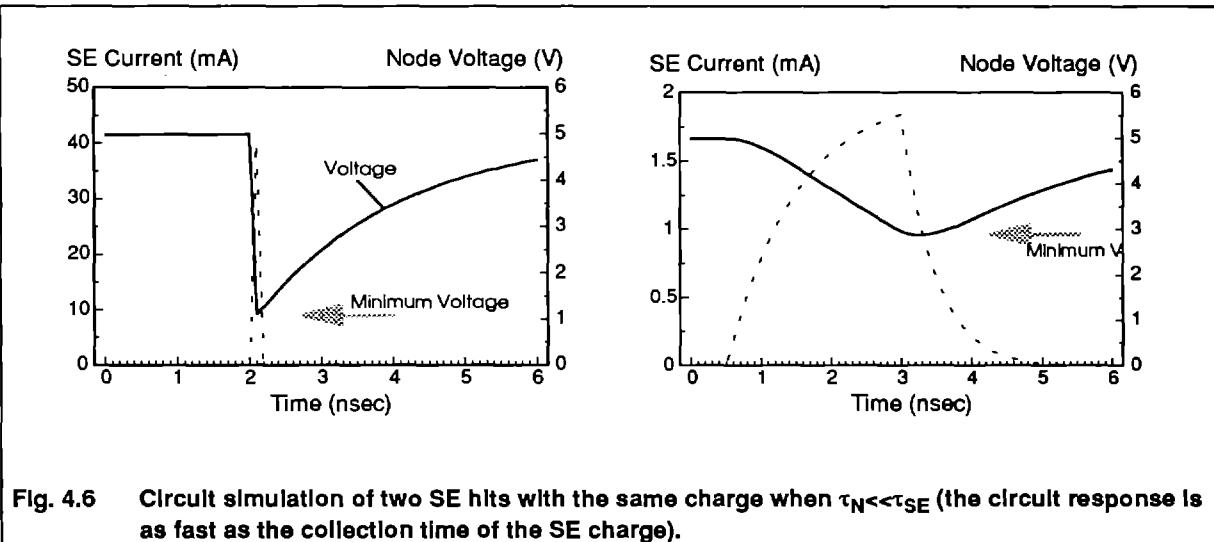


Fig. 4.6 Circuit simulation of two SE hits with the same charge when $\tau_N \ll \tau_{SE}$ (the circuit response is as fast as the collection time of the SE charge).

Having set up circuit modeling and the restrictions of which the modeler needs to be aware, in the following subsections we will briefly discuss circuit-level modeling of SE effects in various circuit designs and technologies. We will discuss both hard and soft errors, but will concentrate on the salient circuit modeling issues for simulating soft errors.

4.3 Destructive Effects

Single-event upsets belong to a broader class of effects known as **single-event phenomena (SEP)**. The distinction between the subclasses of SEP is rather context driven, but here we will distinguish effects which can lead to permanent circuit failures from those events which simply upset the normal circuit operation. In most cases, SEUs refer to the latter, so this section will concentrate on upsets. However, we will briefly introduce destructive effects, since they also play a crucial role in space systems exposed to single-event particles.

The following single-particle effects are either inherently destructive, or given the proper conditions may be destructive, to a circuit. It is not possible to properly cover the complexities of these effects in

this tutorial, but the basic mechanisms will be presented and models introduced. For further details, the reader is directed to the references.

4.3.1 Latchup and Snapback

Latchup is a regenerative current-flow condition which can be induced in any semiconductor structure which possesses a parasitic n-p-n-p path; but it has been of most concern in bulk CMOS integrated circuits. The basic cross-sectional structure of a latchup path and its circuit equivalent are shown in **Figure 4.7**. The charge collection (photocurrent) from a single-event hit in either the base-emitter junction of the parasitic npn transistor, or the emitter-base junction of the pnp transistor, can trigger the regenerative circuit -- **single event latchup (SEL)**. Current flow in R_W (the spreading resistance of the well region to the V_{SS} contact) or in R_S (the spreading resistance of the substrate to the V_{DD} contact) can forward bias the parasitic transistors, leading to more base current and greater forward bias on the transistors, in a regenerative action. This circuit provides a path for large amounts of current flow between the power supply rails. If the energy created by this current path exceeds the thermal dissipation capacity of the surrounding material, melting and electromigration can occur, leading to a destructive breakdown -- a hard error. Even if destructive breakdown does not occur, the latched path will persist until power is removed from the circuit, causing a catastrophic failure of the circuit.

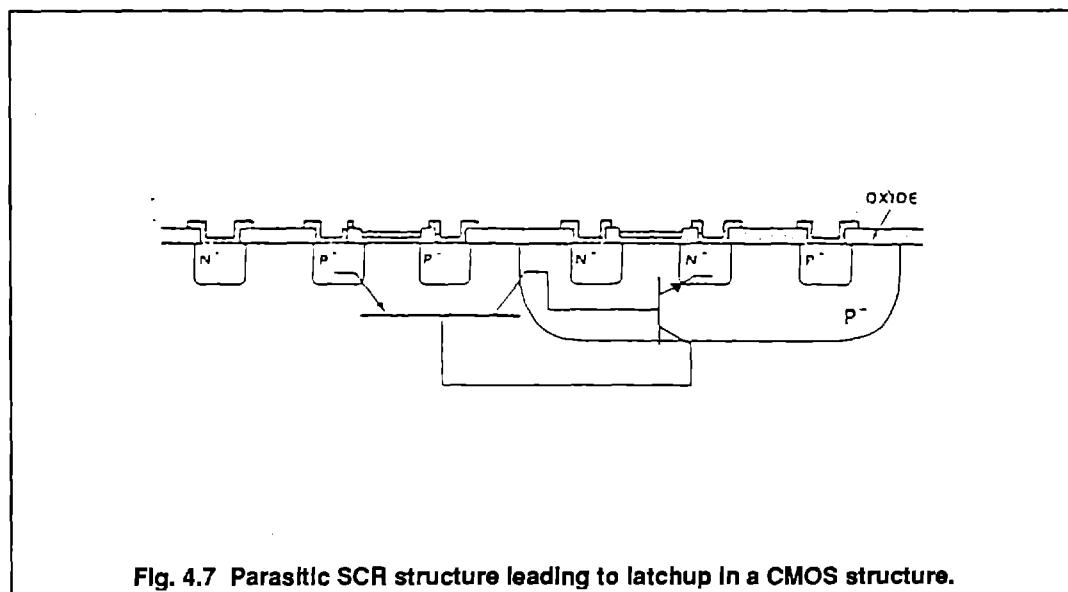


Fig. 4.7 Parasitic SCR structure leading to latchup in a CMOS structure.

Single-event latchup is most often modeled via a circuit model as in Figure 4.8 where the resistances are ohmic paths between the regions and device spreading resistances for the geometry shown in Figure 4.7 [Joh90]. However, analysis is complicated by the three-dimensional geometry involved in the parasitic spreading resistances and the 3-d characteristics of the parasitic bipolar transistors [Och81]. 2-d or 3-d modeling are often used to characterize these lumped parameters for the circuit model.

Similar to latchup, snapback is also a regenerative current mechanisms, but does not require the p-n-p-n structure [Sun78]. In n-channel MOS transistors with large currents, such as IC output driver devices, a parasitic npn bipolar action can be triggered by SE-induced avalanche multiplication near the drain junction of the device [Och83], shown in Figure 4.9. Avalancheing in the depletion region of the drain, due to the large current flow, causes holes to be injected into the p-substrate region under the gate. These holes act as base current to the parasitic npn transistor, causing electrons to be injected by the source and collected by the drain. This increased current leads to more avalanching, more base current, and the regenerative loop is closed. Figure 4.10 shows a circuit model that can be used to model the snapback effect in a CMOS output inverter [Bei88].

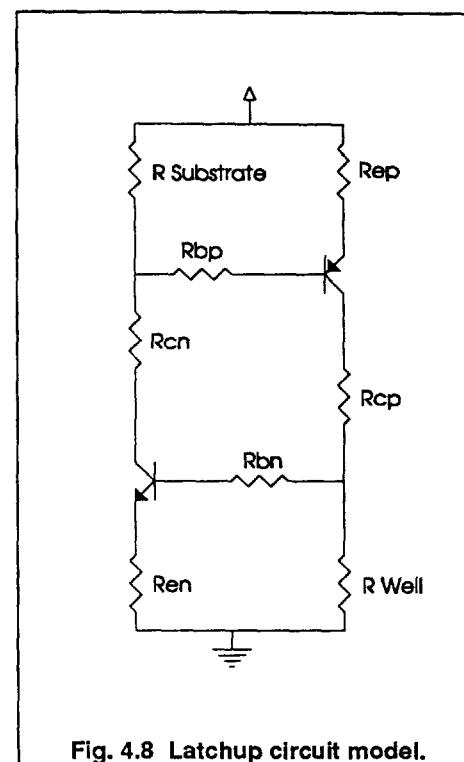


Fig. 4.8 Latchup circuit model.

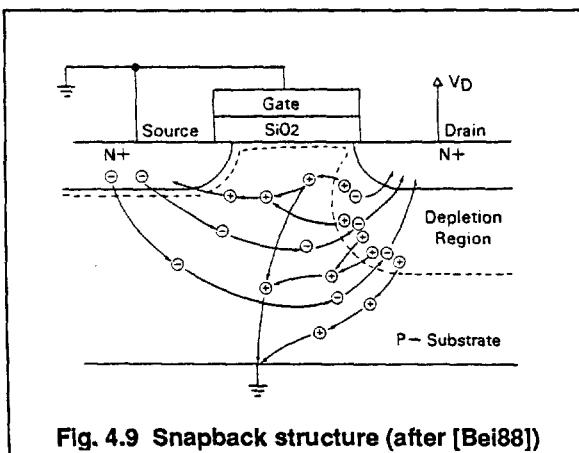


Fig. 4.9 Snapback structure (after [Bei88]).

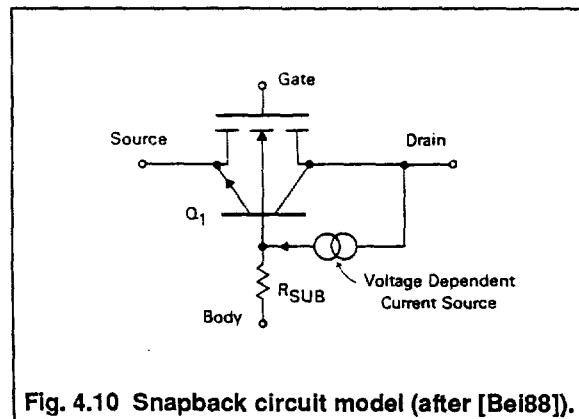


Fig. 4.10 Snapback circuit model (after [Bei88]).

4.3.2 Burnout

Power MOSFET devices, which have large applied biases and high internal electric fields, are susceptible to single-event induced burnout (SEB). As shown in Figure 4.11, the penetration of the source-body-drain region by a single-event can forward bias the thin body region under the source. If the terminal bias applied to the drain exceeds the local breakdown voltage of the parasitic bipolar, the single-event strike can initiate avalanching in the drain depletion region [Hoh87]. Local power dissipation due to the large drain-to-source current leads to destruction of the device. Similar effects have also been seen in power bipolar devices [Tit91].

Models for this effect typically require 2-d or 3-d simulation of the complex field lines and carrier transport in and near the avalanching region; however, a few circuit models and at least one analytical model have been presented [Hoh87]. Figure 4.11 shows the circuit equivalent for the effect. Proper use of this model requires very accurate characterization of the parasitic bipolar, especially the current-dependent avalanche multiplication and the junction breakdown voltage.

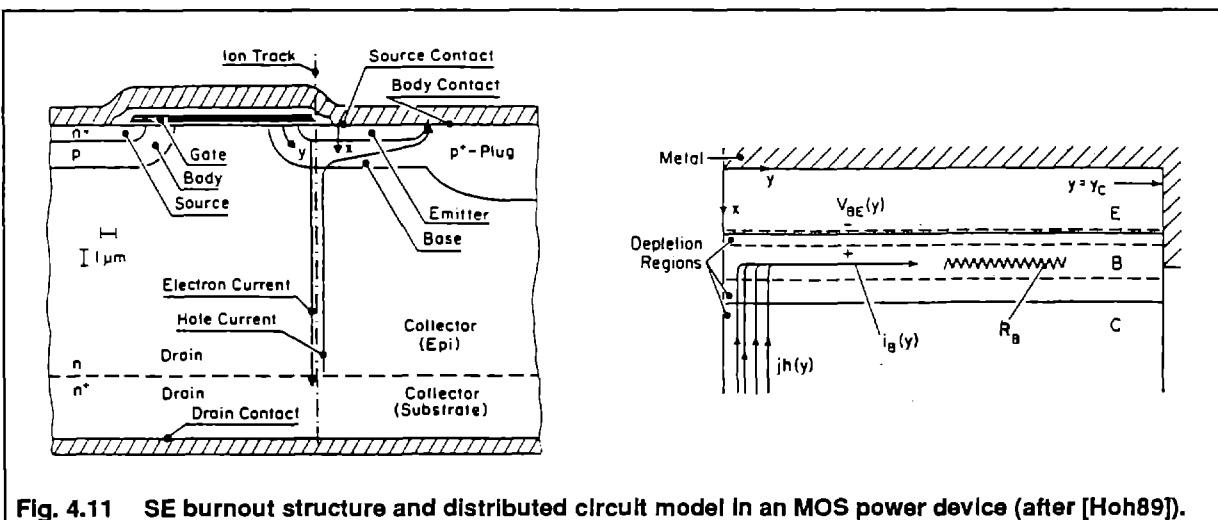


Fig. 4.11 SE burnout structure and distributed circuit model in an MOS power device (after [Hoh89]).

4.3.3 Gate Rupture

If an ionizing particle passes through the gate oxide region of a power transistor, a transient plasma filament can connect the gate conductor and the semiconducting material under the gate oxide. In power transistors, where a large bias is applied to the gate, this filament will allow large currents to flow through the gate oxide, possibly leading to thermal breakdown and destruction of the oxide, a single-event gate rupture (SEGR) [Fis87][Wro87]. There are presently no comprehensive models for this effect, although it is currently an area of active modeling research.

4.4 Non-Destructive Soft Errors Defined

The most basic, colloquial view of a nondestructive single-event circuit effect might be: the deposition of charge at a circuit node where it should not be, or the depletion of charge from a node where it should be. If the particular circuit node affected by the ion contains electrical information, either a propagating signal or a latched bit, then the possibility of an upset exists. Historically, the terms single-event upset, single-event errors, soft errors, and bit-flips have been used rather loosely and interchangeably. However, a bit-flip in an analog circuit is meaningless; and a memory error in a static RAM is an entirely different beast than a logic error in an ALU. Therefore, in order to attempt to keep a consistent standard in this work, a **single-event upset (SEU)** is defined as the SE-induced corruption of an information signal or bit of stored information. The **single-event soft error** is defined as the erroneous response of the circuit to the upset. In a single-bit static RAM cell, upset and error are synonymous since there exist only two stable states (0 or 1) and corruption implies a bit-flip, i.e. a circuit error. On the other hand, for an analog comparator in a analog-to-digital converter (A/D), an upset might be defined as an erroneous high-to-low or low-to-high output transition of the comparator; but if the timing is such that this upset does not affect the conversion process, then an error has not occurred.

These distinctions are drawn so that we can define a single upset level for a particular isolated circuit without the variabilities of what the surrounding circuitry happens to be doing at the time of the hit. The reader should keep in mind that, as stated above, upset and soft-error are interchangeable terms (and are often used interchangeably in the literature) when referring to memory circuits, the most prevalent single-event study in the field.

Electrical information that is static (i.e. not propagating through a circuit, but is stored in some fashion) is stored either passively or actively. **Passive storage** refers to charge (or equivalently, voltage) stored on a capacitance. Circuits belonging to this category include dynamic random-access memories (DRAMs), resistive-load static RAMs (RMOS SRAMs) (semi-passive), nonvolatile RAMs (NVRAMs), and charge-coupled device (CCDs). All of these circuits rely on a packet of charge to represent one bit of information -- charge present may denote a binary '1' or HIGH, charge absent may denote a binary '0' or LOW (or vice versa). The conceptual key to SEU models of these circuits is that the information is passive, and any (no matter how small) disturbance of the stored information by a SE ion is persistent until corrected by external circuitry. So, what has been so often referred to as a 'bit flip', the transition from one binary state to the other, is not required in these circuits for an upset to

occur. A degradation of the stored signal to a level outside the noise margin of the supporting circuitry/system can be considered an upset since it will lead to erroneous interpretation and a resultant error. This effect is shown schematically in Figure 4.12.

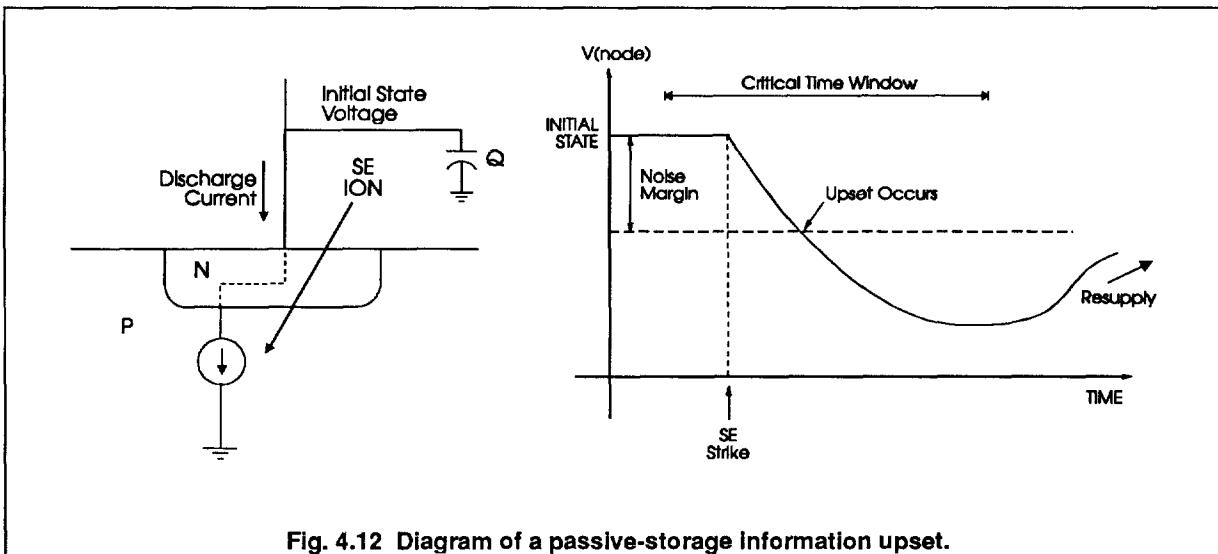


Fig. 4.12 Diagram of a passive-storage information upset.

Active storage, on the other hand, utilizes some form of regeneration to continually reinforce the digital state of the stored information. That is, the electrical information is latched into one of two stable states. This is usually accomplished with high-gain, positive feedback, as shown in Figure 4.13. In this case, signal degradation at the information node is compensated by the regeneration. The SE hit shown in the figure, must not only discharge the signal at the output node, but must also overcome the self-compensation current of the circuit. An upset occurs if and only if the circuit can be

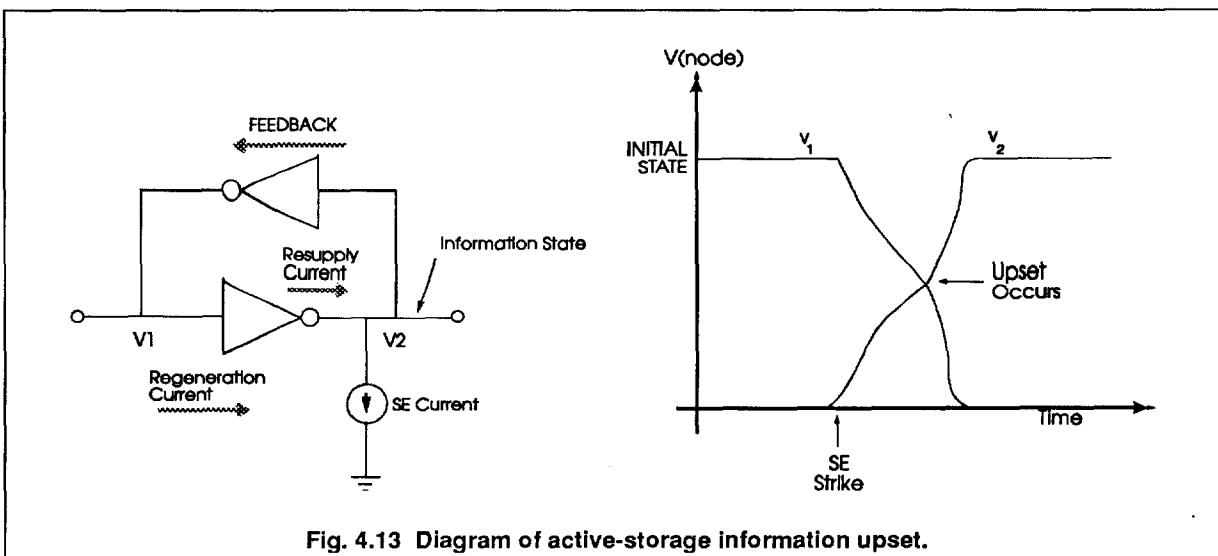


Fig. 4.13 Diagram of active-storage information upset.

thrown from one stable state to the other (i.e. a bit flip). Since this is a *race condition*, timing is much more important to this upset process than the passive upset just studied. Both the signal level to which the SE perturbs the node and the time this perturbation persists are important to modeling the upset process.

4.5 Upset Models for Passive Storage Devices

It has generally been accepted that passive storage memories are not good candidates for space-based applications because of SEU susceptibility. Charge-based storage is inherently SEU prone since most of the schemes perform almost perfect charge integration at the collecting node and there is no method for regenerating the information once an SE hit has occurred. Recently, however, hardening techniques have appeared for these circuits. SE modeling is essential to assessing the viability of these techniques.

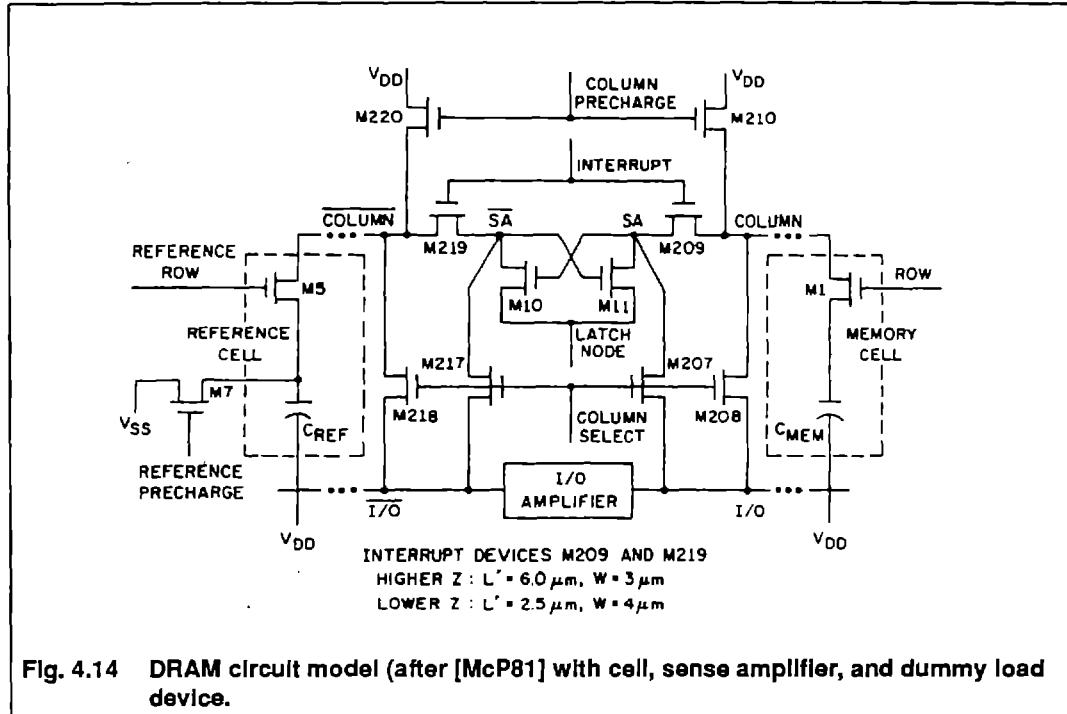
The following subsections describe the SEU modeling of two common passive storage technologies.

4.5.1 DRAMs

Modern one-transistor dynamic RAMs store binary information via a charge packet on an isolated capacitive node, where one bit of information is recognized as the presence or absence of charge with respect to an identical reference capacitance. Since this stored charge will leak off of the capacitance, DRAMs must be periodically refreshed by actively rewriting the stored information.

Figure 4.14 shows a simple view of the one-bit DRAM cell, column, and sense amplifier [McP81]. When writing information to the cell, the sense amplifier latch is forced into a 0-1 or 1-0 state by the I/O lines, the cell access transistor M_1 is turned on, and the high or low voltage state of SA is written to the cell by charge transfer to C_{MEM} . In its holding mode, M_1 is OFF and the information charge is stored on C_{MEM} . As this stored charge leaks and dissipates, the cell is periodically refreshed by performing a read operation. To read this cell, both the columns (C and \bar{C}) are precharged to the HIGH voltage level by activating signal P. Simultaneously, the reference capacitance, C_{REF} , is precharged to the midpoint voltage between the HIGH and LOW voltage (i.e. $V_{DD}/2$). Both the cell and reference cell access transistors are turned on (M_1 and M_R). If the voltage stored on C_{MEM} is greater than that on C_{REF} ($V_{DD}/2$), then C_{REF} will discharge \bar{SA} to a lower voltage than C_{MEM} will

discharge SA. In the other case (voltage on C_{MEM} less than C_{REF}), C_{MEM} will discharge SA to a lower voltage than C_{REF} will discharge \bar{SA} . The sense amplifier latch is activated by L, and the Latch captures and reinforces the small differential disparity between SA and \bar{SA} . This process immediately rewrites the original HIGH or LOW signal back onto C_{MEM} as a refresh.



The preceding description of basic DRAM operation is important to SEU modeling of these devices because it illustrates that upset of these circuits is a complex process involving many transient timing signals and temporal states of the cell [Yan79][Sai82]. There are three areas vulnerable to SEUs. They are (1) an SE hit to the storage capacitance C_{MEM} from the source of M_1 , directly perturbing the stored charge; (2) a hit to the drain junction of M_1 or other cell access transistors along the column causing degradation of the column precharge before a read; and (3) a hit to the sense amplifier latch when it is in a floating state, causing an imbalance before a read cycle. These three locations are shown in **Figure 4.14**.

In order to model the worst case upset scenario of these memories, a straightforward approach is to isolate each of the three vulnerable areas listed and analyze each in its most vulnerable timing state. In this way, the complexities of timing can be partitioned from the SE modeling. A simplified circuit schematic to do this is shown in **Figure 4.15**, with single-event generated current pulses included. Transient circuit simulations, using codes such as SPICE (Univ. of Calif. at Berkeley) or TRIGSPICE

(Vanderbilt Univ.) can show the designer the effect of various collected charge values on the noise margins in the circuit.

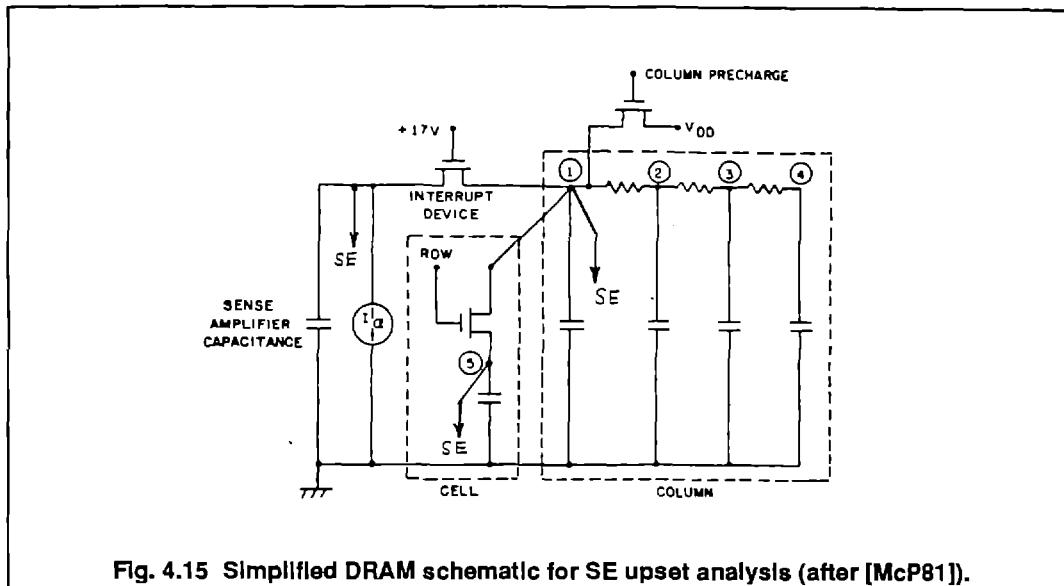


Fig. 4.15 Simplified DRAM schematic for SE upset analysis (after [McP81]).

Analysis of the storage cell itself is straightforward and can even be accomplished analytically, without the need for circuit simulations. Based on the typical cell cross-section and schematic shown in Figure 4.16, the change in voltage on the storage capacitor due to the ion hit is:

$$\Delta V = \frac{Q_{SE}}{C} \quad (18)$$

where Q_{SE} is the total charge delivered to the node by the SE current pulse, i.e. the collected charge at the junction. Using the tolerable noise margin at the node, we have

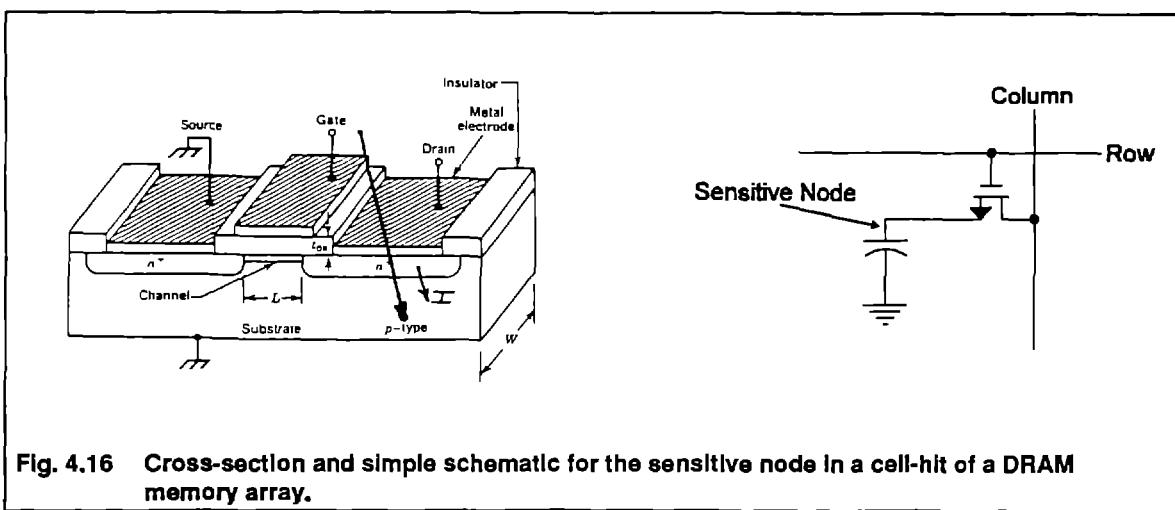


Fig. 4.16 Cross-section and simple schematic for the sensitive node in a cell-hit of a DRAM memory array.

$$Q_C = C \cdot V_{NM} \quad (19)$$

where Q_C [C] is the critical charge needed for an upset, and V_{NM} [V] is the noise margin at the node.

4.5.2 RMOS SRAMs

Resistive-load static RAMs (also called RMOS SRAMs) are designed to have circuit densities close to DRAMs, but to operate without the need for refreshing. Figure 4.17 shows the schematic of an RMOS RAM storage cell. The pull-up resistors, R1 and R2, have very high resistance values (on the order of 100's of MegOhms), so they do not affect the transient operation of the cell; these resistors only supply very small compensation currents to offset the MOSFET junction leakage currents.

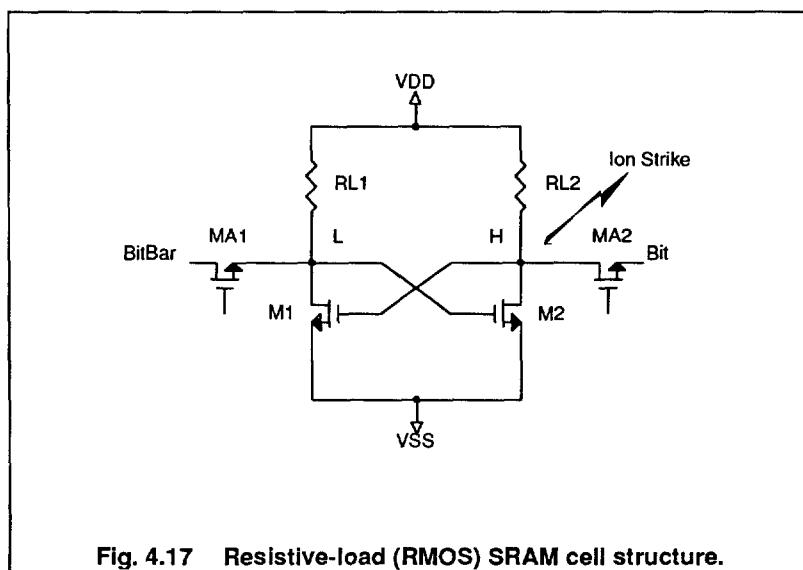


Fig. 4.17 Resistive-load (RMOS) SRAM cell structure.

The circuit arrangement of Figure 4.17 appears to be in a regenerative loop; however, due to the large values of the pull-up resistors, the time constant of the feedback loop is on the order of milliseconds, much longer than any operating transient signal or SE impulse in the circuit. A simplified circuit schematic of the cell which is appropriate for SE simulations is shown in Figure 4.18. As can be seen, the high-voltage side of the cell is roughly equivalent to a DRAM cell, where the storage capacitance is the inherent gate capacitance of the opposite MOSFET transistor. The binary information is stored as a high-low complementary voltage (charge) pair on nodes N1 and N2. Any perturbation of these stored charges will not recover for milliseconds, so can be considered persistent and can lead to a single-event error. Just as with the DRAM cell, the level of voltage (charge) degradation on the hit node which leads to an actual upset is dependent on the noise tolerance and sensitivity of the read sense amplifier and other surrounding circuitry. A maximum noise level

must be defined for the cell in order for the modeler to define a SE upset level. Note that an n-channel RMOS cell is only sensitive to direct upsets on the HIGH storage node, as the SE can only deplete the stored charge, not enhance it.

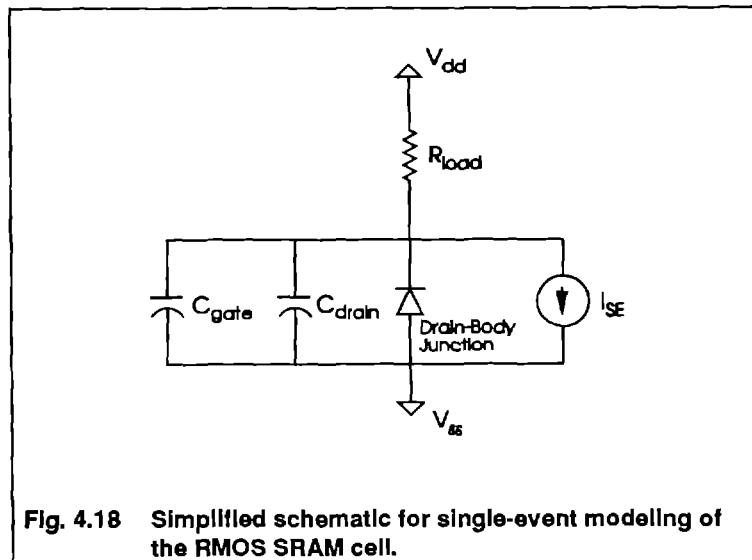


Fig. 4.18 Simplified schematic for single-event modeling of the RMOS SRAM cell.

Results of a SPICE SE circuit simulation of the RMOS SRAM cell are shown in **Figure 4.19** [Mas91]. The single-event strike depletes the stored charge from the HIGH-voltage node, leading to a degradation of the differential information voltage stored. As can be seen, the recovery of the perturbed node happens at times much longer than the nanosecond time frame of interest here.

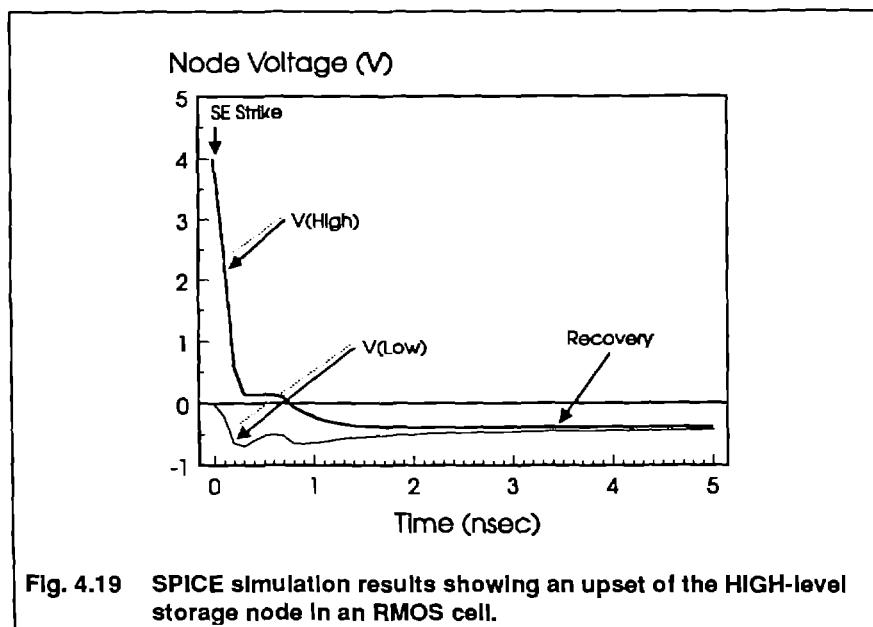


Fig. 4.19 SPICE simulation results showing an upset of the HIGH-level storage node in an RMOS cell.

Even though the hit node eventually recovers to a HIGH state, the possibility of a read operation while the node is depressed is very likely and must be anticipated. Thus, an error will occur when this read operation is fooled by the depressed voltage on the hit node, so this is considered an upset. By defining a minimum acceptable voltage for a successful read operation, the upset threshold can be defined. This is shown in Figure 4.20 as the hit node voltage (minimum value) on the high-side node as a function of increasing collected charge.

When the voltage descends below the noise threshold, the probability of an error is assumed (though not necessarily observed in all cases on the outside) and upset is defined.

Like the DRAM, the simplicity of the charge depletion and the direct relationship of the stored charge and the node voltages allows an approximate analytical model for the hit node voltage as a function of the collected charge. This analytical model is given by [Mas91]

$$V_F = V_I - \frac{I_M \tau_F}{C_D + C_G} \quad (20)$$

where V_F [V] is the final voltage on the node, V_I [V] is the initial voltage on the node, I_M [A] is the peak magnitude of the SE current pulse, τ_F [sec] is the fall time constant of the SE pulse, C_D [F] is the total drain capacitance, and C_G [F] is the opposite transistor gate capacitance.

Given a noise margin threshold which defines upset, the critical charge needed for upset is given by:

$$Q_C = I_M \tau_F = (V_I - V_{NM})(C_D + C_G) \quad (21)$$

where V_{NM} [V] is the noise margin the circuit can tolerate before an upset.

These analytical models assume a constant storage capacitance. This is somewhat valid, since part of the inherent capacitance on node N_1 is due to the fixed gate capacitance of MOSFET N_2 . However, the drain capacitance of N_1 and N_{A1} also contribute to this total capacitance, and these are voltage dependent. Therefore, a more accurate model of the SE hit to the RMOS circuit can be developed using a circuit simulation of the full schematic.

RMOS SRAMs are also sensitive to multiple-bit upsets (MBUs) [Son88]. Because of the long storage time of these cells, diffusion of charge can affect neighboring nodes to the hit junction. Experimental observation of this effect is shown in Figure 4.21. A modeling approach to this phenomenon is to use 2-d or 3-d transport codes to describe the diffusion of charge away from the hit location and to the adjacent nodes, so that a charge collection efficiency (of ratio of charge collected at node X relative to the hit node Y) is determined for all neighboring RAM cells. Then, circuit simulations, as described in the previous paragraph, can be used to find error results for each neighboring cell.

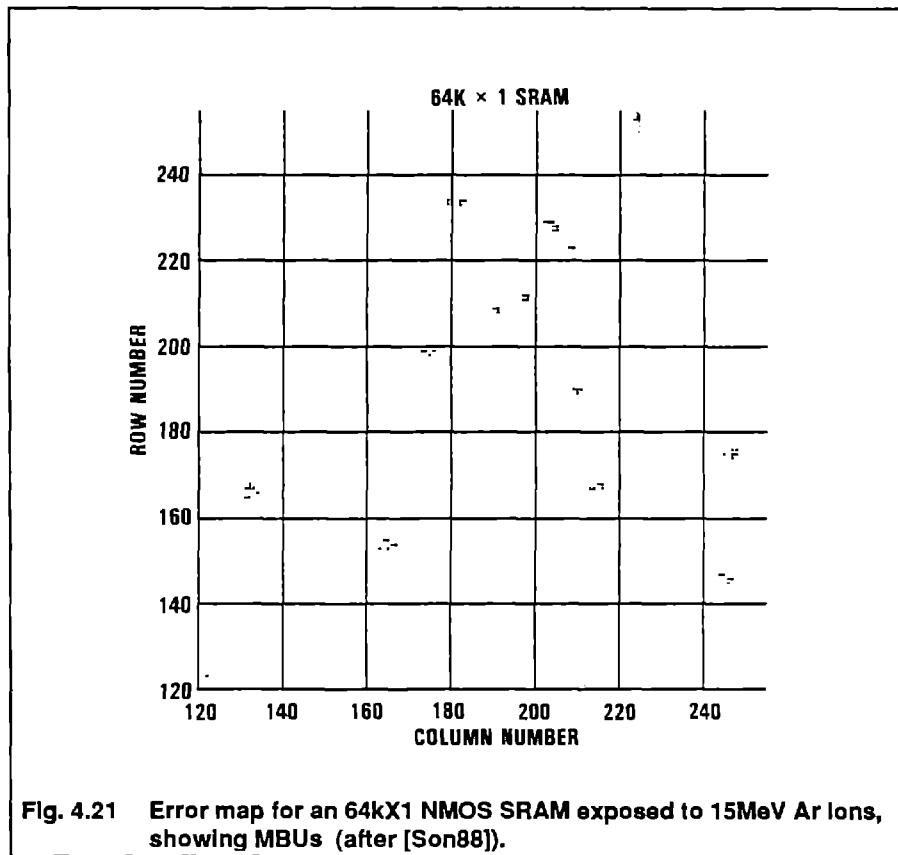


Fig. 4.21 Error map for an 64Kx1 NMOS SRAM exposed to 15MeV Ar ions, showing MBUs (after [Son88]).

The RMOS SRAM has another upset mechanism, the word-line upset [McD89]. A single-event hit to the circuitry which drives the word-line gates to the access transistors can cause a signal transient to propagate down a row of memory cells. A circuit model of this effect and simulation results are shown in Figure 4.22. If the magnitude of this signal is large enough, it can cause the access transistors to turn on, forcing an unwanted write of erroneous information into the affected cells. This phenomenon has been seen experimentally in some RMOS SRAM ICs [McD89].

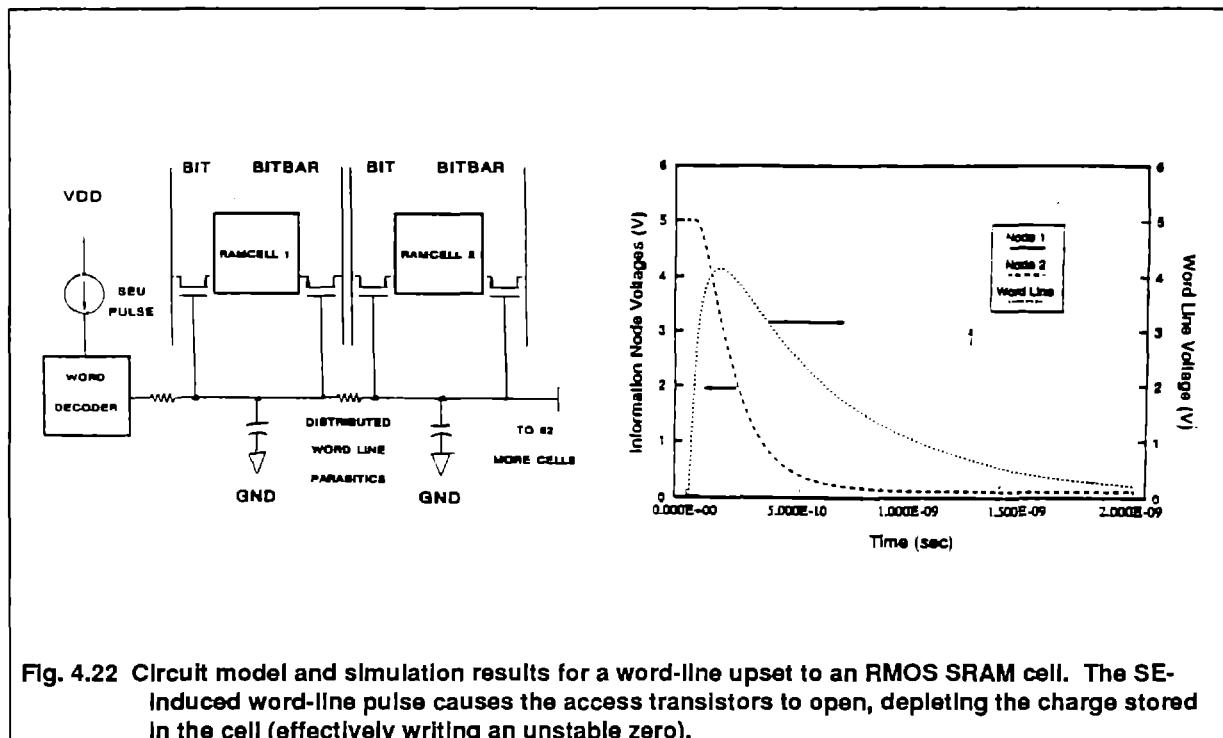


Fig. 4.22 Circuit model and simulation results for a word-line upset to an RMOS SRAM cell. The SE-induced word-line pulse causes the access transistors to open, depleting the charge stored in the cell (effectively writing an unstable zero).

4.6 Upset Models for Active Storage Devices

4.6.1 CMOS SRAMs

The single-event response of full, six-transistor CMOS SRAMs, as shown in Figure 4.23, comprise the basis of considerable study over the past ten years; there are several reasons for this. First, these cells are inherently harder to SE strikes than their DRAM or RMOS SRAM cousins, so there has been considerable interest in these memories for rad-hard military/space applications [Pic78]. Secondly, the

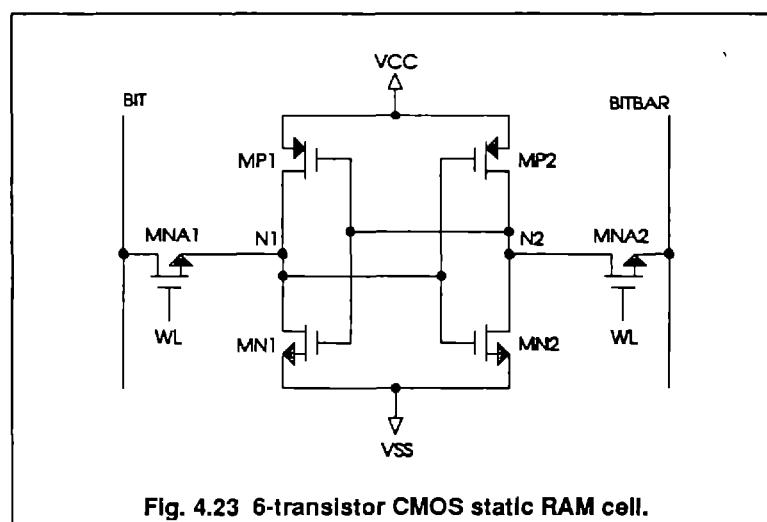


Fig. 4.23 6-transistor CMOS static RAM cell.

6-t SRAM was the first memory cell effectively hardened by circuit design techniques [Die82] to virtual single-event immunity. Thirdly, this cell design has been ported to other processing technologies (e.g. SOI) for even greater hardness to SEUs.

The cell of Figure 4.23 is basically two back-to-back CMOS (n-channel, p-channel) inverters forming a simple latch. It has two and only two stable states: a LOW voltage at node N_1 and a HIGH voltage at node N_2 , and vice versa. The regenerative, high-gain, positive feedback of each inverter driving the input to the other provides static (non-refreshed) operation and a high degree of noise immunity. In this discussion, we will assume that a binary '1' is stored by LOW-HIGH complementary voltages at nodes N_1 and N_2 , and a binary '0' is stored by a HIGH-LOW complementary pair. The particular physical arrangement of these voltage levels in relation to binary logic levels is arbitrary.

Since this circuit contains both n-channel and p-channel devices, it is susceptible to ion strikes in two places, as shown in Figure 4.24. In the figure, an ion strike to the HIGH-side n-channel drain

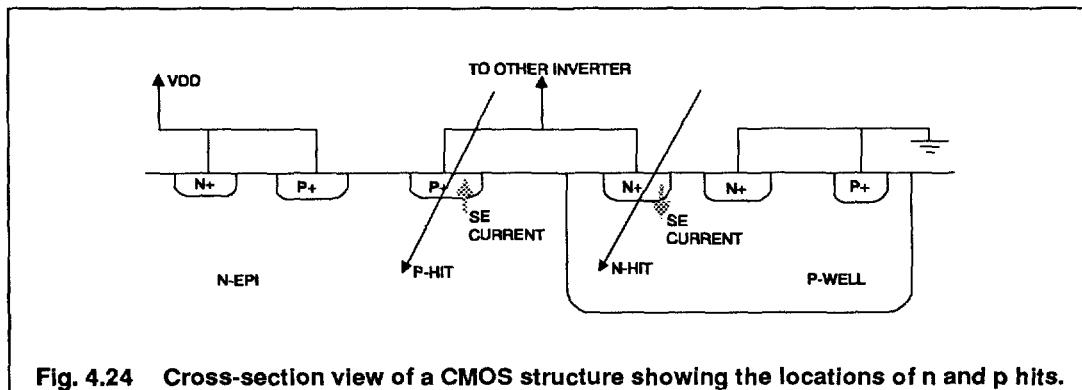


Fig. 4.24 Cross-section view of a CMOS structure showing the locations of n and p hits.

tends to degrade this voltage, while an ion strike to the LOW-side p-channel drain tends to enhance this voltage. Of course, the LOW-side/HIGH-side distinction is dependent on the binary bit stored at the time of the ion strike.

Single event modeling of this circuit involves the inclusion of current sources representative of the SE charge collection process at the sensitive junctions of Figure 4.24. A circuit model for these simulations is shown in Figure 4.25.

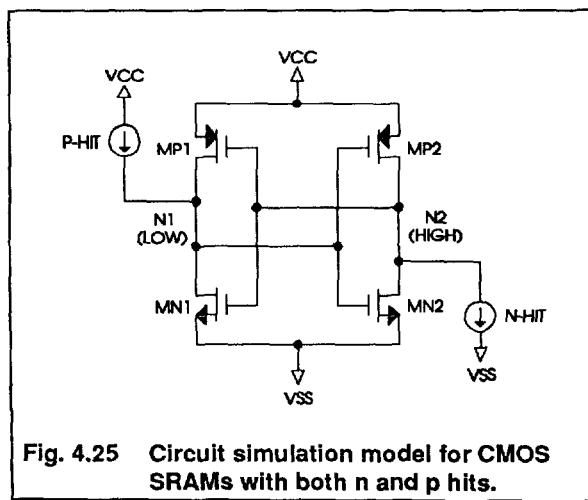


Fig. 4.25 Circuit simulation model for CMOS SRAMs with both n and p hits.

Unlike the passive storage cells discussed in the previous section, the CMOS SRAM cell will actively latch in one of its stable states following any disturbance of the critical nodes (N_1 and N_2). Thus, the definition of an upset is simply any perturbation which causes the cell to switch from one stable state to the other. Simulation results of two SE hits to node N_2 , one not causing upset and the other causing upset are shown in Figure 4.26. As can be seen in the latter case, the SE photocurrent perturbs the HIGH node to a level which causes the cell to regenerate to the other stable state. This regeneration process is very quick, and an externally observable error occurs.

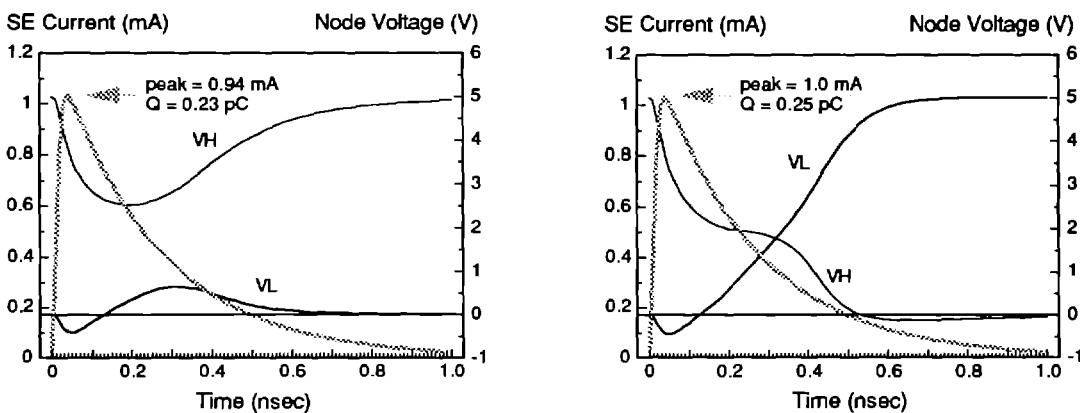


Fig. 4.26 Simulation results showing no upset (left) and upset (right) of the cross-coupled information nodes in a CMOS SRAM cell. In this case, the critical charge is 0.25pC .

The minimum charge needed to cause this upset, Q_C , for this example can be seen in Figure 4.26 as 0.25pC . Finding the Q_C for such a circuit is usually accomplished via iterative circuit simulations to determine the minimum current pulse magnitude (and resultant charge) needed to cause the upset. There exist no accurate closed-form analytical models for determining Q_C in these circuits.

The upset process in CMOS static RAMs is dominated by the regenerative current process. This is evident in Figure 4.26, as a very small change in the current pulse leads to very different results. Essential to accurate circuit modeling of upsets are: 1) accurate models for the dynamic (voltage-dependent) capacitances in the devices (this would include both junction and diffusion capacitances) and 2) accurate device current models for the transient regeneration currents [Mni83].

In very fast CMOS circuits, the equivalent time constant associated with the hit node can be on the order of the SE time constant. In these cases, accurate modeling of the SE time profile is essential to the upset results (see Sec. 4.2).

4.6.2 SOI SRAMs

Silicon-on-insulator (SOI) technology has become a choice technology for radiation-hardened electronics. The buried insulator effectively limits the amount of SE charge that can be collected at the critical junctions of the circuit, as illustrated in Figure 4.27. Since this is a rad-hard CMOS processing technology, standard CMOS designs can be ported to this technology, usually with an improved radiation response. The schematic for a 6-transistor RAM cell on SOI is identical to that of Figure 4.23 with each device fabricated on an isolated mesa of semiconductor surrounded by insulating material. Electrically, the only difference in bulk CMOS SRAM cells and SOI CMOS SRAM cells is the reduced source and drain junction areas of SOI and a common substrate region in bulk versus the isolated substrate (body) regions in SOI (see Figure 4.27). However, the SE modeling of these circuits is complicated by both the limiting geometry involved and the isolated body region under the MOS gate [Mas90].

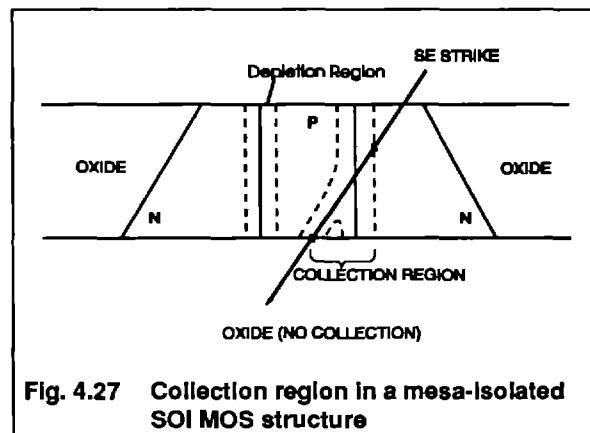


Fig. 4.27 Collection region in a mesa-isolated SOI MOS structure

A most critical aspect of the proper modeling of an SE strike to a SOI MOS device (either n or p-type) is the isolated body region. Whether this region is electrically floating, or connected to the source region by a low-resistance strap (body tie), any charge deposited in this region by a SE must either recombine or exit the region via three methods: (1) across the body-source junction, (2) across the body-drain junction, or (3) out the body-to-source tie. Because both junctions are reverse biased, most of the ionic charge exits the region via the body-to-source tie. This result, shown in Figure 4.28,

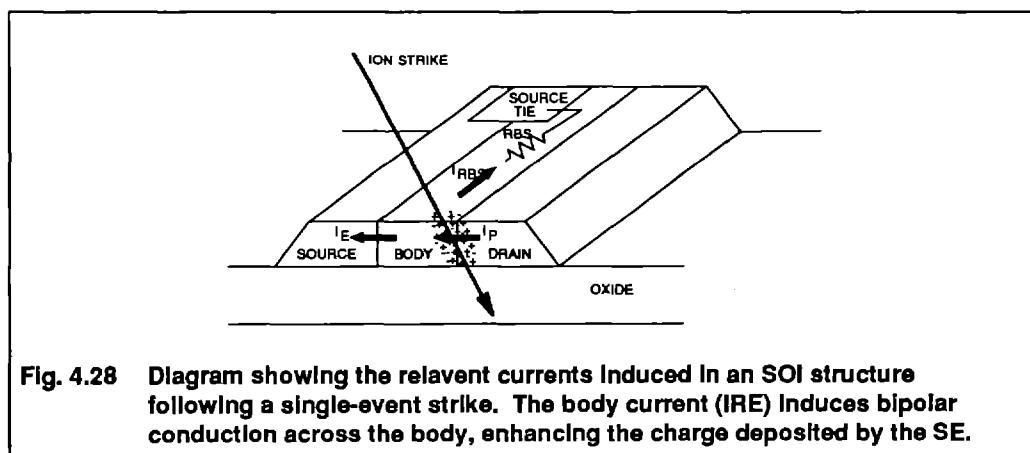


Fig. 4.28 Diagram showing the relevant currents induced in an SOI structure following a single-event strike. The body current (I_{RE}) induces bipolar conduction across the body, enhancing the charge deposited by the SE.

can lead to a potential gradient from the hit location to the body tie (which is clamped at V_{SS} by the source). If the potential near the hit is large enough to cause minority carrier injection across the body-source junction, then parasitic bipolar action can be created between the source and drain. The current created by the ion strike is in effect the base current for this parasitic bipolar transistor; this base current can be amplified to create a large collector current at the sensitive node. Thus, the SE current is amplified with the gain of the parasitic bipolar transistor. A lumped-parameter circuit model for this device response to the SE hit is shown in Figure 4.29 [Ker89].

Circuit simulation results for an SOI RAM cell is shown in Figure 4.30 [All90]. Since the parasitic bipolar current is in concert with the SE current affecting the hit node voltage, the effect enhances the effect of the ion hit, leading to an increased susceptibility to the ion charge (i.e. a reduced Q_C needed for upset).

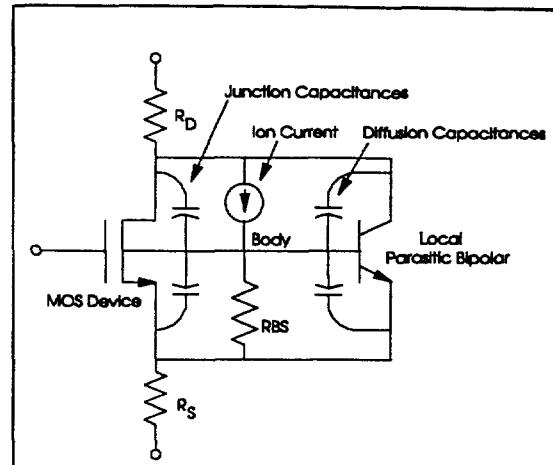


Fig. 4.29 Lumped-parameter SOI device model for single events.

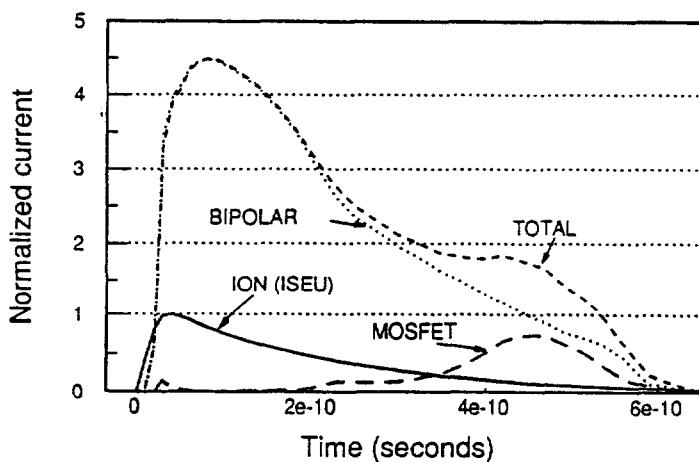


Fig. 4.30 Modeling results showing the relative contributions of currents to the total nodal current discharging the HIGH information node in the SOI structure of the text. The direct charge collection from the single-event-deposited charge is significantly enhanced by the bipolar action.

The SOI charge enhancement mechanism described here involves a 3-dimensional, distributed effect. The potential along the length of the body region varies with position, and the electrical currents then depend on this potential. Bipolar action is very localized; only a small portion of the body region is biased high enough to maintain bipolar conduction. In order to model the geometrical nature of the body region, a distributed circuit model has been used. The width of the body region is modeled by a distributed resistance/capacitance network and local bipolar devices. A single-event induced current pulse can be included at any point along the network to simulate hits at various positions along the width of the body.

Because of the three-dimensional nature of the SOI single-event problem and the complicated nature of the bipolar charge movements in the body region, single-event modeling of these circuits is amenable to approaches which combine macroscopic, circuit-level simulations with microscopic, device-level simulations of the SE hit and resultant charge movements. Some groups have used coupled device-circuit simulations to address problems like this -- the so called mixed-mode simulations [Fu85][Rol88a]. Another approach is to apply microscopic simulations to a specific device in order to quantitatively understand the properties of the device response to the ion hit. Then to hierarchically build on this information, developing an accurate description of the device to include in the full circuit simulation.

SOI technology in particular illustrates several of the powers of modeling presented in Sec. 1.2. The separation of direct charge collection from the ion track and the enhancement of that charge collection by the parasitic bipolar (as shown in Figure 4.30) would be essentially impossible at the experimental level. Yet, modeling yields clear insight into the importance of the parasitic bipolar effect on the upset susceptibility of the cell. Also, the effect of hit location on the critical charge for the cell is another property which would be very difficult to assess experimentally. Predictions about the effects of device scaling and/or hardening techniques on this parasitic bipolar action also emerge from the modeling. All of this information is illusive on the outside of the IC chip, but becomes clear with the modeling results.

4.6.3 Bipolar SRAMs

A cross-section view of a typical npn bipolar transistor is shown in Figure 4.31 and a static RAM circuit is shown in Figure 4.32 [Hau88][Zha89]. An interesting aspect of the SE modeling of bipolar circuits is that they possess several distinct regions sensitive to ion strikes [Zou88b], as shown in the figure. In the MOS circuits just discussed, the sensitive areas involved the drains of the devices (or the body region in SOI); here, single-events can strike the emitter, base or collector regions with different circuit results. This leads to several critical charge values for one RAM cell, depending on the strike location.

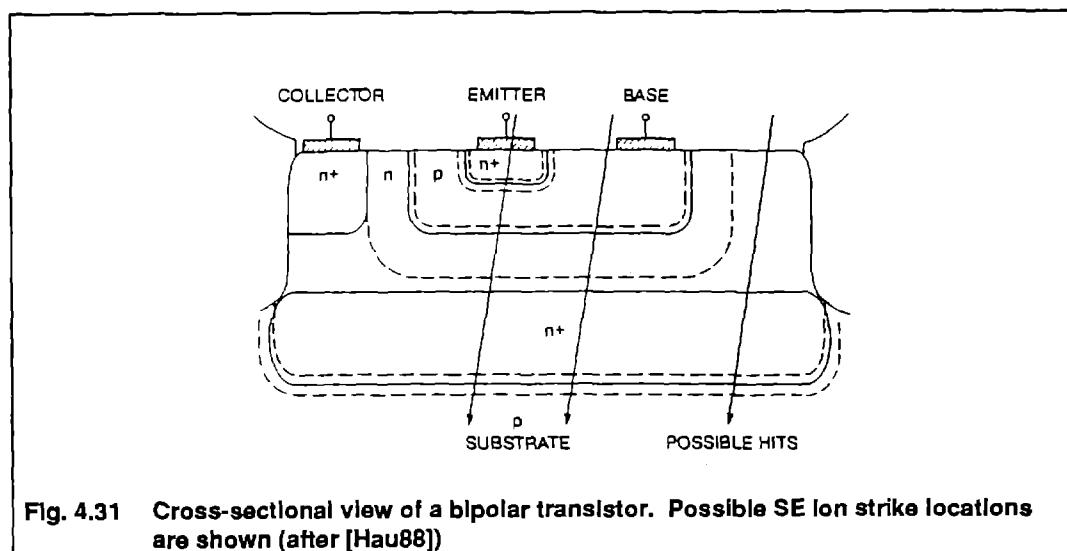


Fig. 4.31 Cross-sectional view of a bipolar transistor. Possible SE ion strike locations are shown (after [Hau88])

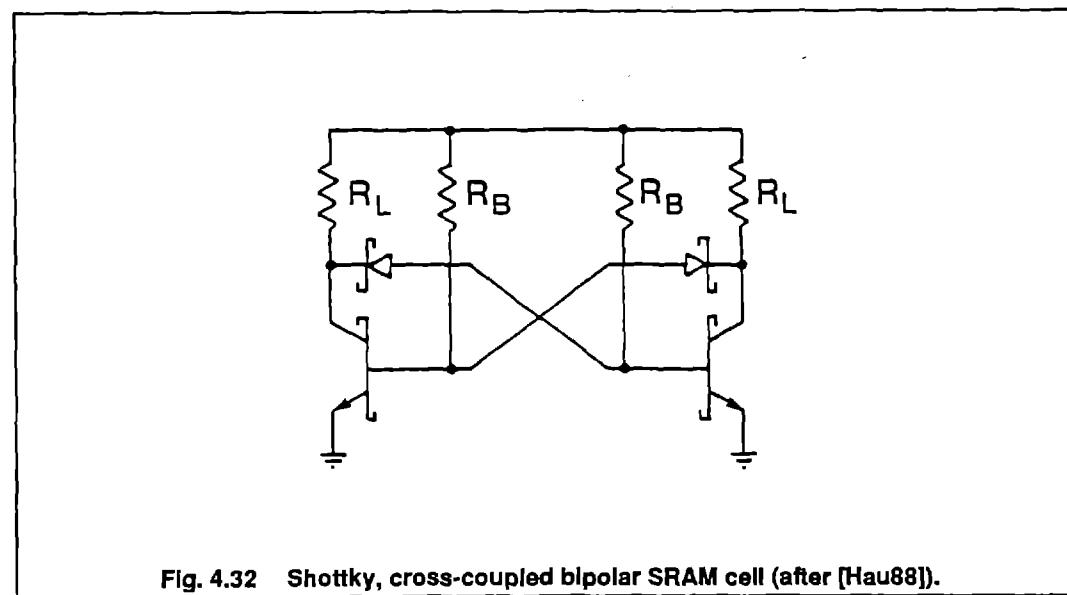


Fig. 4.32 Shottky, cross-coupled bipolar SRAM cell (after [Hau88]).

A circuit schematic for modeling ion hits to a single, isolated bipolar transistor is shown in Figure 4.33. This circuit can be included in the RAM cell circuit of Figure 4.32 to simulate the cell upset and find the SE current pulse (and Q_C) needed for upset.

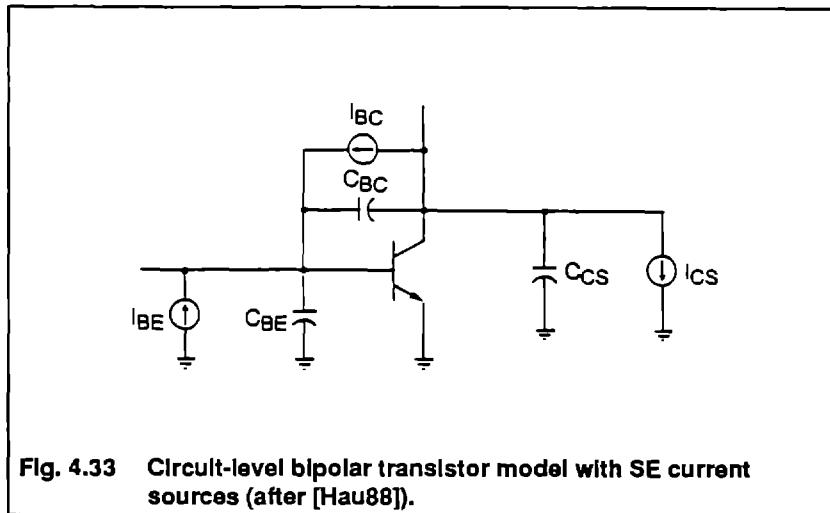


Fig. 4.33 Circuit-level bipolar transistor model with SE current sources (after [Hau88]).

4.6.4 GaAs SRAMs

Figure 4.34 shows a GaAs C-EJFET RAM cell. The cell is similar to the 6-T CMOS RAM, with junction devices for the inverters. Figure 4.35 shows the possible SE strike locations and the resulting photocurrents generated in the device model [Zul84][Zul88]. Since JFET gates are not insulated from the underlying semiconductor, as is the case in CMOS, a key issue in the modeling of GaAs upsets is the proper recognition of gate strikes [Wea85][Wea86]. An SE-induced photocurrent can be generated

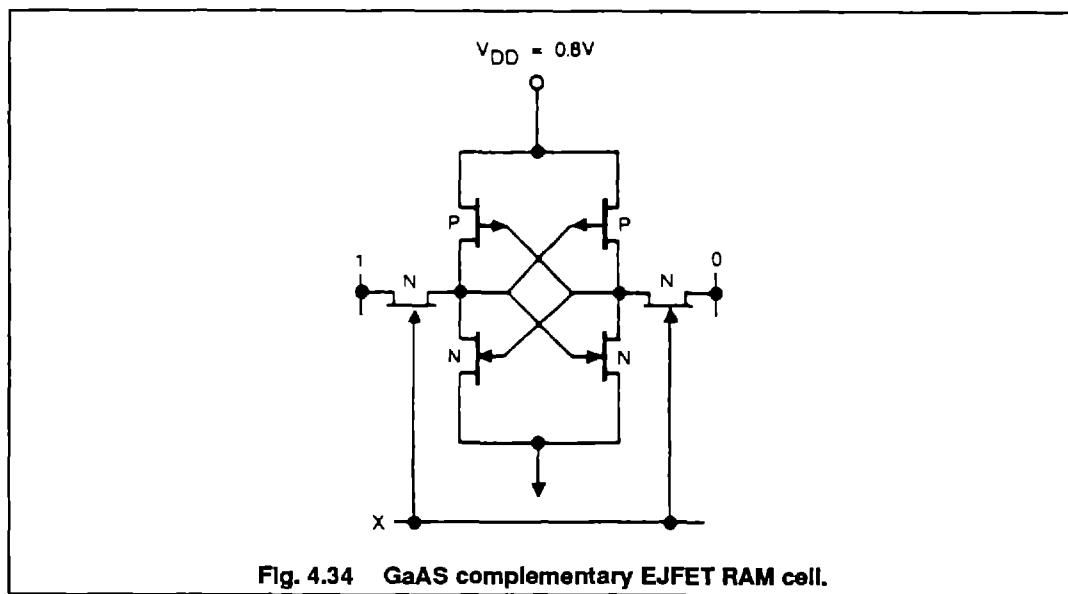


Fig. 4.34 GaAs complementary EJFET RAM cell.

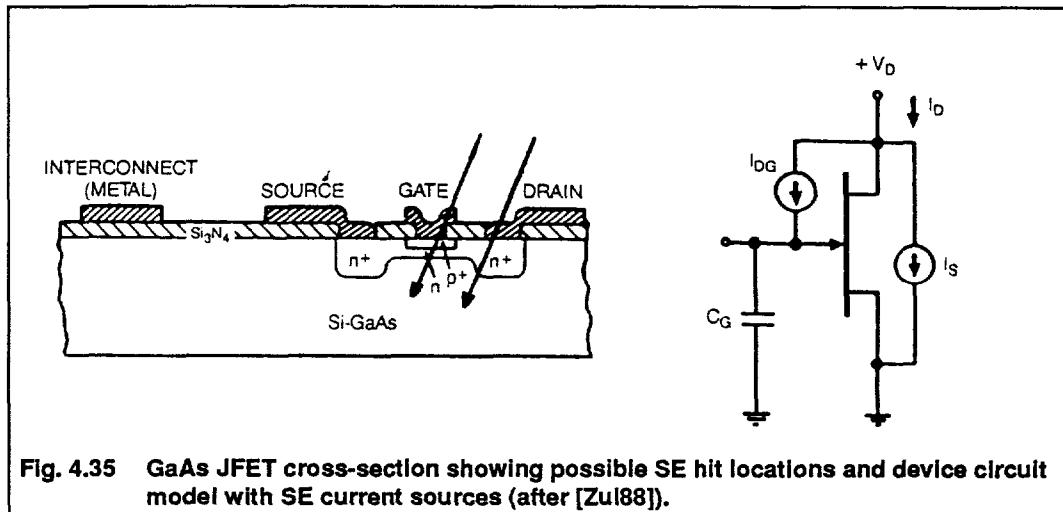


Fig. 4.35 GaAs JFET cross-section showing possible SE hit locations and device circuit model with SE current sources (after [Zul88]).

across the drain-substrate or across the gate-drain junctions. In Figure 4.36, the current paths for gate and drain hits are shown [Wea86].

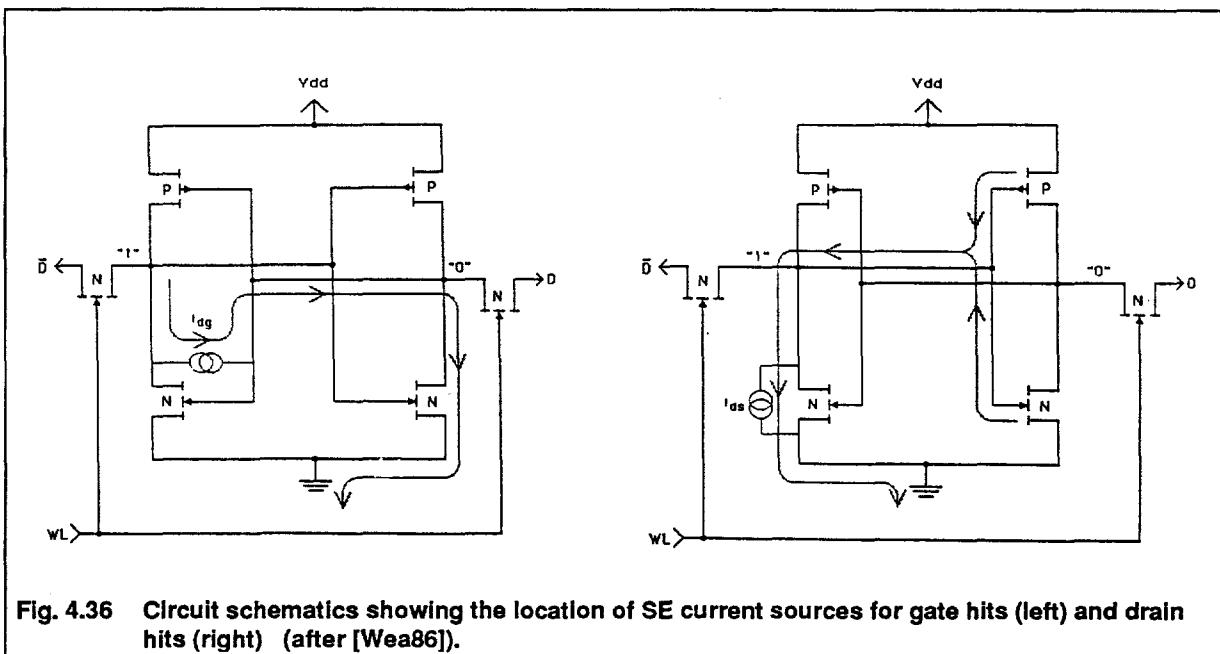


Fig. 4.36 Circuit schematics showing the location of SE current sources for gate hits (left) and drain hits (right) (after [Wea86]).

Other than the gate hit, the analysis of single-events in GaAs and the determination of critical charge follows the same procedure as that of CMOS, presented in Sec. 4.4.2. The modeler must be aware, however, that GaAs circuits can be fast enough that the shape of the SE current pulse is important to the circuit response. We have previously discussed the time constants of the circuit relative to the SE pulse; this is a case where just critical charge is not an entirely appropriate figure of merit. Analysis results should include the time profile of the charge delivery by the SE strike.

4.7 Transients in Combinational Logic

The analysis of single-event upsets in digital combinational logic is considerably more involved than that of memories or latches. The reason is threefold: (1) the circuits in combinational logic are not repeated or arrayed, in contrast to memories where one cell can be analyzed and the results applied to the whole array, so regularity assumptions do not hold (2) timing of the circuit's legitimate signals relative to the SE current pulse is crucial, race conditions exist and the temporal characteristics of the single event determine errors; (3) the many paths an error signal can follow in its propagation are influenced by the fact that only certain paths are active at any one time, depending on the input vector to the logic and the timing; (4) multiple errors can be generated by a single hit, requiring simulation of many error paths [Kau91]. Figure 4.37 shows a simple combinational circuit with a SE hit. The figure depicts the complexities listed above.

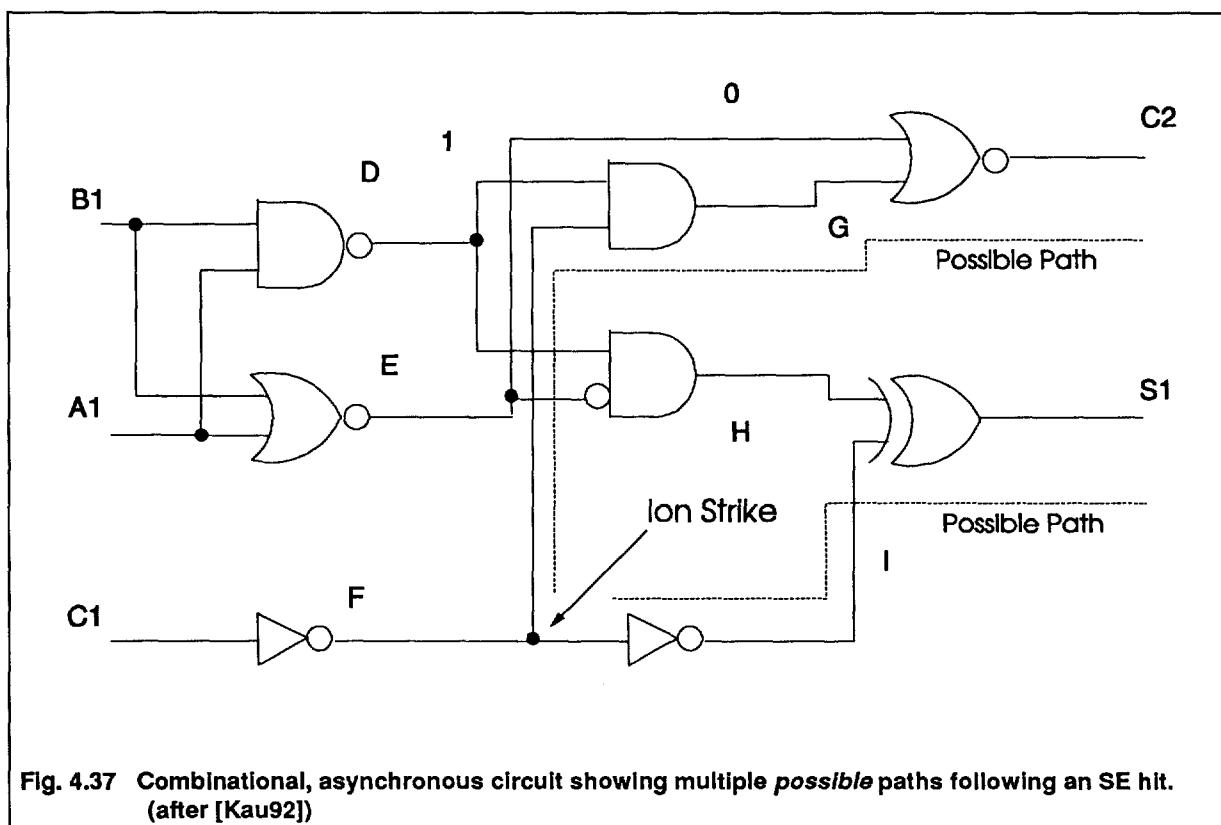


Fig. 4.37 Combinational, asynchronous circuit showing multiple *possible* paths following an SE hit.
(after [Kau92])

One approach to performing combinational logic SEU simulations is to use analytical methods to estimate error probabilities based on possible paths, as in [Die84]. This approach yields efficient results, but with a corresponding sacrifice in accuracy. Circuit-level simulations can be performed on all possible hit locations and resulting paths [Lea91], but the computational complexity is prohibitive

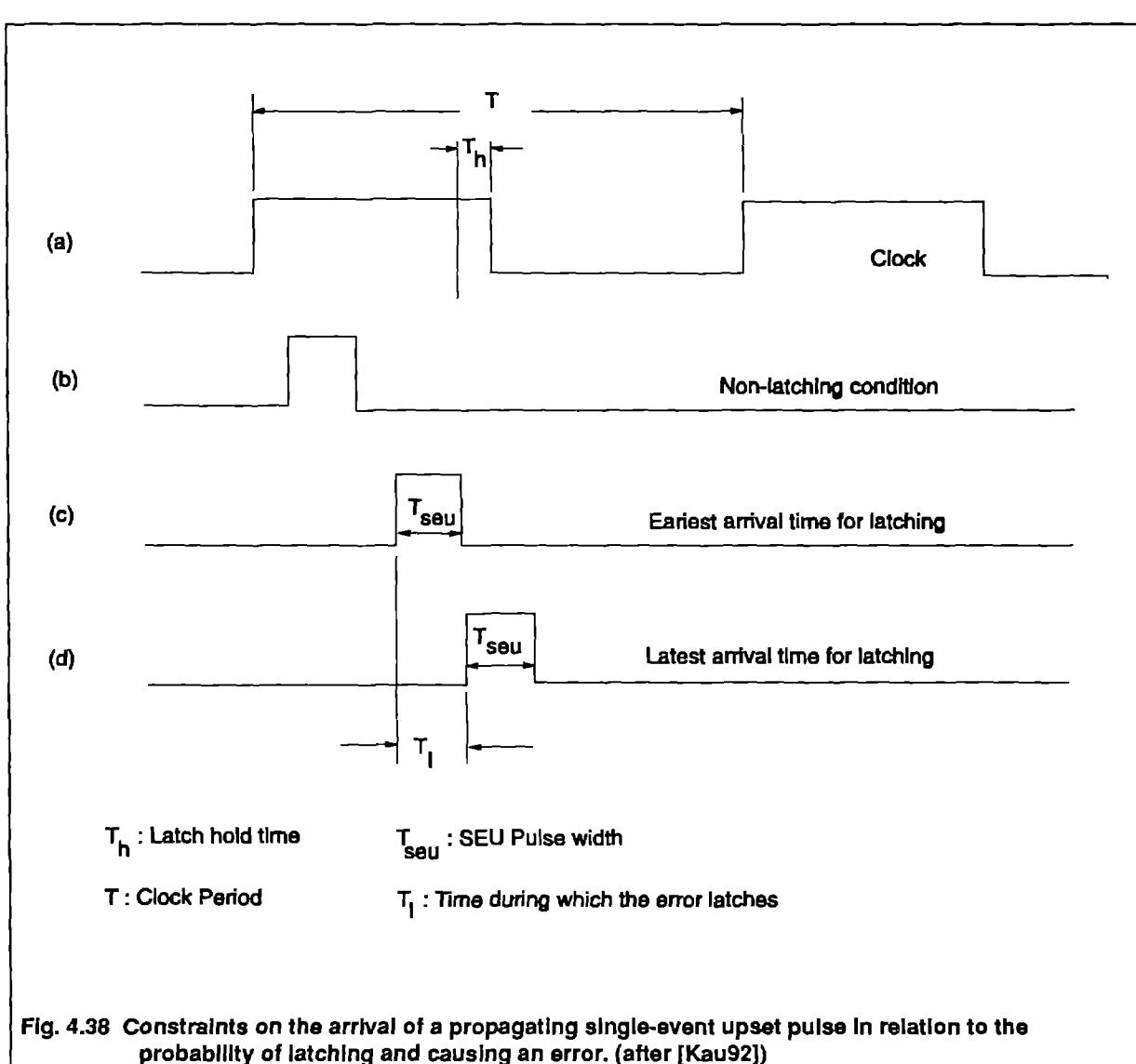
on complex logic structures. Li et al have simulated a TI SBR9000 microprocessor (I^2L technology) using a hardware description language (HDL) model of the microprocessor functional blocks and internal latches [Li84]. In that work, errors were introduced into internal latches. Error progression, as test software ran an instructional mix of opcodes, was monitored. While this approach effectively models the logic-level progression of errors through the system from latch to latch, it does not model the effects of an ion hit to random nodes causing signal glitches to propagate through the system.

A recent approach for this complex simulation task has been introduced by Kaul et al [Kau91]. This approach is to divide the analysis into several serial procedures. The first is to model the ion-hit as a voltage pulse at the affected node, while identifying the most vulnerable nodes in the circuit. These nodes will produce the largest voltage perturbations due to the SE hit. Circuit-level simulations are used to simulate the temporal pulse shape as it propagates through the connected logic paths. The pulse may decay, or may be amplified by subsequent gates. Properly modified switch-level simulations, which account for the pulse shape, can be applied to this part of the problem.

The next task is to determine the set of input vectors which facilitate the upset propagation through available signal paths. In order for the upset to become an error, it must propagate to either an output pad or an internal latch. This analysis task is similar to identifying the test vectors for 'stuck-at' and 'timing' circuit faults; many algorithms have been developed for estimating these vector sets. The most popular employ D-algorithm or back-tracing/back-tracking methods. Once the set of vectors facilitating the error propagation is obtained, the probability of their presence at the inputs when the ion hit occurred can be evaluated. Combined with the device-level simulation results at the hit node, the probability for an error generation and propagation can be estimated with these results.

If the error reaches a storage node, the possibility of latching involves the analysis of error pulse characteristics, latch parameters, and clock frequency as shown in Figure 4.38. This is a complex synergy of temporal characteristics, which cannot be evaluated in total detail. However, statistical probabilities of latching can be evaluated using random variables.

Once the error is stored in a latch, it may spread to the rest of the circuit during subsequent clock cycles. The effect of such a spread of errors, which is circuit topology dependent, must be evaluated to predict vulnerability. This is a logic-level analysis task; many logic-level algorithms have been developed for such error spread calculations.



While stating the necessary steps to model the conception and progression of a SE upset through a combinational logic system, as listed above, is relatively straightforward; the actual implementation of these steps is monumental. The development and combination of such analysis tasks as those presented above for the SE vulnerability of a complex combinational circuit (such as a microprocessor, microcontroller, or ALU) is an area of on-going research. As yet, no complete simulation methodology exists for SE upsets in combinational logic.

4.8 Other Circuit Types

The vulnerability analysis of other types of circuits to SEUs, such as analog subcircuits [Tur90], CCD imagers [Lom90], nonvolatile RAMs, and sensors, are beyond the scope of this work. As yet, none of these areas have seen extensive SE modeling. However, future space systems may require accurate assessment of the vulnerability of these important subsystems, so modeling in these areas may soon mature.

4.9 Recap of Circuit Modeling

This section introduced the concept of circuit-level modeling of single event effects. The section presented the engineer with a circuit description of the SE current pulse which can be used in SPICE-like simulations of SE hits. In order to measure the vulnerability of a particular circuit, critical charge was presented and defined as the minimum charge deposited on a critical node which causes a circuit upset. The engineer was presented the important issue of the circuit time constant at the collection node, and its constraint on the applicability of Q_C as a valid SE measure was discussed.

The definition of upset as a corruption of stored or propagating information was contrasted to the definition of error as the misinterpretation of that corrupt information by the other circuitry. Destructive errors (hard errors) due to single events were briefly outlined, then nondestructive errors in both passive (charge) and active (regenerative) storage circuits were discussed. Several families of circuits were presented with appropriate circuit models for upset analysis. Finally, transient upsets in combinational logic (non-storage) circuits were discussed.

The ultimate result of this circuit modeling is a vulnerability measure for any particular circuit to a single-event hit. This measure may be in the form of critical charge (Q_C) for circuits with long integration times compared to the SE time constant; or in the form of a more complete charge collection profile for circuits with short integration times compared to the SE time constant. This measure of a circuit's upset threshold is used in subsequent sections in the prediction of the error rate of the part in a space environment.

5. GROUND-BASED EXPERIMENTAL SEUS

This section will very briefly touch on the use of ground-based testing to characterize the response of IC chips to heavy ions: in effect, 'modeling' the response of these circuits expected in space environments. Volumes of material could be compiled on the many excellent and enlightening results obtained by heavy-ion bombardment of ICs in the laboratory, unfortunately this lies beyond the scope of this course. In addition, many excellent treatises on these aspects of single-event study have been written, including the 1992 NSREC short course by Sexton [Sex92]. The reader is referred to the literature for a more complete discussion of experimental single-event studies.

The purpose of this section is to introduce the basic results of experimental SEU testing, so that (s)he will be familiar with the terminology and can compare/verify modeling results with experimentally observed upsets levels.

5.1 Measured LET_C versus Simulated Q_c

The typical laboratory test of an IC for SEUs is to bombard the part with a flux of particles of controlled LET, in order to monitor the errors seen at the outputs. For RAMs, a pattern is stored in the memory, the chip is bombarded, and the pattern is checked - giving a count for the number of errors generated by the particles. The effective LET of the particles can be varied by changing the particle species, the energy, or the angle of incidence with the IC under test.

An idealized measurement of upsets in a 16k SRAM is shown in Figure 5.1. In this figure, the beam has no effect on the errors seen at the output pins until a critical LET is reached. At that point, the particle causes enough charge to be collected to upset a RAM cell. That is, the collected charge from the particle equals the critical charge of the circuit. Thus, the cell (and all of its identical sister cells in the memory, which are also being hit by identical particles from the beam fluence) will upset, and we see the total number of upsets increase to 16,384

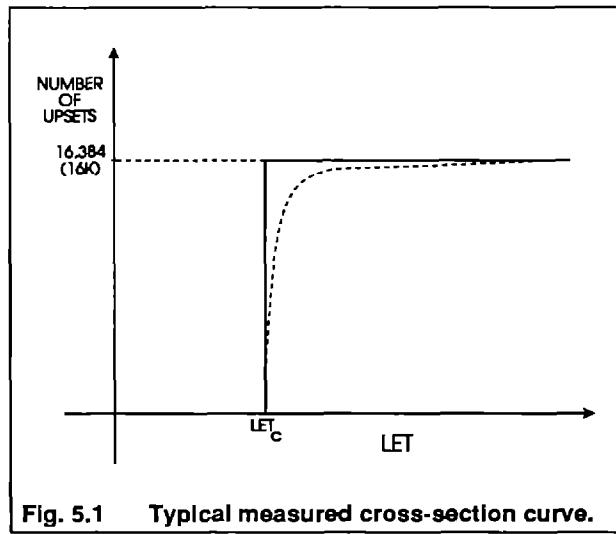


Fig. 5.1 Typical measured cross-section curve.

(every cell in the memory). The critical LET (LET_C) measured experimentally can be related to the critical charge (Q_C) for the circuit, and the terms are often used interchangeably, but they are not conceptually equivalent.

5.2 Measured Cross-section

Of course, the results of Figure 5.1 assume that every cell in the memory receives at least one particle from the beam fluence. If the beam fluence is lower than this 'one-hit-each-cell' value, not all cells will be struck by the beam and have the opportunity to be subsequently upset. To account for the cells not exercised by the beam, the beam fluence is usually normalized out of the measured upsets:

$$\sigma = \frac{N}{F \cos(\theta)} \quad (22)$$

where N is the total number of upsets observed, F [particles/cm²] is the beam fluence, and the $\cos(\theta)$ accounts for the fluence variation with angle of incidence. This parameter has units of area (cm²) and is called the **upset cross-section** for the circuit. In most cases, the saturated cross-section corresponds to the sensitive (or SEU vulnerable) area of the circuit, and is often used as such in the error-rate prediction codes of Sec. 6.

5.3 Limitations on the Interpretation of Experimental Results

Care must be exercised when comparing simulated and measured critical charge (LET) and upset cross-section (sensitive area). Nonideal effects, especially in modern scaled devices, complicate the direct comparisons of the measured and modeled parameters. For example, the approximation of a single sensitive area is not valid in circuits where the critical charge depends on the hit location (see Sec. 4.6.2). Also, statistical spreads in the circuit and process parameters across a chip can cause 'identical' cells in a memory to have different critical charges, so a single critical charge for the chip is not valid.

Another limitation is the assumption that LET can be varied via angle of incidence of the particle. That is, a particle with a certain LET at a high angle of incidence can be equivalenced to a normally-incident particle with a higher LET by the form:

$$LET_{eff} = \frac{LET}{\cos(\theta)} \quad (23)$$

where LET_{eff} is the effective LET that a normally-incident particle would have to deposit the same total charge, LET is the LET of the actual particle, and θ is the angle of incidence measured from the normal. This assumption obviously falls apart at high angle of incidence, since an LET_{eff} of infinity is predicted at $\theta=90$ degrees. However, it also has questionable validity in modern IC technologies where drawn dimensions are very small. When collection regions are very closely packed, a particle with an appreciable angle of incidence spills over into a nearby region, or may be truncated by boundary regions. Thus, the equivalence to a particle with normal incidence is not valid.

A discussion of these and other nonideal effects is beyond the scope of this tutorial, but can be found in the literature. The modeler should be aware that nonideal effects exist when comparing model results with beam experiments.

6. PREDICTING SPACE ERROR RATES

So far, we have studied models for the deposition of charge by an ion, the collection of that charge at a circuit node, and the modeling of the effect of that charge on the operation of a microelectronic circuit. We have introduced the concept of critical charge, which is the minimum amount of charge needed at a sensitive node to cause an information upset. So, critical charge tells us about the vulnerability of the circuit, but it does not tell us anything about the number or rate of errors we would expect if we put a particular part into a particular hostile space environment.

In order to extend the modeling tools presented here to the prediction of actual errors expected in the field, we introduce the concept of environmental **error rate predictions**. Conceptually, the technique of predicting error rates progresses along the following steps.

- (1) determine the flight environment (altitude, inclination, orbit, shielding, etc.)
- (2) determine the distribution of ions and energies in that environment (how many ions of each energy are expected)
- (3) determine the *sensitive area* and the *critical charge* for the circuit of interest (by simulation or experimentation) and extrapolate to the total *sensitive volume* for the circuit (within which charge is collected at the sensitive node),
- (4) determine the number of ion strikes (of each type and energy) which would hit the sensitive area per day
- (5) determine the percentage of the strikes of (4) (on average) which would cause enough charge to be collected to cause an error (based on the critical charge of the circuit)
- (6) the result of (5) is the errors/day expected for the particular circuit or subcircuit. If this subcircuit is one cell of a memory, for example, the result would be errors/bit-day.

A part in a real environment is bombarded by radiation from all directions with LET values over a wide spectrum. Any one of these particular hits might deposit enough charge in a sensitive volume to

cause the circuit to upset and produce an error. Whether a particle deposits enough charge to cause an upset is dependent on both the LET of the particle and its particular random path length (also called chord length) through the collection volume. That is, both a high enough LET ($\text{pC}/\mu\text{m}$) and a long enough path length (μm) are needed to deposit Q_C (pC) in the volume. But both of these quantities are random variables in the space environment -- we don't know when a particle will hit any particular collection volume, what the particle energy (LET) will be, or at what angle the particle will hit and penetrate (giving the path length in) the volume.

When facing random variables such as these, we turn to statistics and begin talking about probabilities of each of these events, rather than actual events. We don't know precisely when a particle will hit a sensitive volume, but based on the number of particles in the environment, we know the number of times, on average, that the volume will be hit per unit time. We don't know the LET the particle may have when it hits the volume, but based on the distribution of LETs in the environment, we can know the percentage of particles with each specific LET value and thus the probability that the volume will be hit by a particle of any specific LET. We don't know the precise path the particle will take through the sensitive volume (thus its path length), but we can analytically model the total distribution of paths (chord lengths) in the volume which are long enough to deposit Q_C based on the LET. Error rate analysis involves the use of these probabilities to predict the number of upsets / unit time expected (given a sensitive volume and a critical charge) based on all the possible combinations of LETs in the environment and chord-lengths in the sensitive volume which could deposit Q_C and cause an upset.

The calculation of error-rate for one memory cell is a summation over all particles in the flight environment of the combination of these percentages (probabilities). One way of stating this summation is:

$$\text{ErrorRate} = A \cdot \sum_{\substack{\text{over} \\ \text{all} \\ E_i}} \left[\begin{array}{l} (\text{total flux hitting the projected sensitive area}) \times \\ (\text{probability a particle has energy } E_i) \times \\ (\text{probability the particle hits in such an angle to produce} \\ \text{a path length equal to, or greater than, needed to} \\ \text{deposit } Q_C) \end{array} \right]$$

where A is the total projected area of the collection volume to the incident flux.

In this form, the energy distribution (LET distribution) is a differential spectrum. The chord length distribution is an integral distribution since all the chord length greater than the length needed to deposit Q_C will also cause upset and must be counted.

Unfortunately, since the flux spectrum is a measured distribution of the space environment, it can possess numerical noise, discontinuities, and other quantization properties as a differential spectrum (it's real bumpy). However, these numerical effects can be reduced and the measured curve smoothed if an integral flux spectrum is used, and this is common practice. Mathematically, the differential LET and integral chord-length distributions can be reversed to integral LET and differential chord length distributions with equivalent results in the summation. This formulation of the error rate is then:

$$\text{ErrorRate} = A \cdot \sum_{\substack{\text{over} \\ \text{all} \\ \text{path} \\ \text{lengths}}} \left[\begin{array}{l} (\text{total flux hitting the sensitive area}) \times \\ (\text{probability a path length in the volume is equal to } l_i) \times \\ (\text{probability a particle along that track has energy equal to,} \\ \text{or greater than, } E_i \text{ needed to deposit } Q_C) \end{array} \right]$$

Since we can relate path length with LET via the critical charge (see Eq. 1), this summation can also be carried out over LET values. In a simplified mathematical form, this is:

$$\text{ErrorRate} = A_P \int_0^{\infty} F(\lambda) P(\lambda) d\lambda \quad (24)$$

where ErrorRate is the **upset rate** in upsets/sec for one sensitive volume, A_P is the projected sensitive areas [in cm^2], $F(\text{LET})$ is the integral flux of particles [in particles/ $\text{cm}^2\text{-sec}$] in the environment as a function of LET, and $P(\text{LET})$ is the differential distribution of chord lengths in the collection volume which deposit Q_C as a function of LET [in units of LET^{-1}]. Note that $P(\lambda)$, the integral flux, is the flux of particles with LET equal to or greater than λ . $P(\lambda)$ is the differential of path lengths in the volume, which given an LET of λ , can deposit Q_C charge. Integrated over all LET values in the environment spectrum, this gives the total error rate for the collection volume.

If the collection volume used in Eq. 24 represents the sensitive volume for one bit in a memory array, the total error rate would be the summation of all contributions from each of the bits.

Figure 6.1 shows a schematic flow chart of this classical error-rate prediction technique. The following sections will refer to this chart and to the mathematical form of Eq. 24 [Pic81][Pet82][[Mas92]].

Of course, the diagram of Figure 6.1 and the preceding description is a gross simplification of the error-rate calculation. However, it does give a qualitative feel for the procedure. In particular, the measured space spectrum is in units of particles/(solid-angle)(sec) since the space flux is isotropic, impinging on the volume from all angles. Proper linkage from this solid angle to the actual parallelepiped collection structure must be carried out by appropriate surface integration; this is not included here for simplicity of the qualitative presentation.

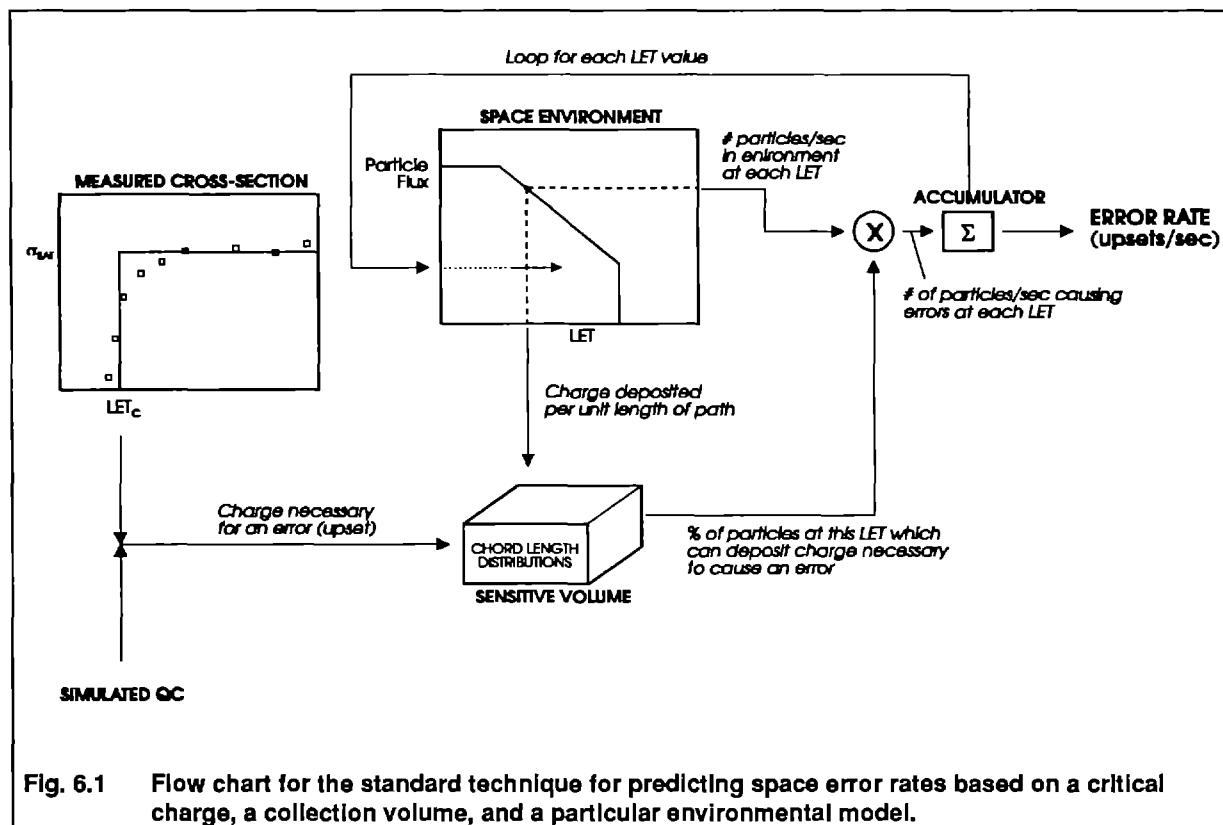


Fig. 6.1 Flow chart for the standard technique for predicting space error rates based on a critical charge, a collection volume, and a particular environmental model.

6.1 Models for the Orbital/Space Environment – $F(\lambda)$ function

Data sets from many satellites have been compiled over the years to construct statistical models for the space environments. The modeler can make use of these databases as inputs to his/her error-rate calculations.

Proton Belts: the most common model for trapped protons is the AP8 model developed by the U.S. National Space Sciences Data Center (NSSDC) at the NASA Goddard Space Flight Center.

Solar Cycle: historically, the solar cycle has been described as a sinusoidal function of 11 year period [Ada87]; recently this model has been improved by Nymmik [Nym91] to account for the variations in the peaks of this cycle. **Solar Flares:** solar flares can be predicted based on the empirical model of Stassinopoulos and King [Sts74], or on a statistical simulation model by King [Kin74]. Also, Stassinopoulos has developed a predictive model, SOLPRO, based on satellite measurements of flares.

Models for cosmic ions are based on the work of Adams et. al. who have developed spectrums for heavy ions at various orbits [Ada83]. In order to keep these spectrum data manageable (since the cosmic spectrum consists of limitless different particles and energies), these data are most often combined into a form introduced by Heinrich, with particles grouped by LET value. Such a Heinrich integral spectrum is shown in Figure 6.2.

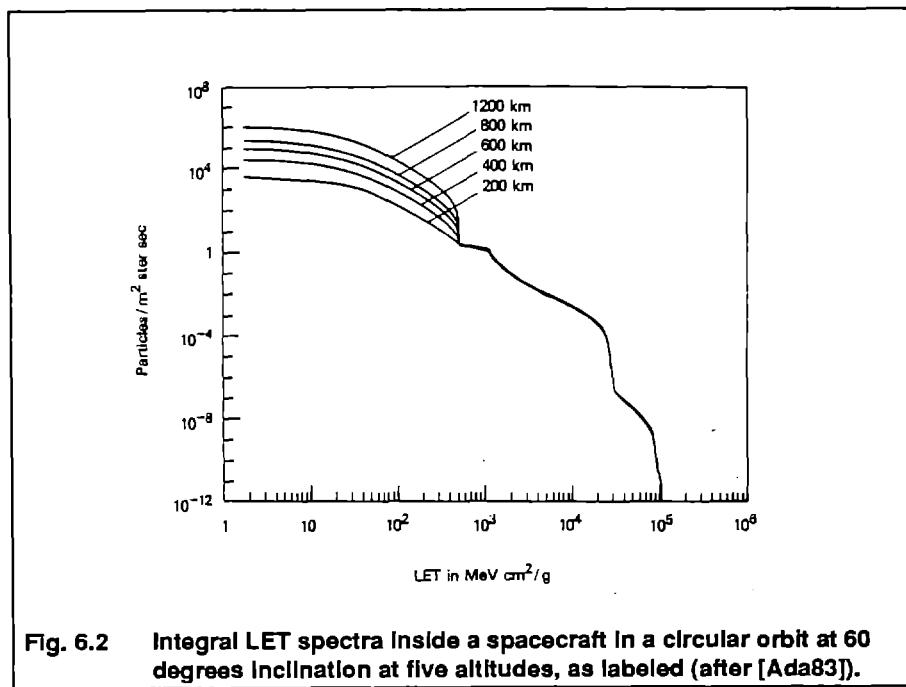
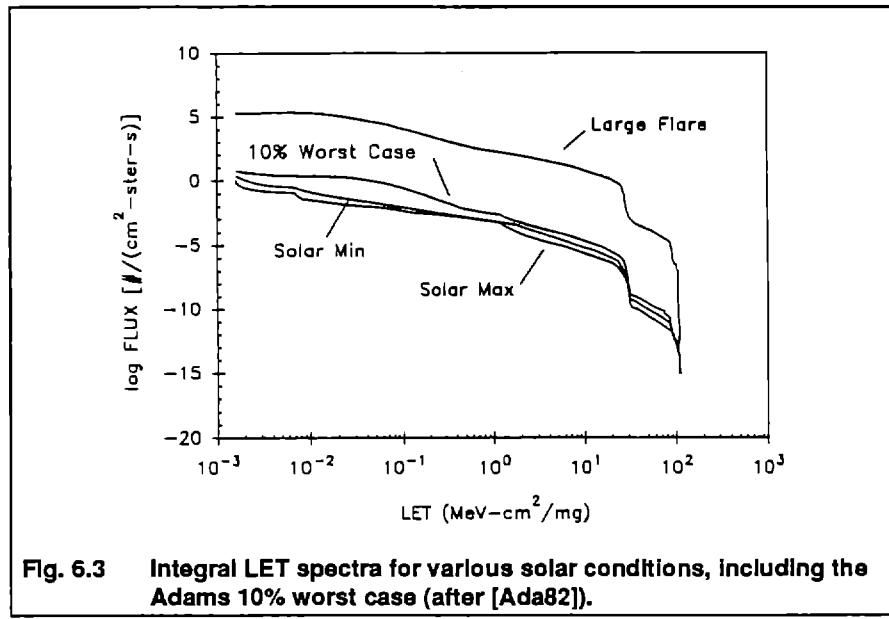


Fig. 6.2 Integral LET spectra Inside a spacecraft in a circular orbit at 60 degrees Inclination at five altitudes, as labeled (after [Ada83]).

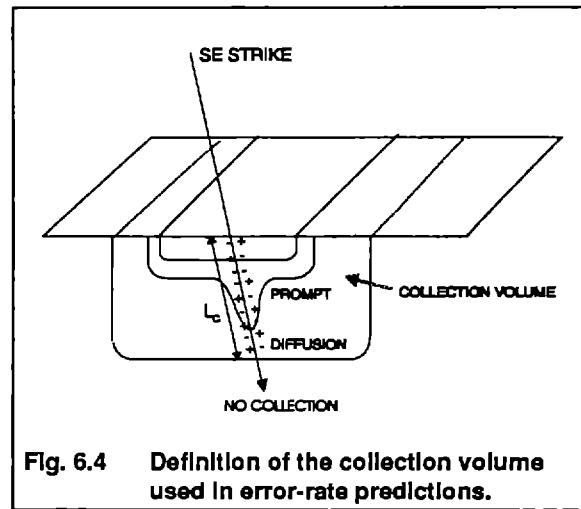
Adams has used historical solar cycle and flare histories to produce a Heinrich spectrum in which the actual galactic environment is more intense only 10% of the time [Ada82]. The so called 'Adams 10% worst case' spectrum is shown in Figure 6.3. This spectrum has become a defacto standard environment for error rate calculations and comparisons over the years. Recently, however, Petersen and Pickel have recommended that a solar minimum spectrum be used for more reliable error rate calculations [Pet92].



6.2 Sensitive Area, Sensitive Volume, and Critical Charge

Given a particular circuit of interest, say a RAM cell, the critical charge and sensitive area can be determined by simulation, as described in Section 4; or extracted from experimental results (in the ideal case), as described in Section 5. The collection volume for a particular sensitive node must be determined.

Approximating the collection volume is a precarious process, since this volume describes the inclusive region surrounding the sensitive node within which the node collects charge from the SE strike. It is assumed that any track length outside this volume does not contribute to charge collection at the node, as shown in Figure 6.4. This volume must include the effects of funneling, if present, and diffusion collection, if important to the circuit of interest. The difficulty in determining the appropriate collection volume lies in the fact that funneling and diffusion of charge to the collecting node depends on the ion stopping power; thus, the collection volume is not constant, but varies with ion type and energy. Several techniques have been developed to account for funneling in error-rate calculation using empirical models.



Another limitation on the collection volume concept is that the collecting region in a modern IC is not a parallelepiped, but is a complex, 3-d structure [Abd87]. Defining a simple collection volume for a typical microelectronic structure is difficult.

More research is needed into the effect of non-ideal geometries on accurate error-rate calculations, and much of this work is presently ongoing.

6.3 Chord-length distributions in the sensitive volume -- $P(\lambda)$ function

When an IC is in a space environment, it can be struck by a cosmic ion from any direction. The probability of any particular vector is assumed to be fairly random. Recall that a particle LET describes its charge deposition per unit length, so the amount of charge deposited by a particle track in a particular volume depends on the path length in that volume. Some of these random penetration vectors are long enough that sufficient charge is delivered and collected by the sensitive volume to cause upset of the circuit (i.e. Q_C is deposited and collected), while other vectors are not, as shown in **Figure 6.5**. So, for a given particle in the environment, the probability that it will cause an upset if it penetrates the sensitive volume depends on the probability it will strike with the proper vector to give it a long enough path length in the volume. The distribution of path lengths (or chord lengths) in a rectangular parallelepiped is a well characterized analytical function, which can be evaluated in closed form [Bra79][Pic80]. This function can be used to describe the probability, given a random path through a rectangular volume, that the path length will be equal to (within the incremental $d\lambda$ of) any predetermined length. This distribution is the differential chord-length distribution of the collection volume.

If one knows the charge needed within the volume (Q_C) to cause an error, and the charge deposition rate of a particular particle; then the chord length integral can be used to determine the probability that

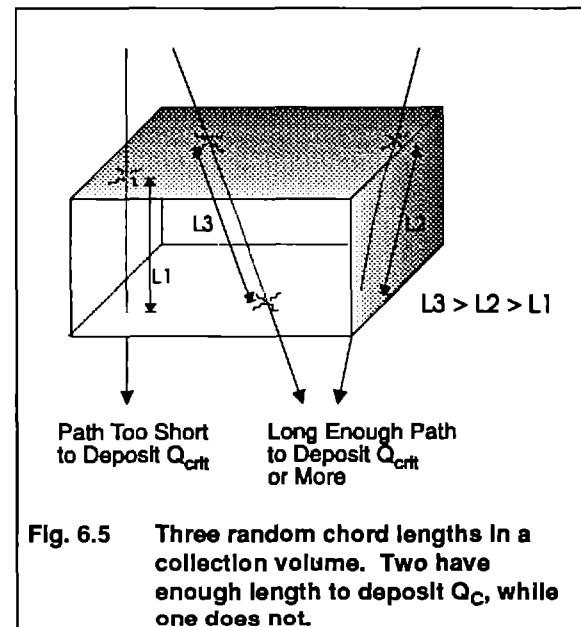


Fig. 6.5 Three random chord lengths in a collection volume. Two have enough length to deposit Q_C , while one does not.

a random path of the particle through the volume will be long enough to deposit enough charge to cause an error.

6.4 Single-Threshold Error-rate Calculation

Upset rate prediction codes, such as NRL's CRUP or CREME, perform the function flow in Figure 6.1 with the requisite chord-length integrations for a specific environment flux to predict the number of errors for that sensitive volume per unit time (usually error/bit-day).

These error-rate analysis models assume one fixed critical charge and one well-defined collection volume for the circuit of question. Both of these are questionable assumptions when applied to modern IC technologies, especially rad-hard designs [Mas92] [Pet92] [Mas93].

6.5 Integral Error-Rate Prediction Technique

The preceding analysis technique determines the upset error rate for one particular subcircuit. In most cases, such as memories, it is assumed that all the subcells are identical, and the results of one are applied to all to arrive at the error rate for the entire IC chip. However, this ideal case is not always representative of real circuits. It has been observed in modern memories that parameter distributions across the IC chip can lead to disparate critical charge values for individual cells within a memory [Koh89][Mas92][Mas93]. In the most rigorous sense, one could perform the error-rate steps outlined above for each individual cell within the memory array, and then sum all of the results to arrive at an error rate for the chip. However, this can be accomplished mathematically by integrating across the distribution of individual cell Q_C 's.

In 1988, Binder laid the groundwork for such an integral error-rate technique [Bin88]. More recently, Petersen [Pet92] and Massengill [Mas92] have concurrently presented analytical methods for performing error-rate calculations in parts which exhibit a distribution of Q_C values. The method of this author is a convolution integral given by:

$$Rate = \int_0^{Q_{max}} R(\lambda) \frac{d\phi}{dQ}(\lambda) d\lambda \quad (24)$$

where $R(Q)$ is the upset rate at each critical charge value given by Eq. 24, $\phi(Q)$ is the number of cells among the total which possess a critical charge of Q_C (which includes the collection volume chord-

length integrations for the LET needed for Q_C), Q is collected charge (which can be related to LET), and Q_{\max} is the value of collected charge where one or both of the convolution terms approach zero.

6.6 Recap of Space Error-Rate Predictions

In this section we presented the basic concepts involved in predicting a chip's error rate in a space environment. The calculation is based on statistically-averaged models of the space environment and probability distributions of hits to a sensitive volume. The entire process revolves around a *known sensitive collection volume and a critical charge for a circuit* -- quantities that can be extracted from experimental results or determined by simulation.

It may not be very evident in the tutorial nature of this section, but error rate calculations involve extremely complex mathematics, many approximations on many levels, and environmental models which must average the unpredictable nature of solar weather. It is impossible for this course to adequately touch on even a few of these complications. However, the reader is referred to an excellent review of error-rate techniques by Petersen in [Pet92].

7. COMPARISONS WITH SATELLITE DATA

Until recently, actual single-event errors in space systems have been difficult to quantify. They have certainly been observable, since these errors have led to real problems in many space systems. Consider Don Vinson: involved in the ground control of the TDRS-I tracking satellite while it experienced frequent SEUs in the attitude control system, he once exclaimed "If this keeps up, TDRS will have to be equipped with a joy stick!" [Wil91]. But even though these errors have been observable, direct comparisons with ground-based testing and/or computer modeling has been hampered by the fact that the errors are usually deeply imbedded in the logic system. Tracing a manifestation of an error in a complex system back to its fundamental origin can be a formidable task in the laboratory, even more so when the system is operating in the field.

For this reason, several rad-effects experimental systems have been launched into earth orbit, or carried piggy-back on other systems, in the past five years. These experiments have been specifically designed to give clear, traceable information on circuit responses to the many radiation environments. One of the limitless uses of the data, which is becoming available from these missions, is the validation and improvement of modeling techniques for single-events.

A few of these systems are:

- the Engineering Test Satellite-V (**ETS-V**) in geosynchronous orbit launched by the National Space Development Agency of Japan,
- the Marine Observation Satellite-1 (**MOS-1**) in medium altitude orbit launched by Japan
- the **UoSat-3** satellite in sun-synchronous polar orbit launched by the Royal Aerospace Establishment of England
- the Combined Release and Radiation Effects Satellite (**CRRES**) in highly elliptical orbit dropping below the proton belts and extending above geosynchronous launched by NASA of the USA.

Data and analyses from these systems can be found in the *IEEE Trans. on Nuclear Science* in recent years, especially the "Space Environments and Effects" section of the 1991 and 1992 issues [CRR91].

8. FUTURE OF MODELING

If present trends continue, future advances in single-event modeling will be fueled by advances such as:

- rapidly improving computing technologies. Engineering workstations are reaching levels of computational power that just a few years ago were available only on the fastest mainframes. The opportunity exists for engineers and scientist to run complex numerical calculations, such as 2-d and 3-d simulations, on a desktop system. As this trend continues, our level of sophistication in SE modeling should also continue. The concept of having to buy 'Cray-years' of computing time to perform SE analysis on a part is no longer a threat in all but the most detailed computer modeling of single-events.
- advancing mathematical and numerical analysis techniques. These advances find their way into better, and more efficient, models for radiation effects.
- data from recent space radiation experiments onboard satellites such as those mentioned in Sec. 7 will improve environmental models. Continued refinement of the descriptions of the known space environment will produce more accurate error rate predictions.

The challenges and areas for continued research in SE modeling are many. A few of these are:

- the improvement of circuit models for scaled, submicron devices and technologies, including advances in charge collection models for these structures,
- the development of models for emerging devices and technologies as they appear,
- a reassessment of the basic assumptions involved in SE modeling when applied to the high speed and dense circuits continuing to emerge,
- the development of comprehensive analysis techniques for single-event in combinational logic,
- more work in the area of single-event modeling of analog subsystems,
- the advancement of system-level analysis techniques,
- the continued improvement in error-rate calculation methods, and
- true integration of SE modeling into the early design phases of microelectronic design (even commercial), as part of the integrated engineering CAD environment.

9. SUMMARY AND CONCLUSIONS

This course has attempted to introduce the reader to single-event effects, the motivation for modeling these effects, and the basic terminology needed to dig further into the literature on the issue of single-event analysis. A bottom-up hierarchical approach to SE modeling has been presented which begins with the physical device response to a single-event particle, then builds on these results to move to the electrical circuit response, then expands to the total microelectronic chip response in a space environment. Each level of the hierarchy uses the results from the previous level as input data, thus controlling the complexity of the problem by limiting the details needed from lower levels. Accuracy is maintained by experimental corroboration of the results at each level. The course has also attempted to outline the critical assumptions and limitations involved with the popular models seen in the literature, so that the user may gauge the applicability to his particular problem.

It is hoped that this course has given the reader a flavor for the many levels of complexity involved in a complete simulations of SE effects in microelectronics. Much enlightened scientific work has been performed and compiled over the past years by many groups and individuals. The reader is referred to the literature, especially the *IEEE Transactions on Nuclear Science*, the *IEEE Journal of Solid State Circuits*, and the *IEEE Transactions on Electron Devices*, for more information on this field of study.

10. ACKNOWLEDGMENTS

The author thanks the many people, without whom, this work would not have been possible. First, I thank my colleagues at Vanderbilt University in the Space Electronics Research Group, both faculty and students. I especially acknowledge the diligent work of Mr. V.J. Sananda, who helped with many of the time-consuming tasks of circuit/device simulation and figure preparation.

I also thank Dan Fleetwood of Sandia National Laboratories and Lew Cohn of the Defense Nuclear Agency for research contract support during the period this course was in the works. Without their support for my research program, the bills could not have been paid, and this work would not have been possible. Ed Petersen gave me helpful comments and suggestions for this course, and I thank him for these insights. I thank Allan Johnston and the 1993 IEEE NSREC committee for helpful suggestions and support.

My sincerest appreciation also goes to my editor and proofreader, CLM.

11. REFERENCES

- [Abd87] W.G. Abdel-Kader, P.J. McNulty, S. El-Teleaty, J.E. Lynch, and A.N. Khondker, "Estimating the Dimensions of the SEU-Sensitive Volume," *IEEE Trans. on Nuclear Science*, vol. NS-34, no. 6, pp. 1300-1304 (1987).
- [Ada82] J.H. Adams, Jr., "The Natural Radiation Environment Inside Spacecraft," *IEEE Trans. on Nuclear Science*, vol. NS-29, no. 6, pp. 2095-2100 (1982).
- [Ada83] J.H. Adams, Jr., "The Variability of Single Event Upset Rates in the Natural Environment," *IEEE Trans. on Nuclear Science*, vol. NS-30, no. 6, pp. 4475-4480 (1983).
- [Ada87] J.H. Adams, Jr., "Cosmic Ray Effects on Microelectronics, Part IV," *NRL Technical Memorandum Report 5901*, December (1987).
- [All90] M.L. Alles, K.L. Jones, J.E. Clark, J.C. Lee, W.F. Kraus, S.E. Kerns, and L.W. Massengill, "SOI/SRAM Rad-Hard Design Using a Predictive SEU Device Model," *GOMAC Conference Digest of Papers*, Las Vegas, November (1990).
- [Bei88] B.A. Beitzman, "N-Channel MOSFET Breakdown Characteristics and Modeling for p-well Technologies," *IEEE Trans. on Electron Devices*, vol. ED-35, no. 11, pp. 1935-1941 (1988).
- [Ben83] W.L. Bendel and E.L. Petersen, "Proton Upsets in Orbit," *IEEE Trans. on Nuclear Science*, vol. NS-30, no. 6, pp. 4481-4485 (1983).
- [Ben84] W.L. Bendel and E.L. Petersen, "Predicting Single Event Upsets in the Earth's Proton Belts," *IEEE Trans. on Nuclear Science*, vol. NS-31, no. 6, pp. 1201-1205 (1984).
- [Bin75] D. Binder, E.C. Smith, and A.B. Holman, "Satellite Anomalies from Galactic Cosmic Ions," *IEEE Trans. on Nuclear Science*, vol. NS-22, no. 6, pp. 2675-2680 (1975).
- [Bin88] D. Binder, "Analytic SEU Rate Calculation Compared to Space Data," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 1570-1572 (1988).
- [Bra79] J.N. Bradford, "A Distribution Function for Ion Track Lengths in Rectangular Volumes," *Journal of Applied Physics*, vol. 50, no. 6, pp. 3799-3801 (1979).
- [Car87] P.M. Carter and B.R. Wilkins, "Influences on Soft Error Rates in Static RAMs," *IEEE Journal of Solid-State Circuits*, vol SC-22, no. 3, pp. 430-436 (1987).
- [Che88] D.Y. Cheng, C.G. Hwang, and R.W. Dutton, "PISCES-MC: A Multiwindow, Multimethod 2-D Device Simulator," *IEEE Trans. on Computer Aided Design*, vol. 7, no. 9, pp. 1017-1026 (1988).
- [CRR91] Much data from the CRRES microelectronic test package has been published in the Space Environments and Effects section of the *IEEE Trans. on Nuclear Science*, vol. NS-38, no. 6, pp. 1642-1718 (1991) and the *IEEE Trans. on Nuclear Science*, vol. NS-39, no. 6, pp. 1804-1868 (1992).
- [Die82] S.E. Diehl, A. Ochoa, P.V. Dressendorfer, R. Koga, W.A. Kolasinski, "Error Analysis and Prevention of Cosmic Ion-Induced Soft Errors in Static CMOS SRAMs," *IEEE Trans. on Nuclear Science*, vol. NS-29, no. 6, pp. 2032-2039 (1982).

- [Die84] S.E. Diehl-Nagle, J.E. Vinson, and E.L. Petersen, "Single-Event Upset Rate Predictions for Complex Logic Systems," *IEEE Trans. on Nuclear Science*, vol. NS-31, no. 6, pp. 1132-1138 (1984).
- [Fis87] T.A. Fischer, "Heavy-Ion-Induced, Gate Rupture in Power MOSFETs," *IEEE Trans. on Nuclear Science*, vol. NS-34, no. 6, pp. 1786-1791 (1987).
- [Fu85] J.S. Fu, C.L. Axness, and H.T. Weaver, "Memory SEU Simulations Using 2-D Transport Calculations," *IEEE Electron Device Letters*, vol. EDL-6, no. 8, pp. 422-424 (1985).
- [Gov86] J.E. Gover and J.S. Browning, "Radiation-Hardening Systems Considerations : Electrical Systems and Radiation Environments," Sandia National Labs Technical Report No. SAND86-1737 (1986).
- [Gru84] H.L. Grubin, J.P. Kreskovsky, and B.C. Weinberg, "Numerical Studies of Charge Collection and Funneling in Silicon Device," *IEEE Trans. on Nuclear Science*, vol. NS-31, no. 6, pp. 1161-1166 (1984).
- [Hau85] J.R. Hauser, S.E. Diehl-Nagle, A.R. Knudson, A.B. Campbell, W.J. Stapor, and P. Shapiro, "Ion Track Shunt Effects in Multi-Junction Structures," *IEEE Trans. on Nuclear Science*, vol. NS-32, no. 6, pp. 4115-4121 (1985).
- [Hau88] J.R. Hauser, "SEU Effects in Bipolar RAMs," *Journal of Radiation Effects Research and Engineering*, vol. 6, no. 2, pp. 44-55 (1988).
- [Hoh87] J.H. Hohl and K.F. Galloway, "Analytical Model for Single Event Burnout of Power MOSFETs," *IEEE Trans. on Nuclear Science*, vol. NS-34, no. 6, pp. 1275-1280 (1987).
- [Hoh89] J.H. Hohl and G.H. Johnson, "Features of the Triggering Mechanism for Single Event Burnout of Power MOSFETs," *IEEE Trans. on Nuclear Science*, vol. NS-36, no. 6, pp. 2260-2266 (1989).
- [Hsi81] C.M. Hsieh, P.C. Murley, and R.R. O'Brien, "A Field-Funneling Effect on the Collection of Alpha-Particle-Generated Carriers in Silicon Devices," *IEEE Electron Device Letters*, vol. EDL-2, no. 4, pp. 103-105 (1981).
- [Hsi83] C.M. Hsieh, P.C. Murley, and R.R. O'Brien, "Collection of Charge From Alpha-Particle Tracks in Silicon Devices," *IEEE Trans. on Electron Devices*, vol. ED-30, no. 6, pp. 686-693 (1983).
- [Hu82] C. Hu, "Alpha-Particle-Induced Field and Enhanced Collection of Carriers," *IEEE Electron Device Letters*, vol. EDL-3, no. 2, pp. 31-34 (1982).
- [Iwa92] H. Iwata and T. Ohzone, "Numerical Analysis of Alpha-Particle-Induced Soft Errors in SOI MOS Devices," *IEEE Trans. on Electron Devices*, vol. 39, no. 5, pp. 1184-1190 (1992).
- [Joh90] A.H. Johnston and B.W. Hughlock, "Latchup in CMOS from Single Particles," *IEEE Trans. on Nuclear Science*, vol. NS-37, no. 6, pp. 1886-1893 (1990).
- [Kau91] N. Kaul, B.L. Bhuva, and S.E. Kerns, "Simulation of SEU Transients in CMOS ICs," *IEEE Trans. on Nuclear Science*, vol. NS-38, no. 6, pp. 1514-1520 (1991).
- [Kau92] N. Kaul, "Computer-Aided Estimation of Vulnerability of CMOS VLSI Circuits to Single-Event Upsets," PhD Dissertation, Dept. of Electrical Engineering, Vanderbilt University (1992).
- [Ker89] S.E. Kerns, L.W. Massengill, D.V. Kerns, M.L. Alles, T.W. Houston, H. Lu, and L.R. Hite, "Model for CMOS/SOI Single-Event Vulnerability," *IEEE Trans. on Nuclear Science*, vol. NS-36, no. 6, pp. 2305-2310 (1989).

- [Kin74] J.H. King, "Solar Proton Fluences for 1977-1983 Space Missions," *J. Spacecraft and Rockets*, vol. 11, pp. 401-408 (1974).
- [Kir79] S. Kirkpatrick, "Modeling Diffusion and Collection of Charge from Ionizing Radiation in Silicon Devices," *IEEE Trans. on Electron Devices*, vol. ED-26, no. 11, pp. 1742-1753 (1979).
- [Knu84] A.R. Knudson, A.B. Campbell, P. Shapiro, W.J. Stapor, E.A. Wolicki, E.L. Petersen, S.E. Diehl-Nagle, J. Hauser, and P.V. Dressendorfer, "Charge Collection in Multilayer Structures," *IEEE Trans. on Nuclear Science*, vol. NS-31, no. 6, pp. 1149-1154 (1984).
- [Knu86] A.R. Knudson, A.B. Campbell, J.R. Hauser, M. Jessee, W.J. Stapor, and P. Shapiro, "Charge Transport by Ion Shunt Effect," *IEEE Trans. on Nuclear Science*, vol. NS-33, no. 6, pp. 1560-1564 (1986).
- [Knu91] A.R. Knudson and A.B. Campbell, "Comparison of Experimental Charge Collection Waveforms with PISCES Calculations," *IEEE Trans. on Nuclear Science*, vol. NS-38, no. 6, pp. 1540-1545 (1991).
- [Kog88] R. Koga, W.A. Kolasinski, J.V. Osborn, J.H. Elder and R. Chitty, "SEU Test Techniques for 256k Static RAMs and Comparisons of Upsets Induced by Heavy Ions and Protons," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 1638-1643 (1988).
- [Koh89] R.A. Kohler and R. Koga, "SEU Characterization of Hardened CMOS SRAMs Using Statistical Analysis of Feedback Delay in Memory Cells," *IEEE Trans. on Nuclear Science*, vol. NS-36, no. 6, pp. 2318-2323 (1989).
- [Kre85] J.P. Kreskovsky and H.L. Grubin, "Simulation of Charge Collection in a Multilayer Device," *IEEE Trans. on Nuclear Science*, vol. NS-32, no. 6, pp. 4140-4144 (1985).
- [Kre86] J.P. Kreskovsky and H.L. Grubin, "Numerical Simulation of Charge Collection in Two- and Three-Dimensional Silicon Diodes - A Comparison," *Solid State Electronics*, vol. 29, no. 5, pp. 505-518 (1986).
- [Lea91] J.F. Leavy, L.F. Hoffmann, R.W. Shovan, and M.T. Johnson, "Upset Due to a Single Particle Caused Propagated Transient in a Bulk CMOS Microprocessor," *IEEE Trans. on Nuclear Science*, vol. NS-38, no. 6, pp. 1493-1499 (1991).
- [Li84] K.W. Li, J.R. Armstrong, and J.G. Tront, "An HDL Simulation of the Effects of Single Event Upsets on Microprocessor Program Flow," *IEEE Trans. on Nuclear Science*, vol. NS-31, no. 6, pp. 1139-1144 (1984).
- [Lom90] T.S. Lomheim, R.M. Shima, J.R. Angione, W.F. Woodward, D.J. Asman, R.A. Keller, and L.W. Schumann, "Imaging Charge-Coupled Device (CCD) Transient Response to 17 and 50 MeV Protons and Heavy-Ion Irradiation," *IEEE Trans. on Nuclear Science*, vol. NS-37, no. 6, pp. 1876-1885 (1990).
- [Mar87] R.C. Martin, N.M. Ghoniem, Y. Song, and J.S. Cable, "The Size Effect of Ion Charge Tracks on Single Event Multiple-Bit Upset," *IEEE Trans. on Nuclear Science*, vol. NS-34, no. 6, pp. 1305-1309 (1987).
- [Mas90] L.W. Massengill, D.V. Kems, S.E. Kems, and M.L. Alles, "Single-Event Charge Enhancement in SOI Devices," *IEEE Electron Device Letters*, vol. EDL-11, no. 2, pp. 98-99 (1990).
- [Mas91] L.W. Massengill, "SEU-Hardened Resistive-Load Static RAMs," *IEEE Trans. on Nuclear Science*, vol. NS-38, no. 6, pp. 1478-1485 (1991).

- [Mas92] L.W. Massengill, "SEU Characterization of Modern, Hardened Technologies," presented at the *1992 SEU Symposium*, Los Angeles (1992).
- [Mas93] L.W. Massengill, M.L. Alles, S.E. Kerns, and K.L. Jones, "Analytical Correlation Between Observed Non-Ideal SEU Cross-Section Data and Process Parameter Distributions," presented at the *1993 Nuclear and Space Radiation Effects Conference*, Snowbird, Utah, July (1993).
- [May79] T.C. May and M.H. Woods, "Alpha-Particle-Induced Soft Errors in Dynamic Memories," *IEEE Trans. on Electron Devices*, vol ED-26, no. 1, pp. 2-9 (1979).
- [McD89] P.T. McDonald, W.J. Stapor, A.B. Campbell, and L.W. Massengill, "Nonrandom Single-Event Upset Trends," *IEEE Trans. on Nuclear Science*, vol. NS-36, no. 6, pp. 2324-2329 (1989).
- [McL82] F.B. McLean and T.R. Oldham, "Charge Funneling in n- and p-type Si Substrates," *IEEE Trans. on Nuclear Science*, vol. NS-29, no. 6, pp. 2018-2023 (1982).
- [McN85] P.J. McNulty, W.G. Abdel-Kader, and J.M. Bisgrove, "Methods of Calculating SEU Rates for Bipolar and NMOS Circuits," *IEEE Trans. on Nuclear Science*, vol. NS-32, no. 6, pp. 4180-4184 (1985).
- [McP81] R.C. McPartland, "Circuit Simulations of Alpha-Particle-Induced Soft Errors in MOS Dynamic RAMs," *IEEE Journal of Solid State Circuits*, vol. SC-15, no. 1, pp. 31-34 (1981).
- [Mes82] G.C. Messenger, "Collection of Charge on Junction Nodes from Ion Tracks," *IEEE Trans. on Nuclear Science*, vol. NS-29, no. 6, pp. 2024-2031 (1982).
- [Mes86] G.C. Messenger and M.S. Ash, *The Effects of Radiation on Electronic Systems*, van Nostrand Reinhold Co., New York (1986).
- [Mni83] T.M. Mnich, S.E. Diehl, B.D. Shafer, R. Koga, W.A. Kolasinski, and A. Ochoa, "Comparison of Analytical Models and Experimental Results for Single Event Upset in CMOS SRAMs," *IEEE Trans. on Nuclear Science*, vol. NS-30, no. 6, pp. 4620-4623 (1983).
- [Nym91] R.A. Nymmik, M.I. Panasyuk, T.I. Pervaja, and A.A. Suslov, "Model of Galactic Cosmic Ray Fluxes," in *NRL Technical Publication 209-4154* (1991).
- [Och81] A. Ochoa and P.V. Dressendorfer, "A Discussion of the Role of Distributed Effects on Latch-Up," *IEEE Trans. on Nuclear Science*, vol. NS-28, no. 6, pp. 4292-4297 (1981).
- [Och83] A. Ochoa, F.W. Sexton, T.F. Wrobel, G.L. Hash, and R.J. Sokel, "Snapback : A Stable Regenerative Breakdown Mode of MOS Devices," *IEEE Trans. on Nuclear Science*, vol. NS-30, no. 6, pp. 4127-4130 (1983).
- [Old83] T.R. Oldham and F.B. McLean, "Charge Collection Measurements for Heavy Ions Incident on n- and p-type Silicon," *IEEE Trans. on Nuclear Science*, vol. NS-30, no. 6, pp. 4493-4500 (1983).
- [Old86] T.R. Oldham, F.B. McLean, and J.M. Hartman, "Revised Funnel Calculations for Heavy Particles with High dE/dx," *IEEE Trans. on Nuclear Science*, vol. NS-33, no. 6, pp. 1646-1650 (1986).
- [Pet82] E.L. Petersen, P. Shapiro, J.H. Adams, and E.A. Burke, "Calculation of Cosmic Ray Induced Soft Upsets and Scaling in VLSI Devices," *IEEE Trans. on Nuclear Science*, vol. NS-29, no. 6, pp. 2055-2063 (1982).

- [Pet92] E.L. Petersen, J.C. Pickel, J.H. Adams, and E.C. Smith, "Rate Prediction for Single Event Effects -- A Critique," *IEEE Trans. on Nuclear Science*, vol. NS-39, no. 6, pp. 1577-1599 (1992).
- [Pic78] J.C. Pickel and J.T. Blandford, "Cosmic-Ray-Induced Errors in MOS Memory Cells," *IEEE Trans. on Nuclear Science*, vol. NS-25, no. 6, pp. 1166-1171 (1978).
- [Pic80] J.C. Pickel and J.T. Blandford, "Cosmic-Ray-Induced Errors in MOS Devices," *IEEE Trans. on Nuclear Science*, vol. NS-27, no. 6, pp. 1006-1015 (1980).
- [Pic81] J.C. Pickel and J.T. Blandford, "CMOS RAM Cosmic-Ray-Induced Error-Rate Analysis," *IEEE Trans. on Nuclear Science*, vol. NS-28, no. 6, pp. 3962-3967 (1981).
- [Rol88a] J.G. Rollins and J.C. Choma, "Mixed-Mode PISCES-SPICE Coupled Circuit and Device Solver," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 862-867 (1988).
- [Rol88b] J.G. Rollins, T.K. Tsubota, W.A. Kolasinski, N.F. Haddad, L. Rockett, M. Cerrila, and W.B. Hennley, "Cost-Effective Numerical Simulation of SEU," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 1608-1612 (1988).
- [Sai80] G.A. Sai-Halasz and M.R. Wordeman, "Monte Carlo Modeling of the Transport of Ionizing Radiation Created Carriers in Integrated Circuits," *IEEE Electron Device Letters*, vol. EDL-1, no. 10, pp. 211-213 (1980).
- [Sai82] G.A. Sai-Halasz, M.R. Wordeman, and R.H. Dennard, "Alpha-Particle-Induced Soft Error Rate in VLSI Circuits," *IEEE Trans. on Electron Devices*, vol ED-29, no. 4, pp. 725-731 (1982).
- [Sel80] S. Selberherr, A. Schutz, and H.W. Potzl, "MINIMOS - A Two-Dimensional MOS Transistor Analyzer," *IEEE Trans. on Electron Devices*, vol. ED-27, no. 8, pp. 1540-1549 (1980).
- [Sex92] F.W. Sexton, "Measurement of Single-Event Phenomena in Devices and ICs," 1992 Nuclear and Space Radiation Effects Short Course, Chapter 3, New Orleans, July (1992).
- [Sil87] R. Silberberg, C.H. Tsao, J.H. Adams, and J.R. Letaw, "High-Energy Radiation Environment During Manned Space Flights," *AIP Conference Proceedings 186*, Sanibel Island, FL pp. 146-158 (1987).
- [Son88] Y. Song, K.N. Vu, J.S. Cable, A.A. Witteles, W.A. Kolasinski, R. Koga, J.H. Elder, J.V. Osborn, R.C. Martin, and N.M. Ghoniem, "Experimental and Analytical Investigation of Single Event, Multiple Bit Upsets in Polysilicon Load 64k x 1 NMOS SRAMs," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 1673-1677 (1988).
- [Sta88] W.J. Stapor, P.T. McDonald, A.R. Knudson, A.B. Campbell, and B.G. Glagola, "Charge Collection in Silicon for Ions of Different Energy but Same Linear Energy Transfer (LET)," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 1585-1590 (1988).
- [Sta90] W.J. Stapor, J.P. Meyers, J.B. Langworthy, and E.L. Petersen, "Two Parameter Bendel Model Calculations for Predicting Proton Induced Upset," *IEEE Trans. on Nuclear Science*, vol. NS-37, no. 6, pp. 1966-1973 (1990).
- [Sts74] E.G. Stassinopoulos and J.H. King, "Empirical Solar Proton Model for Orbiting Spacecraft Applications," *IEEE Trans. on AES*, vol AES-10, no. 4 (1974).
- [Sts87] E.G. Stassinopoulos, "Charged Particle Radiation Exposure of Geocentric Satellites," *AIP Conference Proceedings 186*, Sanibel Island, FL pp. 3-63 (1987).

- [Sun78] E. Sun, J. Moll, J. Berger, and B. Adlers, "Breakdown Mechanism in Short Channel MOS Transistors," *IEDM Technical Digest*, pp. 478-482 (1978).
- [Tit91] J.L. Titus, G.H. Johnson, R.D. Schrimpf, and K.F. Galloway, "Single-Event Burnout of Power Bipolar Junction Transistors," *IEEE Trans. on Nuclear Science*, vol. NS-38, no. 6, pp. 1315-1322 (1991).
- [Tur90] T.L. Turflinger, "Understanding Single Event Phenomena in Complex Analog and Digital Integrated Circuits," *IEEE Trans. on Nuclear Science*, vol. NS-37, no. 6, pp. 1832-1838 (1990).
- [Wag88] R.S. Wagner, N. Bordes, J.M. Bradley, C.J. Maggiore, A.R. Knudson, and A.B. Campbell, "Alpha, Boron, Silicon, and Iron Ion-Induced Current Transients in Low-Capacitance Silicon and GaAs Diodes," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 1578-1584 (1988).
- [Wal62] J.T. Wallmark and S.M. Marcus, "Minimum Size and Maximum Packing Density of Nonredundant Semiconductor Devices," *Proc. of the IRE*, vol 50, pp. 286-298 (1962).
- [Wea85] T.R. Weatherford, J.R. Hauser, and S.E. Diehl, "A Study of Single Events in GaAs SRAMs," *IEEE Trans. on Nuclear Science*, vol. NS-32, no. 6, pp. 4170-4175 (1985).
- [Wea86] T.R. Weatherford, J.R. Hauser, and S.E. Diehl, "Comparisons of Single-Event Vulnerability of GaAs SRAMs," *IEEE Trans. on Nuclear Science*, vol. NS-33, no. 6, pp. 1590-1596 (1986).
- [Wil91] D.C. Wilkinson, S.C. Daughridge, J.L. Stone, H.H. Sauer, and P. Darling, "TDRS-1 Single Event Upsets and the Effect of the Space Environment," *IEEE Trans. on Nuclear Science*, vol. NS-38, no. 6, pp. 1708-1712 (1991).
- [Wro87] T.F. Wrobel, "On Heavy Ion Induced Hard-Errors in Dielectric Structures," *IEEE Trans. on Nuclear Science*, vol. NS-34, no. 6, pp. 1262-1268 (1987).
- [Yan79] D.S. Yaney, J.T. Nelson, and L.L. Vanskike, "Alpha-Particle Tracks in Silicon and their Effect on Dynamic MOS RAM Reliability," *IEEE Trans. on Electron Devices*, vol. ED-26, no. 1, pp. 10-16 (1979).
- [Zha89] X. Zhang, "Calculation of Critical Charge of Bipolar Memory Circuits," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 1, pp. 187-189 (1989).
- [Zie79] J.F. Ziegler and W.A. Lanford, "Effect of Cosmic Rays on Computer Memories," *Science*, vol. 206, pp. 776-788 (1979).
- [Zie85] J.F. Ziegler, J.P. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solids*, Pergamon Press, New York (1985).
- [Zou83] J. A. Zoutendyk, "Modeling of Single-Event Upset in Bipolar Integrated Circuits," *IEEE Trans. on Nuclear Science*, vol. NS-30, no. 6, pp. 4540-4545 (1983).
- [Zou88a] J.A. Zoutendyk, H.R. Schwartz, and L.R. Nevill, "Lateral Charge Transport from Heavy-Ion Tracks in Integrated Circuit Chips," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 1644-1647 (1988).
- [Zou88b] J.A. Zoutendyk, E.C. Secrest, and D.F. Berndt, "Investigation of Single-Event Upset (SEU) in an Advanced Bipolar Process," *IEEE Trans. on Nuclear Science*, vol. NS-35, no. 6, pp. 1573-1577 (1988).

- [Zou90] J.A. Zoutendyk, L.S. Smith, and L.D. Edmonds, "Response of a DRAM to Single-Ion Tracks of Different Heavy-Ion Species and Stopping Powers," *IEEE Trans. on Nuclear Science*, vol. NS-37, no. 6, pp. 1844-1848 (1990).
- [Zul84] R. Zuleeg, J.K. Notthoff, and D.K. Nichols, "SEU of Complementary GaAs Static RAMs due to Heavy Ions," *IEEE Trans. on Nuclear Science*, vol. NS-31, no. 6, pp. 1121-1123 (1984).
- [Zul88] R. Zuleeg and J.K. Notthoff, "SEU Hardening of GaAs E-JFET SRAMs," *Journal of Radiation Effects Research and Engineering*, vol. 6, no. 2, pp. 136-145 (1988).

Chapter 4

Novel Devices and Sensors

James Pickel
S-Cubed Division of Maxwell Laboratories
Mission Viejo, CA 92691

NOVEL DEVICES AND SENSORS

James C. Pickel

S-Cubed Division of Maxwell Laboratories
San Diego, California

- 1. INTRODUCTION**
- 2. OVERVIEW OF OPTICAL SENSORS**
- 3. SYNOPSIS OF DETECTOR AND READOUT ELECTRONICS TECHNOLOGIES**
 - 3.1 PHOTOCONDUCTIVE DETECTORS**
 - 3.2 PHOTOVOLTAIC DETECTORS**
 - 3.3 METAL-INSULATOR-SEMICONDUCTOR DETECTORS**
 - 3.4 CRYOGENIC CMOS**
 - 3.4 SUPERCONDUCTING ELECTRONICS AND DETECTORS**
- 4. STATUS OF COMPONENT TECHNOLOGIES**
 - 4.1 DETECTOR TECHNOLOGY**
 - 4.2 ELECTRONICS FOR SENSOR APPLICATIONS**
 - 4.2.1 CRYOGENIC IC DEVICE TECHNOLOGY**
 - 4.2.2 SIGNAL PROCESSING TECHNOLOGY**
 - 4.3 SUBSYSTEM INTERCONNECTS AND DATA BUS**
 - 4.4 CRYO-COOLING TECHNOLOGY**
- 5. UNIQUE RADIATION EFFECTS**
 - 5.1 TRANSIENT EFFECTS IN DETECTORS**
 - 5.2 TOTAL DOSE EFFECTS IN CRYOGENIC MOS ELECTRONICS**
 - 5.3 TOTAL DOSE EFFECTS IN DETECTORS**
 - 5.4 DISPLACEMENT DAMAGE EFFECTS IN DETECTORS**
- 6. ADDITIONAL CONSIDERATIONS**
 - 6.1 NOISE**
 - 6.2 READOUT OF LARGE ARRAYS**
 - 6.3 RELIABILITY OF HYBRIDS**
 - 6.4 CRYO-COOLING**
- 7. CONCLUSIONS**

IV-2

1. INTRODUCTION

Optical sensors are an emerging application of electronics with potential for wide ranging use in space systems. Similar to requirements discussed in the previous papers, considerations for the radiation environment in space and the system requirements must be made for successful application of this novel class of electronics.

This paper is a tutorial providing a broad overview of passive optical sensors and considerations for their use in space. The paper is written from the perspective of a device physicist rather than a systems engineer. The information presented is from the open literature and from private communication with technologists in the rapidly emerging field. Much of the detailed information on IR detector technology is documented in the Proceedings of the annual IRIS Detector Specialty Conference [1].

Because the optical wavelength band spans a broad range, approximately from 0.1 μm to 100 μm , various detector technologies are utilized. The discussions center on modern device technologies for detectors and microelectronic components, as well as some of their support systems such as cryo-coolers; detailed discussion of optics, optical sources and optical propagation is not included. Also, the paper concentrates on natural environment radiation effects rather than nuclear weapon effects.

The discussion begins with an overview of optical sensors, followed by a brief synopsis of detector and readout electronics technologies. Areas where there are unique device physics considerations are pointed out. Summaries on the status of component technologies, including infrared detectors, visible and ultraviolet detectors, and readout and processing electronics are then given. Radiation effects that are unique to sensors are discussed and compared to well known radiation effects in conventional electronics. At appropriate locations throughout the paper, the practical considerations for use of these devices in space are discussed. The paper concludes with discussion of additional considerations and trade-offs facing the sensor designer.

2. OVERVIEW OF OPTICAL SENSORS

What can we do with optical sensors in space systems? Figure 1 shows that there is a wide range of photon wavelengths of interest for optical sensors to detect, spanning from the ultraviolet (UV) to the visible (VIS) to the short wavelength (SWIR), medium wavelength (MWIR), long wavelength (LWIR), very long wavelength (VLWIR) and far infrared (FIR). There is a wealth of information that can be obtained from the distribution of these photons provided they are appropriately detected and the electronic signals appropriately processed.

If we have a sensor in orbit, we can either view objects against the earth background or against the background of deep-space. For earth viewing, a nice job for a spaceborne sensor is to monitor natural resources and environmental changes. Also the military has a keen interest in observing the enemy's position on earth and particularly in having an early warning that they are beginning to shoot missiles. There are many applications for scientific instruments to measure the photonic emissions from planet Earth.

RT-36074-M

WAVELENGTHS OF INTEREST (μm)								
FIR	VLWIR	LWIR	MWIR	SWIR	NIR	VIS	UV	
26-100	14-26	8-14	3-8	1-3	0.8-1	0.4-0.8	0.1-0.4	
EARTH MONITORING								
• Resource/environment surveillance		X	X	X	X	X	X	X
• Military surveillance		X	X	X	X	X	X	
• Scientific instruments		X	X	X	X	X	X	
SPACE MONITORING								
• Strategic defense								
- Early warning				X	X			
- Midcourse surveillance		X	X	X	X			
- Boost to midcourse intercept		X	X	X	X			
- Midcourse intercept		X	X	X	X			
- Midcourse to reentry intercept		X	X	X	X			
• Scientific								
- Astronomy, spectroscopy	X	X	X	X	X	X	X	X

- Imaging and surveillance and astronomy
- Detection, tracking and discrimination
- Spectroscopy

Figure 1. Applications of optical sensors in space systems.

If we consider viewing into space, an application that has received a lot of attention over the past decade is strategic defense against intercontinental ballistic missiles. Also astronomers love to place their telescopes and instruments above the atmosphere and look out at space. There has been revolutionary growth in the astronomy data base because of improved sensors.

So we see that the applications can be grouped into 1) imaging, surveillance, and astronomy 2) detection, tracking and discrimination, and 3) scientific measurements such as spectroscopy and interferometry. The detailed considerations for designing a sensor for each of these various applications are different and mission oriented; however there are many common considerations which we will address in this paper.

IR sensor systems can be grouped into active and passive systems. Active systems operate by illuminating a target with an IR source and detecting the reflected radiation. Passive systems rely on detecting the photons that are emitted from the target by nature of the temperature of the target, or from reflected sunlight in the case of visible detectors. We will only address passive sensors.

Photon Sources

All objects at a temperature above absolute zero emit photons by radiation. The radiant emittance is proportional to temperature to the fourth power and is described by the Stefan Boltzman law,

$$R = e\sigma T^4 \quad (1)$$

where

- e = emissivity
- σ = Stefan-Boltzmann constant
- T = absolute temperature.

The photons are emitted with a distribution in wavelength that is described by the Plank formula,

$$Nd\lambda=2\pi c d\lambda/\{\lambda^4[\exp(hc/\lambda kT)-1]\} \quad (2)$$

where

- c = speed of light
- λ = wavelength
- h = Planck's constant
- k = Boltzmann's constant.

Figure 2 shows the spectral radiant emittance from a blackbody as a function of temperature. The notable effects are that the total radiant emittance increases strongly with increasing temperature and that the peak wavelength of the spectrum decreases with increasing temperature. An object at any temperature above absolute zero can be detected by its passive radiant emittance against a background that is comparable or colder, provided a detector with sufficient sensitivity is used. With the average temperature of the earth's upper atmosphere near 250 K, hot objects such as rocket exhaust can be readily distinguished from the terrestrial background and also against the cold background of space. For colder objects, a more sensitive detector is required since the total radiant energy is less and the detector must be able to detect longer wavelengths. The application for cold body detection is against the low background of space, which has an effective temperature near 3 K.

Photon Detectors

The central element in any infrared detection system is the detector, the device which transforms the energy of the electromagnetic energy falling upon it into some other form, usually electrical. Most detectors are based upon photoeffects in semiconductors.

Semiconductor photon detectors can be grouped into thermal detectors and photodetectors. Thermal detectors measure the rate at which energy is absorbed. They respond only to the intensity of the absorbed radiant power, disregarding the spectral content of it. Types of thermal detectors

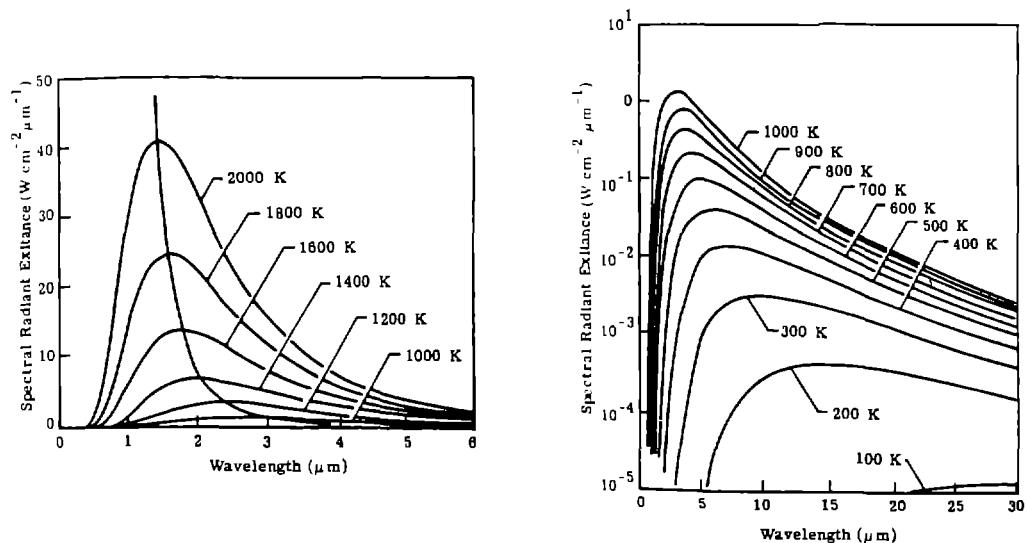


Figure 2. Blackbody radiation curves.

include bolometers, whose operation depend on heating produced by the radiation, and thermoelectric detectors, such as a thermocouple formed by joining two metals or semiconductors having different thermoelectric powers and producing a voltage in response to temperature change.

In contrast, photon detectors measure the rate at which quanta are absorbed rather than the absorbed energy. The energy of the photons is related to the wavelength by the basic quantum equation,

$$E = hc/\lambda \quad (3)$$

where

- h = Planck's constant
- c = speed of light
- λ = wavelength.

A convenient relation is

$$E = 1.24/\lambda \quad (4)$$

where

- E = photon energy in eV
- λ = photon wavelength in μm .

Photon detectors are selective detectors responsive only to those photons of sufficiently short wavelength (high enough energy). Their response at any wavelength is proportional to the rate at which photons of that wavelength are absorbed. Figure 3 shows the photon energy as a function of wavelength. It can be seen that in the wavelengths from 0.1 to 100 μm , the energy of the photons is in the range of semiconductor bandgaps and energy levels, in the range from 0.01 to 10 eV.

For space applications semiconductor photodetectors are used almost entirely since thermal detectors are much less sensitive and are not selective regarding wavelength. The photo-generated carriers can be detected by semiconductor devices either by photoconductive or by photovoltaic effects. Photoconductivity relies on a change of conductivity induced by the radiation. Photovoltaic effects rely on change in a voltage at a potential barrier; the potential barrier can be produced either by a p-n junction or induced by electric fields with a metal-insulator-silicon structure.

Figure 4 summarizes the principal detector technologies applicable to space systems and their range of operation. The top section divides the wavelengths applicable to optical sensors into the conventional designations for the wavelength bands, ranging from ultraviolet to visible to various

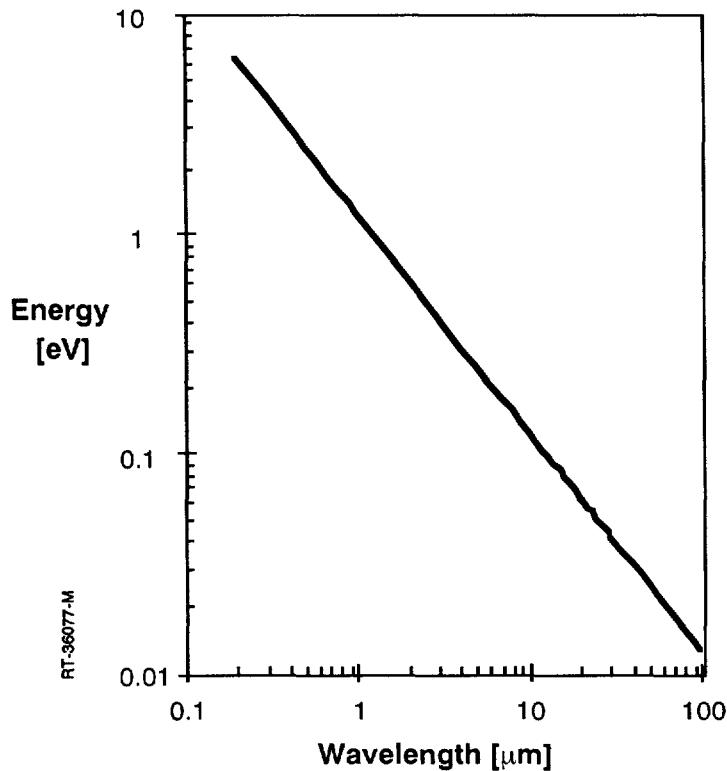


Figure 3. Photon energy versus wavelength.

bands in the infrared. The divisions are based somewhat on requirements for different detector technologies for the different wavelengths, as will be discussed later. The next section on the figure shows the atmospheric transmission a function of wavelength. There are preferred bands for maximum transmission depending on the wavelength. The VLWIR band is not suitable for detection within the atmosphere. The next section of the figure shows the wavelength bands for emission from various objects. All objects reflect sunlight and offer a target in the visible and near infrared. Hot objects present an optimum target in the SWIR band with gradual transition to MWIR, LWIR and VLWIR as we go from warm to cold bodies. The lower section of the figure shows the principal detector technologies and their wavelength ranges of operation. Thermal detectors, including thermocouples, thermistors and pyroelectric devices, are based on heating effects and are generally not sensitive enough for space applications. The photodetectors include intrinsic and extrinsic detectors. These detectors are based on excitation of carriers in the semiconductors by interaction of optical photons. They have the highest sensitivity, and various device technologies are available for detection that span the wavelength bands of interest. In the following discussions, we will concentrate on the semiconductor photodetectors listed in Figure 4.

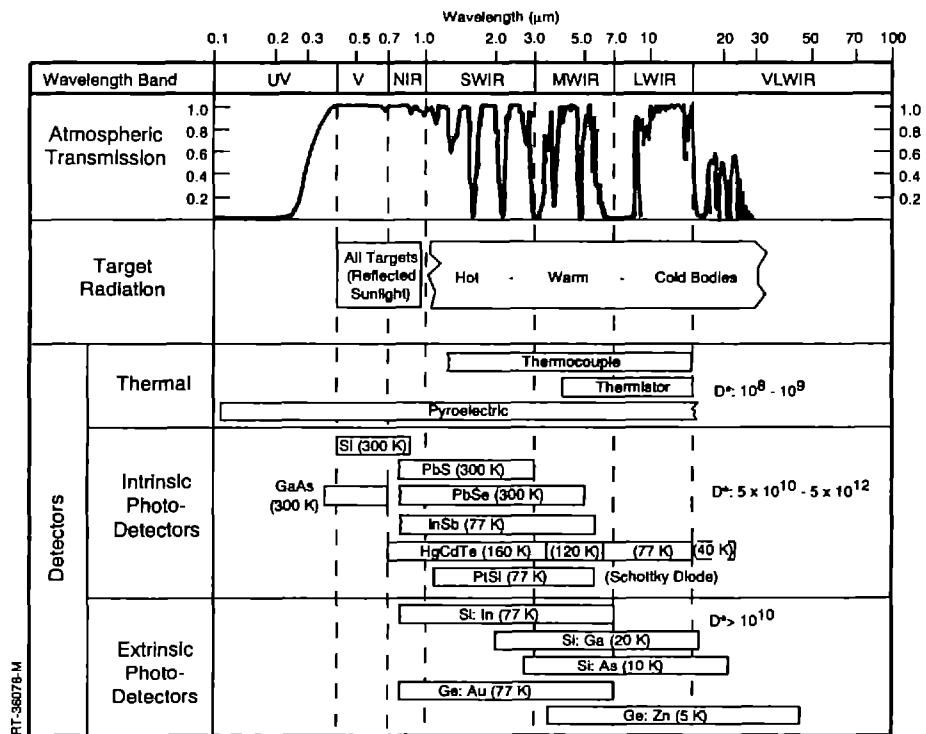


Figure 4. Principal detector technologies and ranges of operation.

Description of Generic Sensor

Before we delve into the details of the various detector technologies, it will be useful to consider some of the aspects of an optical sensor system. The detector is the “heart” of the sensor and has the primary job — to convert photons to electrons. It needs a lot of support in doing this job. Figure 5 illustrates all the surrounding support elements and subsystems that provide the necessary environment and support to a hybrid infrared focal plane array.

The task of the sensor is to collect photons from a target or region of interest, analyze their intensity, wavelength and spatial distribution on the focal plane and produce useful information about the scene in the form of digital bits that can be communicated to the user. The photons are collected by a telescope and routed to the detector by a system of optics consisting of mirrors and lenses. For IR system optics, mirrors are used more than lenses. Control of stray light is performed by baffle designs that minimize unwanted reflections. The detector array is enclosed in a vacuum chamber. The vacuum chamber is not needed in space, of course, but is necessary for ground test. Within the vacuum chamber is a cold shield which surrounds the detector. The cold shield is necessary because the detector responds to photons that are emitted by any surface within its field of view. Cold shields are typically at temperatures sufficiently low that the black body emission is negligibly low. Another consideration for controlling the background radiation is the aperture

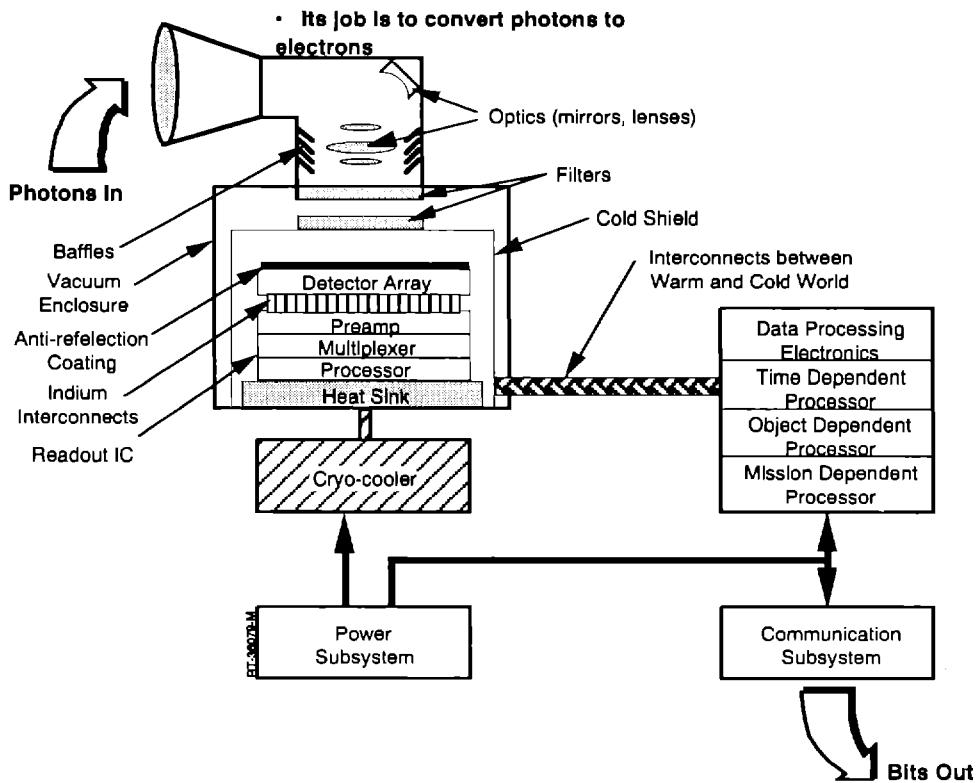


Figure 5. Generic sensor components for an infrared hybrid focal plane array.

through the cold shield which allows the detector to view the target scene. For many designs, a cold band-pass filter is placed over the aperture which allows only radiation within bands that are tuned to the target. The filter itself must be cold to limit self emission.

The detector array requires a readout integrated circuit (ROIC) to amplify the small signal current from the detector, multiplex the signals from the various detectors on the array and perform some signal processing. These tasks are accomplished by a ROIC which, for the case of hybrid focal plane arrays, is electrically and physically mated to the detector array by means of indium column interconnects between each detector site on the detector array and each unit cell on the readout array. Most IR detectors must be operated at cryogenic temperatures as indicated in Figure 4. The hybrid detector / readout assembly is thermally bonded to a heat sink which is maintained at the proper cryogenic temperature. The heat sink requires some form of cooling. Several approaches to accomplishing the cryogenic cooling are discussed in a following section, but the practical approaches require a powered mechanical refrigeration cycle.

The detector and readout must be electrically connected to more powerful data processing electronics in order to make use of the data from the focal plane. The interconnects between the cold world of the hybrid and the warm world of the data processing electronics are no trivial matter for the sensor designer. Because detector arrays can produce a very large data rate, the interconnect to

the data processor can be a bottleneck. There is a limit to the number of physical wires that can be interconnected between the cold and warm areas and still maintain low temperatures, because of heat transfer. Low thermal conductivity wires are required; these have higher electrical resistance than the usual copper wire and this causes increased noise sensitivity to electromagnetic coupling. In addition, use of more wires results in increased complexity with an impact on reliability and cost.

The data processing functions can be divided into three general functions: time dependent processor, object dependent processor and mission dependent processor. The time dependent processor involves processing functions related to temporal aspects of the signal. The object dependent processor is related to spatial aspects of the signal. The mission dependent processor is related to specific functions of the sensor mission.

Additional support for the sensor is required from a power subsystem and a communication subsystem. The power subsystem supports the cryo-cooler as well as the data processing electronics and the communication subsystem. The communication subsystem accepts the digital bits representing the signal information from the data processing electronics and communicates the data to the user.

Example of a Real Sensor Design

It is useful at this point to examine a real sensor design. An example of a scientific application of an IR sensor is the Atmospheric Infrared Sounder (AIRS) instrument, being designed by Loral [2]. AIRS is a key component of NASA's Earth Orbiting System (EOS) program, a spaceborne global observation system designed to obtain comprehensive long-term measurements of Earth processes affecting global change. The role of AIRS will be to provide both new and more accurate data about the atmosphere, land and oceans for application to climate studies and weather prediction. Among the important parameters to be derived from AIRS observations are atmospheric temperature profiles with an accuracy of 1 K in 1 km layer mean averages, and surface temperature with an accuracy of 0.5 K.

The AIRS measurement technique is based on passive IR remote sensing from a 705 km polar orbit using a nadir viewing high spectral resolution grating spectrometer operating in the 3.4-15.4 μm region. It will measure temperature by observing simultaneously as many as 3600 colors in the infrared via spectral dispersion across high-density linear HgCdTe focal plane arrays. NASA algorithms will convert this data to global profiles of air and surface temperature accurate to a fraction of a degree. Such data will be used to investigate global warming phenomena. Figure 6 gives an overview of the AIRS mission.

A more detailed physical view of the AIRS instrument is given in Figure 7. This figure illustrates some of the complexity in a sensor design. Figure 8 shows the complex optics involved in a sensor design. Figure 9 shows the cryocooling concept for AIRS that is based on a pair of Stirling cycle refrigerators.



Mission: Earth Observing System (EOS)
NASA Polar Platform 2 (EOS-PM1)

Science Objective: Determine Global Atmospheric Temperature and Humidity Profiles

Technique: Multispectral IR Remote Sensing

- Vertical Coverage: 0-25 km
- Vertical Resolution: 1 km
- Temperature Accuracy: 1 K
- Ground Resolution: 13 km @ Nadir (1.1°)

Instrument: Array Grating Spectrometer

- IR Spectral Coverage: 3.74-4.61 μm
6.20-8.22 μm
8.80-15.4 μm
- IR Spectral Resolution, $\lambda/\Delta\lambda$: 1200
- IR Spectral Samples: 2378
- Sensitivity (NEΔT): 0.2 K
- 4 Visible Channels: 0.4-1.0 μm
- Operating Lifetime: 5 years

Figure 6. Atmospheric Infrared Sounder (AIRS).

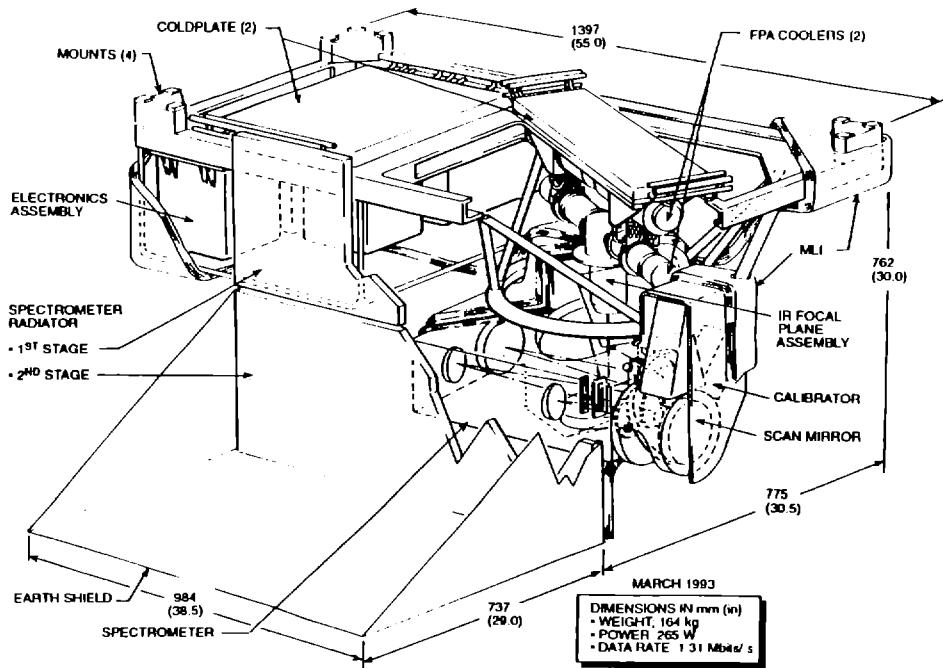


Figure 7. AIRS instrument schematic.

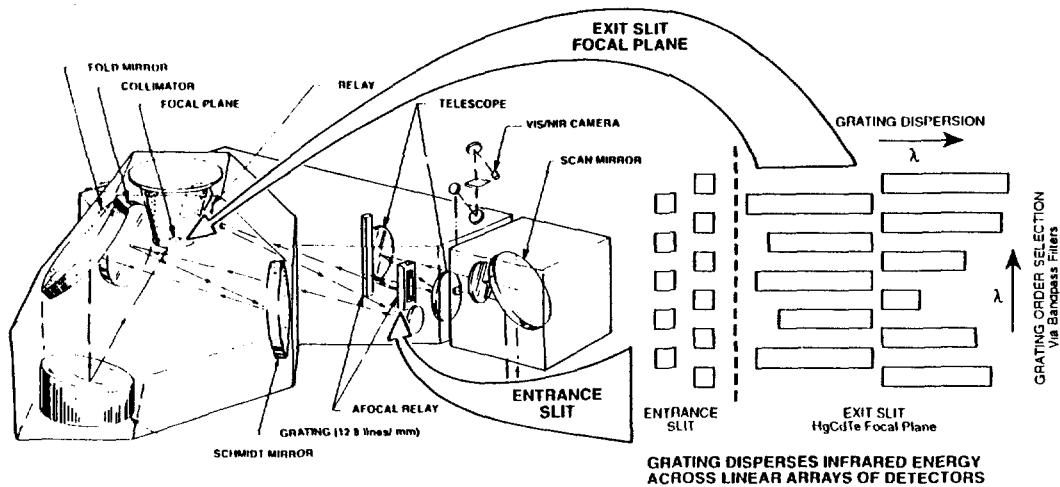


Figure 8. AIRS Array Grating Spectrometer concept.

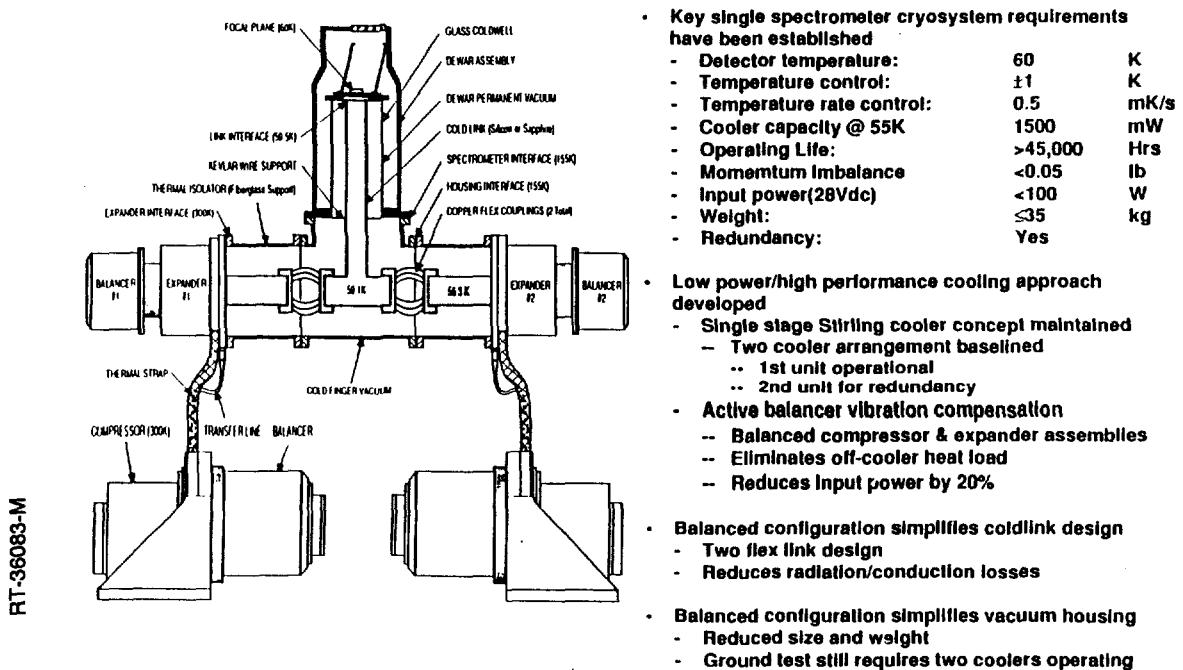


Figure 9. AIRS Cryo-Cooling concept.

Practical considerations in the design of the AIRS sensor include the natural environment radiation effects. AIRS will operate continuously for 5 years in low earth orbit and will be exposed to the degrading effects of the natural space radiation environment, including trapped energetic particles, cosmic rays, solar particle events, ionospheric plasma and auroral electrons. The natural environment radiation effects are being considered in the design through careful choice of shielding materials, component designs that are adequately radiation-hardened and signal processing to mitigate radiation-induced noise effects.

Focal Plane Array Formats

We now consider some aspects of the format of the focal plane array (FPA). The early technology detectors were generally discrete devices. With improved processing capabilities, it is now possible to produce large integrated arrays of detectors. Most applications use integrated arrays of detectors rather than discrete detectors. Most detector arrays require associated preamplifier readout electronics that are physically close, that is, on the FPA. There is unacceptable noise penalty for locating the preamplifier very far from the detector such as in the warm region.

There are two approaches for integrating the detector array and the readout integrated circuit: monolithic or hybrid. In the monolithic case, the detector array and readout electronics are both processed in the same material and produced as a monolithic chip. This is the case for Si CCDs and for PtSi imagers. Initially attempts were made to produce monolithic extrinsic Si IR detectors with Si electronics integrated into the same chip with the Si detectors, but this failed because the processing required for the electronics and for the detectors is incompatible.

The approach that is now universally used for extrinsic Si detectors and for other detector material technologies such as HgCdTe or InSb is a hybrid configuration. The readout is normally based on silicon integrated circuit (IC) technology and is processed separately. The detector array is also processed independently. The detector array and readout IC are then joined in a hybrid assembly with electrical interconnects and physical support through an array of precisely located indium columns that are deposited photolithographically on both the detector array and the readout array. The two chips are then carefully aligned and pressed with sufficient force to deform the soft indium bumps and form a cold weld joint. Indium is the choice material because of its deformation capability and compatibility with photolithographic processing.

Figure 10 shows a typical hybrid focal plane array. Note that the common approach is back side illumination, that is, the IR signal is incident on the detector through the detector array substrate which is transparent to the radiation being detected. The reason for back side illumination is that it maximizes the active area of the detector array that receives the optical radiation. For front side illumination, there is shadowing due to opaque interconnect lines and other structure.

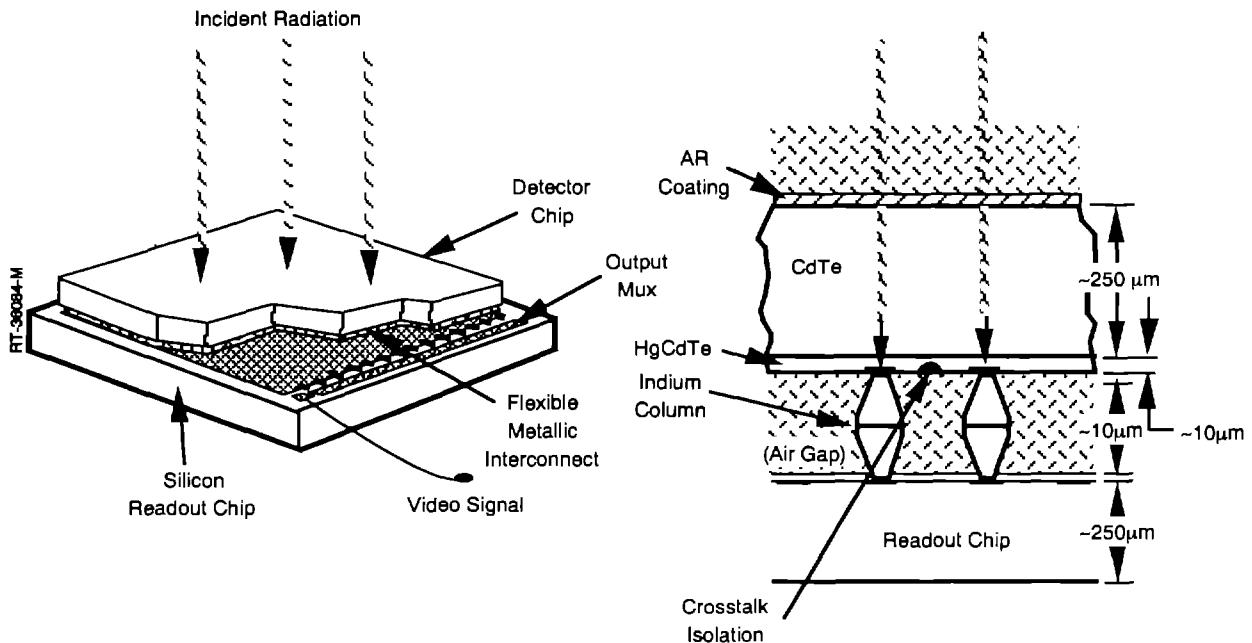


Figure 10. Typical back-side illuminated hybrid FPA.

A variation to the conventional hybrid FPA format is z-plane technology. In this approach, the readout and signal processing electronics are fabricated on multiple chips which are stacked in the plane perpendicular to the detector array. Indium bump bonds on the top of the stack of readout chips connect the detector array. The advantage of this approach is that there is more Si real-estate for fabricating the readout and on-focal-plane signal processing circuitry. The disadvantages include producibility and potential reliability problems.

Another consideration for FPA formats is a scanning system versus a staring system. In a scanning system, the scene is scanned optically across a linear array of detectors. Each detector sequentially sees points on a line though the scene and the data is combined electronically to form an image. In a staring system, the scene is focused on the focal plane and each detector in the array stares at a particular point of the scene for a specified time. Often the staring system will operate in a step-stare mode where the image is optically stepped to a new scene after a specified staring time. The advantage of a scanning system is that a smaller linear detector array can be used, but the disadvantage is that it requires a more complex system to implement the scanning and the detectors are required to have higher sensitivity. The staring approach allows a less sensitive detector to be used because the signal can integrate for a longer time; however a disadvantage is that larger area arrays are needed to achieve the same optical scene coverage.

The primary reason for wanting to increase the array size is to achieve increased resolution. The largest optical arrays at present are 4096x4096 CCDs [3]. Figure 11 (from Reference 3) shows three images of a dollar bill taken with an experimental 4096x4096 CCD. The amazing resolution

achieved with such a sensor is evident. Figure 11-a displays a section of 600 vertical pixels by 1000 horizontal pixels. Figure 11-b is a magnified view of the same scene covering an area of 300x400 pixels. The sub-area represents 0.72% of the FPA. Figure 11-c shows a 112x150 pixel section representing only 0.1% of the active area of the FPA. Individual pixels can be seen at this magnification. The resolution power of this device competes with the resolution capability of photographic film. For example, two football fields set side-by-side could be resolved to 1 inch/pixel.

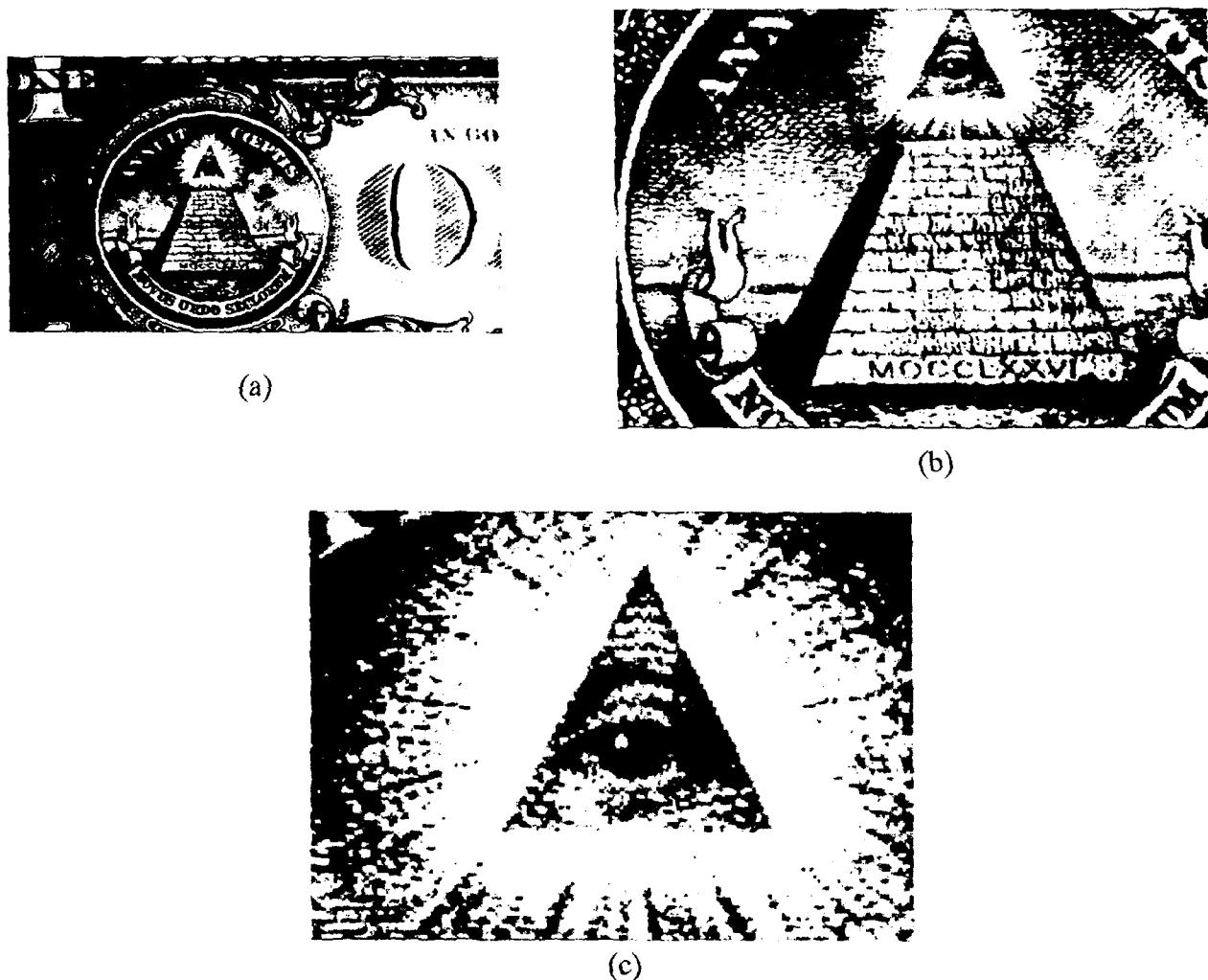


Figure 11. An example of the resolution achieved with a 4096x4096 CCD FPA by imaging a dollar bill. (a) A partial region occupying 600Vx1000H pixel that represents 3.5% of the FPA. (b) Magnified view of a 300Vx400H pixel section representing 0.72% of the FPA. (c) Magnified view showing 112Vx150H pixels representing of 0.1% of the FPA.

Detector and Hybrid Terminology and Figures of Merit

There are several unique terms and figures of merit associated with detectors and hybrids. The principal ones are briefly described in the following.

Responsivity - The current out of the detector per photon power incident on the detector; normally expressed in units of amps/watt.

D* ("dee-star") - Responsivity to noise ratio normalized to area of the detector; units of cm²/Hz/W. This is the primary figure of merit for detectors.

RoA (zero bias resistance - area product) - The slope of the current versus voltage curve for a PV diode at zero bias times the area of the detector; units of Ω-cm². This is a primary figure of merit for PV detectors since it determines D*.

Quantum Efficiency - The ratio of electrons out of the detector to the number of photons incident.

Spectral Response - Normalized response versus wavelength of incident radiation.

Cutoff Wavelength - The maximum wavelength at which the normalized response = 0.5.

Dark Current - The detector current without any incident optical signal. The current is due to defects in the detector which results in leakage current. Obviously lower is better and operation at lower temperature often lowers dark current.

BLIP (Background Limited Infrared Photoconductor) - Operating mode for which the IR background limits the noise performance of the detector. This is the theoretical best performance that can be obtained, and is a function of operating temperature and background of in-band photons. A figure of merit is often quoted in terms of percentage of BLIP.

Pixel - A unit cell in a focal plane array consisting of a detector and a dedicated readout circuit. The name derives from the term "picture element."

Cross-Talk - Coupling of signal or noise between pixels in an array. The coupling can be optical and/or electrical. This is a design consideration in the detailed layout of the detector and readout structures.

Uniformity - The consistency of response, noise, cutoff wavelength, etc., across an array. Imaging systems often have uniformity correction algorithms in the signal processing electronics.

Fill Factor - The percentage of an array that responds to optical signal. For most hybrid FPAs, the signal is incident through the detector substrate (back-side illuminated) to maximize the fill factor.

NEI (Noise Equivalent Irradiance) - The signal flux that gives a signal to noise ratio of one. This is a primary measure of the noise level of the hybrid and the goal is always to minimize this parameter.

NEDT (Noise Equivalent Differential Temperature) - The minimum resolvable temperature difference between two adjacent pixels in an array.

Integration Time - In a staring FPA, this is the time that the signal current is integrated on the detector before readout and reset. Longer integration times allow small signals to be detected provided the noise is sufficiently low.

Linearity - The degree to which the voltage output of the readout preamplifier is linear with photon input to the detector.

Dynamic Range - The maximum signal range divided by the noise level.

Frame Time - Time required to readout the entire array.

CTE (Charge Transfer Efficiency) - The fraction of signal charge remaining after a single stage transfer in a charge coupled device (CCD).

Power - The power dissipated by the FPA is an important figure merit that must be minimized for space-based applications.

3. SYNOPSIS OF DETECTOR AND READOUT ELECTRONICS TECHNOLOGIES

Detector Physics

In this section we will briefly review the basic physics of detectors and relate the detection mechanisms to the primary detector technologies suitable for space applications.

All semiconductor detectors operate on the principle of ionizing interactions of photons with the detector material which generate free carriers in the semiconductor, thereby affecting the electrical properties in a way that can be exploited for detection. There are two fundamental mechanisms for the ionization of carriers: intrinsic and extrinsic excitation. For intrinsic excitation, carriers are ionized across the bandgap of the semiconductor (Figure 12-a). In this case the energy gap E_g determines the cutoff wavelength. Those photons with energy above E_g can ionize carriers from the valence band to the conduction band; those with energy less than E_g (i.e., longer wavelength) do not interact. Si, III-V compounds such as InSb and GaAs, and lead salts such as PbS and PbSe are examples of common detector materials with a fixed bandgap.

In some types of detector materials it is possible to tailor the bandgap to achieve a desired cutoff wavelength. HgCdTe is a common detector material that allows a tailored bandgap by stoichiometric control in the growth of the material. Also, by “bandgap engineering” as is done in superlattice materials, variation in cutoff wavelength can be accomplished.

It is also possible to ionize carriers from shallow impurity levels that are deliberately placed in the bandgap and therefore achieve cutoff wavelengths at longer wavelengths than would be provided by bandgap excitations. This process is termed extrinsic excitation and is illustrated in Figure 12-b. In this case the cutoff wavelength is determined by the impurity level. The advantage of extrinsic detectors is that they can generally operate to longer wavelength but a disadvantage is that they require a lower temperature than intrinsic detectors to achieve a given dark current. This is a result of the requirement to keep the impurities “frozen out” so that they can be ionized only by the optical photons. Extrinsic detector technology was originally developed with impurity dopants in germanium, Ge:X, but the dominant technology is now Si:X.

Once the carriers are excited into the conduction band by either intrinsic or extrinsic processes, we must exploit some change in the electrical properties to detect the radiation. There are three principal photon detection mechanisms. Photoconductive (PC) detectors exploit the change in conductivity by absorption of the photons. Photovoltaic (PV) detectors collect the photocurrent

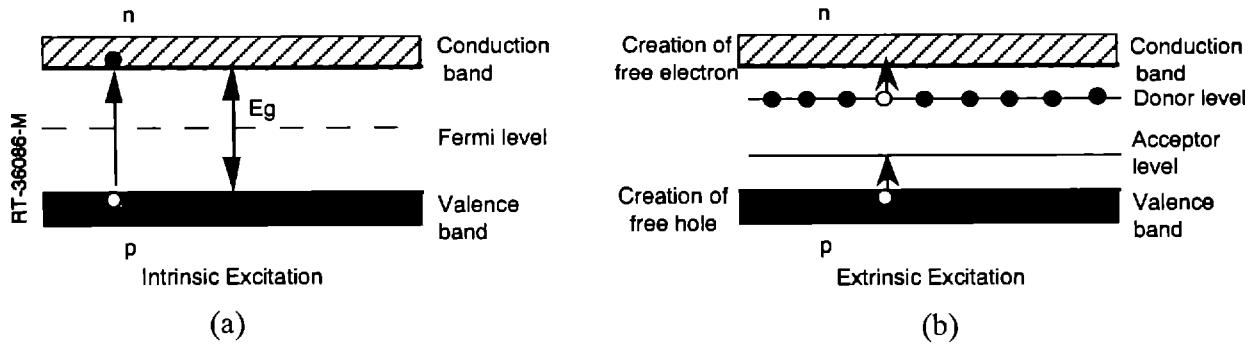


Figure 12. Intrinsic versus extrinsic excitation.

induced by the photons at a built-in potential at a p-n junction. Metal-Insulator-Silicon (MIS) detectors collect photocurrent by integrating into a potential well induced by voltage applied to an MIS capacitor structure.

3.1 Photoconductive Detectors

Figure 13 illustrates the PC detector mechanism. The PC detector is simply a resistor made of photo-sensitive material. When the incident signal photons generate carriers, the resistivity is modulated and this can be detected by an appropriate circuit. A bias voltage must be applied in order to sweep out the generated free carriers before they recombine. Common detector types based on photoconductivity include extrinsic Si, HgCdTe, InSb, Ge, PbS, PbSe, and superlattice detectors. Power dissipation can be a limitation for use of PC detectors in space applications. An exception is extrinsic Si, which has a very low dark current due to freezeout of the impurities.

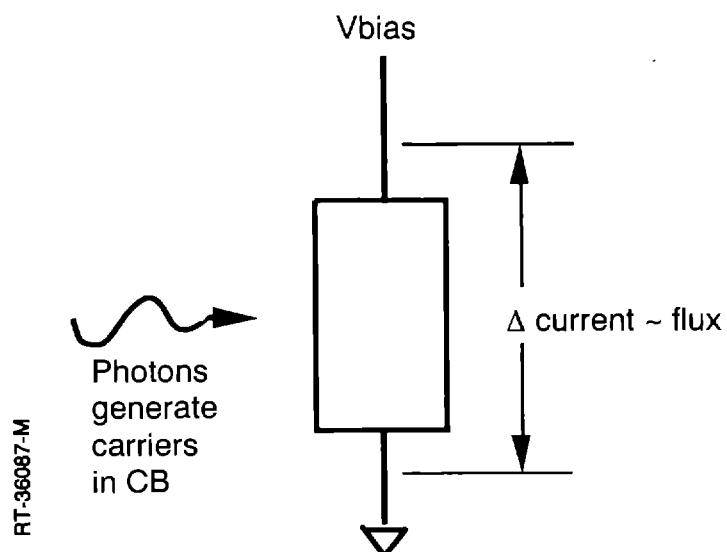


Figure 13. Photoconductive (PC) detectors are resistors made of photo-sensitive material.

Extrinsic Si:As detectors is the dominant technology for the VLWIR wavelength regime because of their high sensitivity and the lack of any competing technology for these wavelengths. This technology is relatively mature, having been developed for military applications. The cutoff wavelength is $\sim 26 \mu\text{m}$ and the operating temperature is $\sim 10 \text{ K}$. Other dopants have been developed for extrinsic Si. Si:Ga operates at $\sim 20 \text{ K}$ with a cutoff wavelength of $\sim 18 \mu\text{m}$; Si:In operates at $\sim 80 \text{ K}$ with a cutoff wavelength of $\sim 8 \mu\text{m}$. However these technologies do not compete in the MWIR and SWIR regimes where HgCdTe and InSb dominate.

For operation in the standard PC mode, the Si:As detector is required to be relatively thick, on the order of $100 \mu\text{m}$, in order to have sufficient interaction probability between the photons and the impurity dopants and give sufficient quantum efficiency. However, for space and military applications, a thick detector is a disadvantage because of gamma or electron induced noise from the radiation environment. To address this ionization-noise issue, the technology was driven toward thinner detectors. However, in order to achieve sufficient quantum efficiency, the doping concentration of the optically active impurity, arsenic in this case, must be increased to the level that impurity banding occurs. That is, the impurities are sufficiently close that there is high probability of hopping of carriers between partially ionized impurities, resulting in impurity band conduction. The consequence of this for a normal PC mode detector is unacceptable dark current. To address this problem, the impurity band conduction (IBC) detector was developed.

IBC detectors are a specialized form of extrinsic Si photoconductor. The active layer is made very thin, on the order of $30 \mu\text{m}$, and is heavily doped with the As impurity to achieve adequate quantum efficiency. The high concentration of As results in an impurity band but the dark current is blocked by addition of a thin layer of undoped Si which serves as a blocking layer to the impurity band conduction. However, extrinsic excitation of electrons from the As donor impurities can still be detected since there is no barrier to these carriers. A schematic cross-section of the IBC detector is shown in Figure 14. The thin active layer is grown by molecular beam epitaxy (MBE) on a thick Si substrate. A transparent contact is provided at the interface between the substrate and the active layer. The IR is incident through the substrate, which is transparent to the long wavelength radiation. An intrinsic blocking layer is grown on the epitaxial active layer. A bias is applied which sweeps out photo-generated electrons and associated ionized donor impurities as the signal. The dark current is blocked by the potential barrier at the intrinsic layer. The key to achieving this device is well controlled epitaxial growth and doping.

By use of precisely controlled epitaxial growth to fabricate thin alternating layers of different materials or of different doping, it is possible to “engineer” the bandgap in some materials to establish potential wells with appropriate energy depths to act as detectors of IR photons. Such semiconductors with engineered bandgaps are known as superlattice. Superlattice detectors can be fabricated to operate in either intrinsic or extrinsic mode. There are several approaches that have been used to

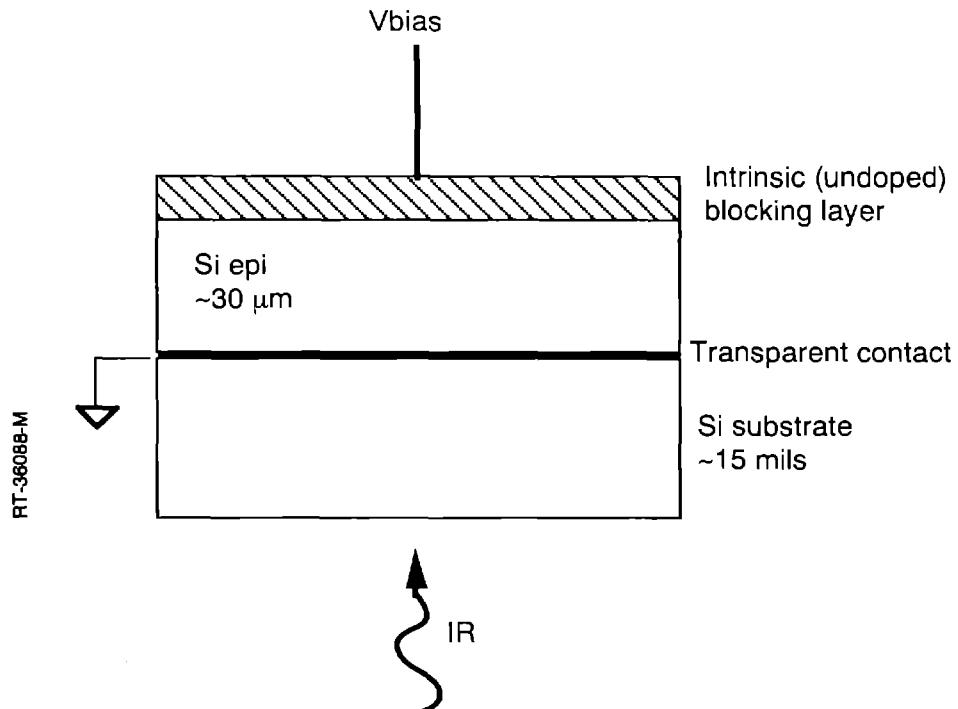


Figure 14. Impurity band conduction (IBC) detectors are a specialized form of extrinsic Si photoconductor.

establish photo-active potential wells for IR detection including multiple quantum wells (MQW), strained layers (SL), and delta doping. It is possible to adjust the depth of the trapping levels based on the structure and dimensions, thereby allowing an adjustable cutoff wavelength.

The AlGaAs/GaAs MQW technology is the most developed of the superlattice detectors; this technology is illustrated in Figure 15. The detection mechanism is extrinsic with a photoconductive detector. The advantages of the MQW detector are superior radiation hardness, both for total dose and for gamma noise, and superior uniformity, especially compared to MWIR and LWIR HgCdTe where this technology is competitive. The principal disadvantage is relatively poor quantum efficiency. Development is continuing in the AlGaAs/GaAs technology as well as other superlattice technologies.

3.2 Photovoltaic Detectors

Photovoltaic (PV) detectors are simply a p-n diode formed in a photosensitive material. The photo-generated carriers give rise to a photocurrent which is collected in the electric field at the junction depletion region. Also carriers generated within the bulk semiconductor which diffuse to the junction are collected. The operation is illustrated in Figure 16. The detectors operate with zero applied bias or a small reverse bias. The change in reverse current that is induced by the photon

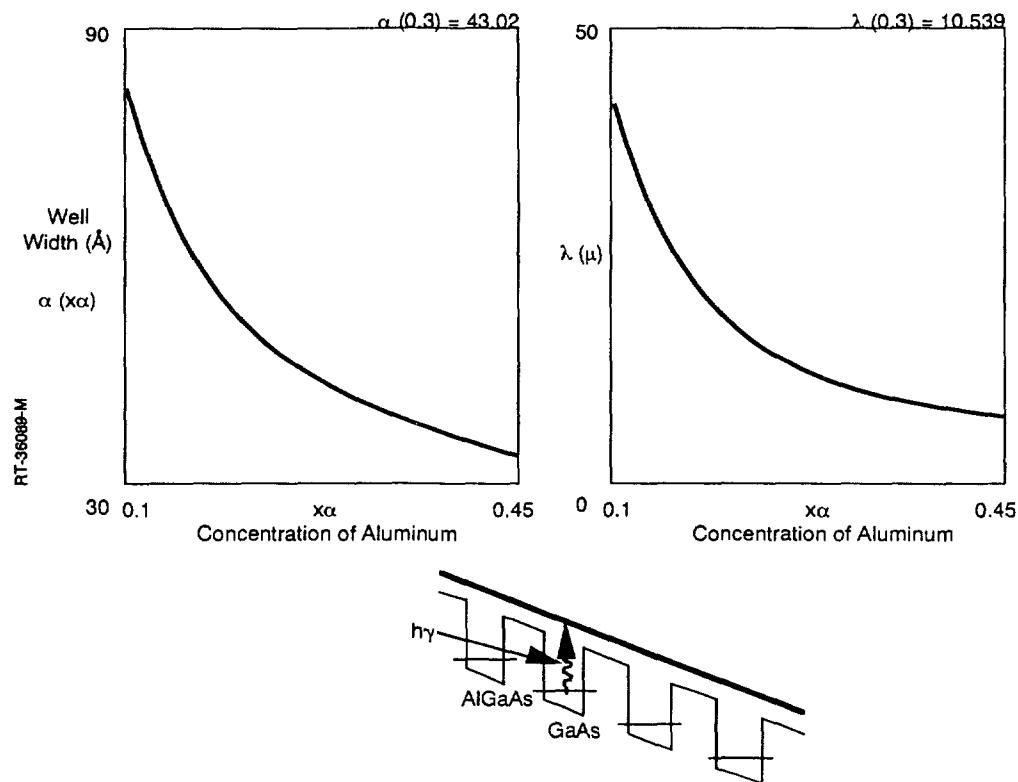


Figure 15. Superlattice detectors rely on bandgap engineering to establish potential wells.

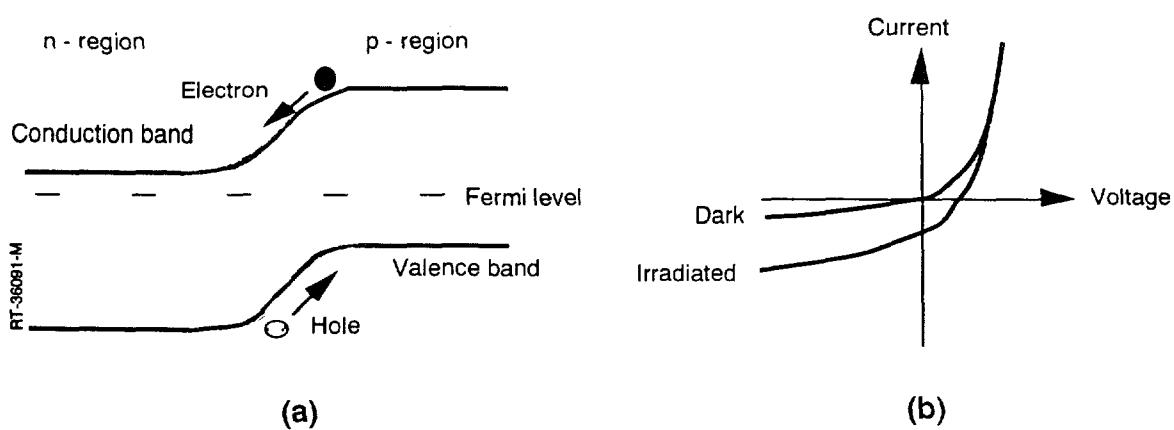


Figure 16. Photovoltaic (PV) detectors rely on built-in voltage.

radiation serves as the detection mechanism. For the IR region, HgCdTe and InSb are the dominant PV detector technologies. For the visible and UV regions, Si and GaAs or GaInAs are the dominant technologies.

HgCdTe detector technology is a fairly well developed technology for the MWIR regime, less so for SWIR and much less so for LWIR and VLWIR. By virtue of the adjustable bandgap in this material, cutoff wavelengths from 1 to 20 μm are possible for PV and MIS modes, and up to 26 μm has been reported for PC mode. However, for cutoff wavelengths $>13 \mu\text{m}$, the technology is very immature.

Development efforts are underway for the LWIR regime. There are really two distinct technologies for LWIR HgCdTe: high temperature and high performance. The high temperature technology operates at temperature $>65 \text{ K}$ and has its performance limited by diffusion and or generation/recombination (g-r) currents. The high performance technology operates at 40-65 K under low background conditions where the performance is limited by leakage current from tunneling.

HgCdTe detectors are used in all three of the principal detection modes, PV, PC and MIS. Since low power is needed for space applications, it is primarily the PV and MIS modes that are appropriate.

The advantages of HgCdTe detectors include high D* and higher operating temperature than Si:X and InSb. However, until recently the material has had problems with low total-dose hardness and with worse nonuniformity and worse producibility compared to the competing technologies. Significant progress has recently occurred in uniformity, producibility and in total-dose hardening and LWIR HgCdTe is now becoming competitive with IBC Si and InSb for some applications, primarily because of its significant advantage of being able to operate at higher temperature than Si or InSb.

InSb detectors are a competing technology with HgCdTe for the MWIR regime. InSb can function as a detector in both the PV and MIS modes, however the most mature technology is in PV mode. This is a relatively mature technology for MWIR array applications. IR cameras based on 256x256 InSb arrays operating in the MWIR are commercially available. A disadvantage of InSb compared to HgCdTe is that it requires a lower operating temperature for equivalent performance. For use in the visible and UV wavelengths, thinning of the substrate is required. Again, InSb has a disadvantage compared to the competing technology of Si-CCDs in the visible/UV regime because it requires more cooling.

GaAs photodiode technology has been developed for detection in the visible, primarily for star-tracking applications. It operates at room temperature and has spectral response up to 0.85 μm . There have been other ternary or quaternary materials based on GaAs developed as detectors for other wavelengths. An example is the fairly mature technology of InGaAsP which is used as a light emitting diode (LED) and PIN photodetector for fiber optic communication.

3.3 Metal-Insulator-Semiconductor Detectors

Metal-Insulator-Silicon (MIS) detectors integrate photocurrent in a potential well that is established in photoactive material. A bias is applied to a conducting gate on an insulator over the detector material, as illustrated in Figure 17. Carriers are generated by interaction of photons, usually incident through the substrate, and are collected by the electric field at the bias-induced depletion region under the gate. Carriers are collected both from ionizations within the depletion region and those nearby that diffuse to the region. Various schemes are used for readout of the charge such as charge transfer in a CCD, connection to an external preamplifier through addressable bus lines as in a charge injection device (CID), or direct connection to a preamplifier within the unit cell (either monolithic or hybridized).

Examples of common detector materials that are used in the MIS mode include Si-CCD, HgCdTe and InSb.

In the Si-CCD imager, the photosignal is converted directly to charge in the detector, as illustrated in Figure 18. A potential well is established by application of a bias voltage and charge generated by the optical signal is collected. The potential well is spatially confined by implanted potential barriers. A MIS capacitor structure with coupled gates is used to transfer charge out of the array. By proper sequencing of the voltage pulses applied to the gates, the signal charge packet can be moved along the array to the edge, where a serial register transfers the charge packets from each column sequentially into the signal processing electronics. Signal processing then allows an image to be restored.

Most imaging applications are back-side illuminated to improve the fill factor. CCD technology is under development for UV detection, based on thinning the substrate so that the UV can penetrate to the active region.

PtSi are a special case of a MIS detector, where the potential well is forced by a Schottky barrier with Pt on Si. Photocarriers are generated in the metal and stored in the potential well of the Si. The concept is illustrated in Figure 19. A PtSi electrode on p-type silicon emits electrons when photons of adequate energy are incident. The cutoff wavelength ranges up to $\sim 6 \mu\text{m}$. The photo-emitted electrons are collected in the depletion region in the Si under the electrode and replicate the IR image on the electrode array. An integrated monolithic detector array includes the PtSi Schottky diode array, a multiplexer network and charge sensing video amplifiers.

A principal advantage of PtSi detector arrays is that they are based on standard Si processing, therefore having very good uniformity and producibility, especially compared to HgCdTe. However, a major disadvantage is a relatively low quantum efficiency, in the range of 0.4 to 1%, and a requirement for cooling to 60-80 K.

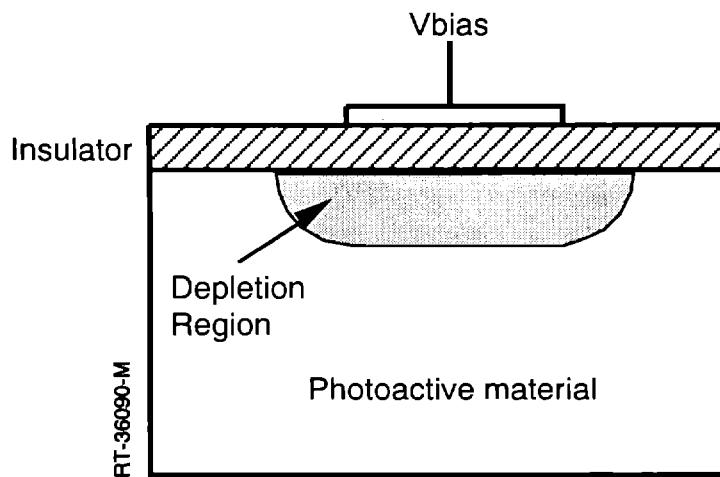


Figure 17. MIS detectors integrate photocurrent in a potential well.

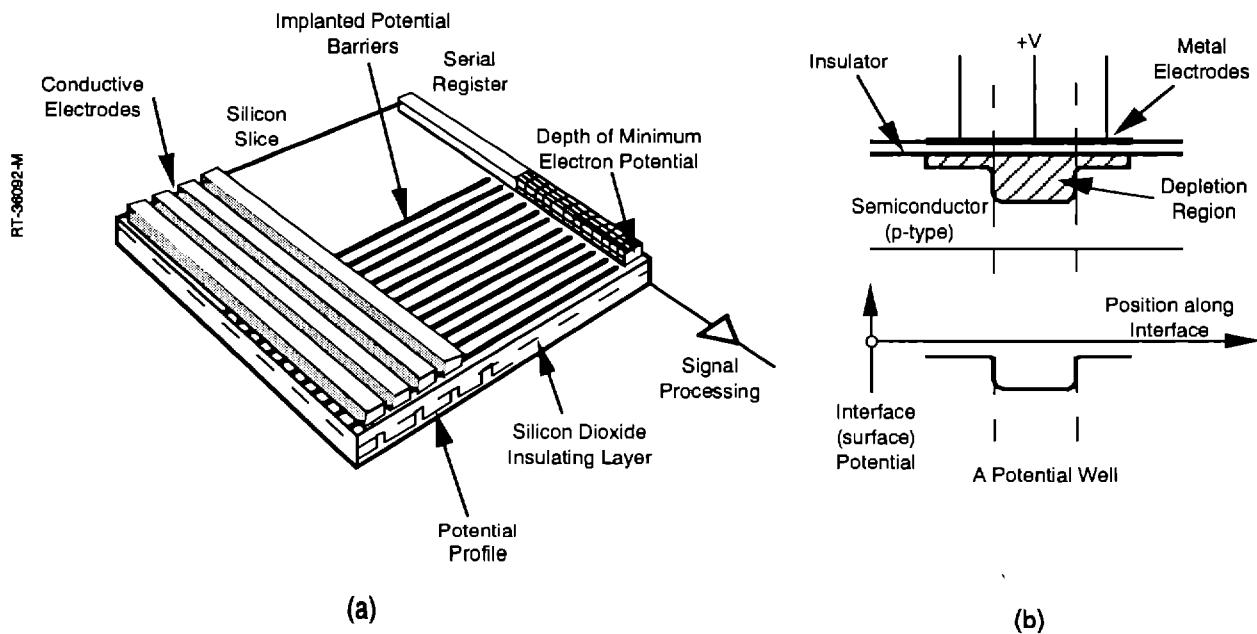


Figure 18. Si CCDs can be used as visible and UV imagers.

3.4 Cryogenic CMOS

Readout electronics for IR detectors are usually required to be operated at cryogenic temperatures since the detectors have to be cooled and there is unacceptable noise penalty for locating the preamplifier electronics distant from the detectors in the warm region. For this reason, most electronics development for IR detector readout applications has been in Si MOS technology, as

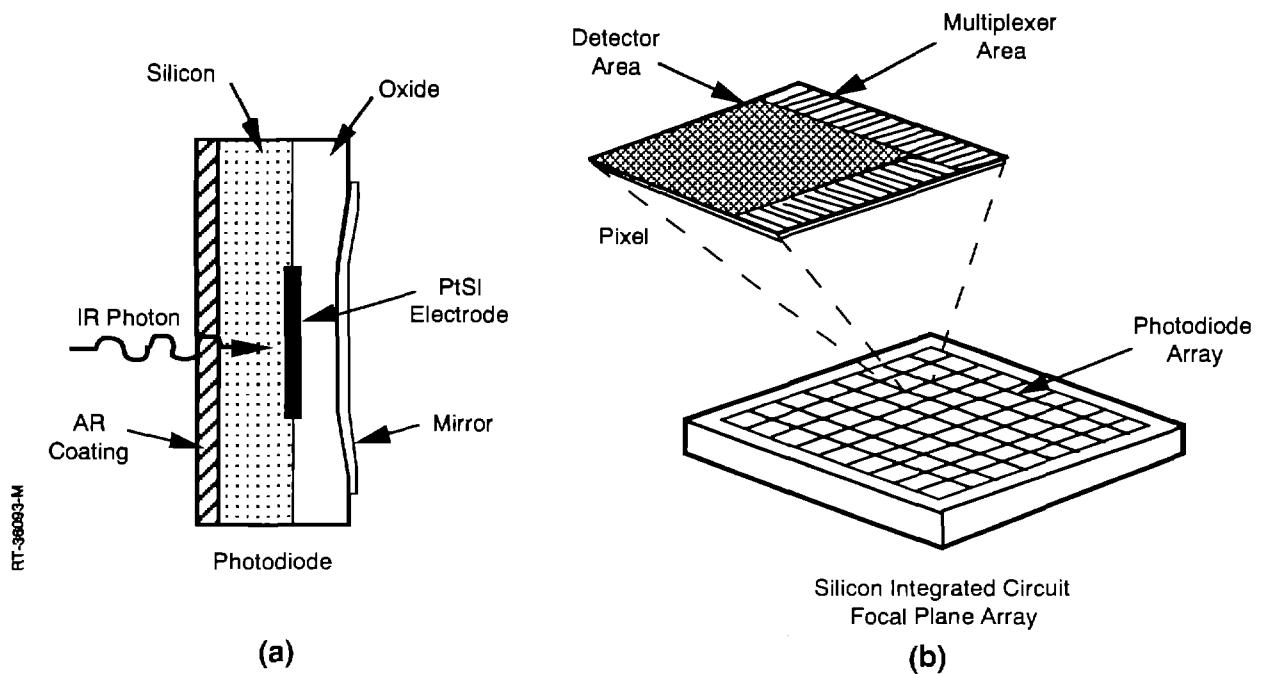


Figure 19. PtSi detectors.

opposed to bipolar, since bipolar suffers from reduced gain at low temperature. Also, since power is a concern for FPAs, CMOS is the preferred technology.

CMOS works quite well at cryogenic temperatures down to ~40 K, and various studies have led to well-understood models for the device temperature dependencies [4,5,6,7,8]. At lower temperatures, carrier freezeout problems begin to compromise the performance, although some technologies with proper design can be made to operate acceptable well to temperatures below 10 K [9]. At temperatures above 40 K, Si CMOS actually performs better than at room temperature as a result of increased mobility due to less phonon scattering of the majority carriers in the channel. However, the process must be designed for operation at cryogenic temperature. The major design issues are: the change in threshold voltage as temperature changes, the necessity to avoid carrier freeze-out at the desired operating temperature and approaches to mitigate hot electron trapping in the gate dielectric [10].

3.5 Superconducting Electronics (SCE) and Detectors

Superconductors can be divided into two types: the original low temperature superconductor (LTS) which require operation at temperatures less than 10 K, and the recently developed high

temperature superconductor (HTS) which can operate at temperatures of 77 K or greater. Until 1987, the only superconductors known operated at temperatures less than 10 K and were required to be cooled with liquid helium. The LTS technology is based on niobium, or niobium nitride which is somewhat less mature. In 1987, superconductivity above 77 K (liquid nitrogen temperature) was discovered in copper oxides and this made superconducting technology considerably more practical.

Superconductors can be used for both electronic signal processing applications and detector applications, although the electronic applications are much more developed than the detector applications.

For electronics, the thin electrode Josephson Junction (JJ) technology has been under development for quite some time. A Josephson Junction consists of two superconducting electrodes separated by an insulator which is thin enough that electrons can tunnel through the barrier. A JJ can serve as a switch because it exhibits zero resistance when the current through it is less than some critical current. If the current exceeds the critical current, the junction will switch to a finite resistance state and remain in that state until the current is reduced to zero. It is possible to build logic devices with very high switching speeds with JJs.

Another important application of superconductors to electronics is in fabricating planar microwave integrated circuits [11]. Conventional microwave and RF components tend to be fairly large and require expensive waveguide and coaxial components. With HTS technology, it is possible to build planar microwave ICs to specifications that exceed those of their waveguide and coaxial counterparts.

Superconducting bolometers have been developed as an IR detection mechanism. Recently, other detection mechanisms have been under development. The potential advantages of superconducting detectors include broadband detection, high speed and low noise. However, the technology is very immature.

A series of space experiments are planned for in-space evaluation of HTS electronics with the launch of the first platform (HTSSE-1) in 1993 [12].

4. STATUS OF COMPONENT TECHNOLOGIES

In this section we will review the status of the major detector and readout electronics that are suitable for space applications.

4.1 Status of Detector Technology

Figure 20 lists the detector technologies that could potentially be used for space applications for the various wavelength bands. The darkened listings are the dominant technologies and the others are alternates. We will discuss the status of each of the dominant technologies.

VLWIR

The primary detector technology for VLWIR is IBC Si:As. There is no alternate technology for this wavelength regime. These detectors require operating temperature in the 8-12 K range and

<u>VLWIR (14-26 μm)</u>	<u>LWIR (8-14 μm)</u>	<u>MWIR (3-8 μm)</u>
Si:As (IBC,PC)	HgCdTe (PV,MIS,PC)	HgCdTe (PV,MIS,PC)
Si:Ga (PC)	Superlattice (PC,PV)	InSb (PV,MIS)
HgCdTe (PC)	Ge:Hg (PC)	PtSi (PV/MIS)
		PbSe (PC)
		Si:In (PC)
		Superlattice (PC,PV)
<u>SWIR (1-3 μm)</u>	<u>VIS/NIR (0.4-1 μm)</u>	<u>UV (0.1-0.4 μm)</u>
HgCdTe (PV,MIS,PC)	Si-CCD (MIS)	Thinned Si-CCD
PbS (PC)	Si-CID (MIS)	GaAs (PV)
	GaAs (PV)	

Figure 20. Major detector technologies suitable for space applications.

generally operate under low background conditions. That is, they are used to look out at space rather than at the earth. The primary technical applications for which this detector technology has been developed are cold-body detection, tracking and discrimination. However, there have also been very successful applications in astronomy. Array sizes up to 128x128 have been produced.

LWIR

PV HgCdTe is the dominant detector technology for LWIR space applications for both high temperature, high background applications and high performance, low background applications. The applications are cold-body detection and tracking and some limited discrimination capability. The high temperature (65-80 K) technology is used in moderate background applications, such as looking toward the earth. Array sizes up to 256x256 have been demonstrated. The high performance technology (40-65 K) is used in low background applications and its performance is limited by tunneling currents. Largest demonstrated array sizes for this technology is 128x128.

Superlattice detectors are an alternate technology for this wavelength regime but lack sufficient sensitivity with the present state of development to be competitive.

MWIR

HgCdTe is also the dominant technology in the MWIR regime. For these wavelengths, both PV and MIS technologies are viable. These detectors usually operate in the 80 to 120 K temperature range under moderate to high backgrounds. A primary application is hot and warm body detection and tracking. Some very limited discrimination capability is also possible. The largest array sizes demonstrated are 512x512 and 640x480.

Alternate technologies for MWIR include PV InSb and PtSi. The disadvantage of InSb is that it generally requires lower operating temperature than HgCdTe to achieve the same performance. The limitation to PtSi is relatively poor quantum efficiency. Array size up to 640x480 has been demonstrated for InSb and up to 1024x1024 has been demonstrated for PtSi.

SWIR

The primary detector technology for SWIR is HgCdTe; again both PV and MIS technologies are possible. These detectors operate at temperatures above 120 K under moderate to high background conditions. The military application is hot body detection. The environmental application is earth resource surveillance.

The alternate technology is PtSi but the disadvantage is low quantum efficiency.

Visible

Silicon CCDs are the dominant detector technology for visible imaging. These can operate at room temperature but improved performance is achieved by cooling to ~270 K. Applications include earth surveillance, space surveillance, and star tracking for navigation. Array sizes up to 4096x4096 have been demonstrated. Such large array sizes provide high resolution imaging and are useful for closely spaced object resolution. There is presently no demonstrated alternate technology for visible imaging.

UV

There is really no mature detector technology for UV imaging. Technology is under development for thinning Si CCDs for use in the UV, however this is immature. A possible technical application would be orbital object tracking.

4.2 Status of Electronics for Sensor Applications

In this section we review the status of electronics for sensor applications, including the cold electronics that are on or near the focal plane and the signal processing on the sensor before communication to the user.

4.2.1 Cryogenic IC Device Technology

Si CMOS

One of the thrusts in sensor technology development is toward putting more of the signal processing electronics on the focal plane array, that is, in the cryogenic regime. The drivers for this are reduced data rate off the focal plane and fewer connections between the warm signal processing area and the cold on-FPA electronics; these factors lead to reduced system weight, power and complexity and ultimately to reduced cost.

Cryogenic analog Si CMOS has been under development for several years and considerable progress has been made. Areas needing further improvements are higher density and better electrical performance at very low (<20 K) temperature for application to extrinsic Si detectors. Total-dose hardness can generally be managed to required levels using thin gate dielectrics and guard bands, although with some density penalty. Reliability may be an issue for cryogenic electronics due to increased hot electrons.

GaAs

GaAs has some potential advantages compared to Si-CMOS for FPA applications — better thermal expansion match to HgCdTe to allow large format arrays, not subject to carrier freezeout at

10 K, inherently high total dose hardness, and higher speed. There have been limited efforts for developing a cryogenic analog GaAs technology suitable for focal plane applications. A low-power, low-noise technology is needed. However, the progress in cryogenic Si-CMOS technology and development of other approaches to mitigate thermal mismatch in hybrids has diminished the drive for GaAs in the sensor FPA application.

4.2.2 Signal Processing Technology

Focal plane array applications and sensor system designs are varied and cover many orders of magnitude in the design parameters for the signal processing [13,14]. Many applications can take advantage of high optical signal levels or use fast optics to simplify the FPA architecture and allow use of simpler detector interfaces such as CCDs or Schottky arrays, as discussed earlier. However when the signal and background levels are low or optical aperture area is expensive, more complex readout circuit considerations are required.

On-Focal-Plane Readout and Preamplification

Figure 21 shows the generic functional diagram for a unit cell readout and preamplifier for a low-background hybrid, as is typical for many IR applications. Readout circuits in these hybrid arrays generally take photodiode current (or photocurrent from a PC or IBC detector) and inject it into a temporal integration circuit for each detector on the array. After the sense node is reset to a fixed voltage level, charge is integrated on a capacitor; this provides a current-to-voltage function. Such circuits are generally called a trans-impedance amplifier (TIA). A low output impedance buffer then feeds the sense node voltage to a switched MOSFET multiplexer which drives the output signal. The output multiplexer reduces the number of output leads and optimizes use of downstream electronics.

The circuitry from the detector output to the multiplexer input is called a unit cell amplifier and is replicated for each detector on the array. The multiplexer cycles through the unit cell amplifiers and the output voltage from each detector channel is sampled. The output of the multiplexer can be sent to signal processing electronics and analog-to-digital converters (ADCs).

Figure 22 shows simplified functional diagrams of four common readout circuits used with IR detectors: the Source Follower per Detector (SFD) [15-17], Direct Injection (DI) [18-20], Buffered Direct Injection (BDI) [18-21], and Capacitive Trans-Impedance Amplifier (CTIA) [21-24]. Each circuit has a temporal integration circuit composed of a capacitor and a FET switch. After the common node is reset to a fixed voltage level, photocurrent from the detector is integrated on the capacitor at the sense node, generating a voltage that varies with time under illumination. The source follower is a near-unity gain amplifier which enhances drive capability for sending signals into the multiplexer and off the FPA for further processing.

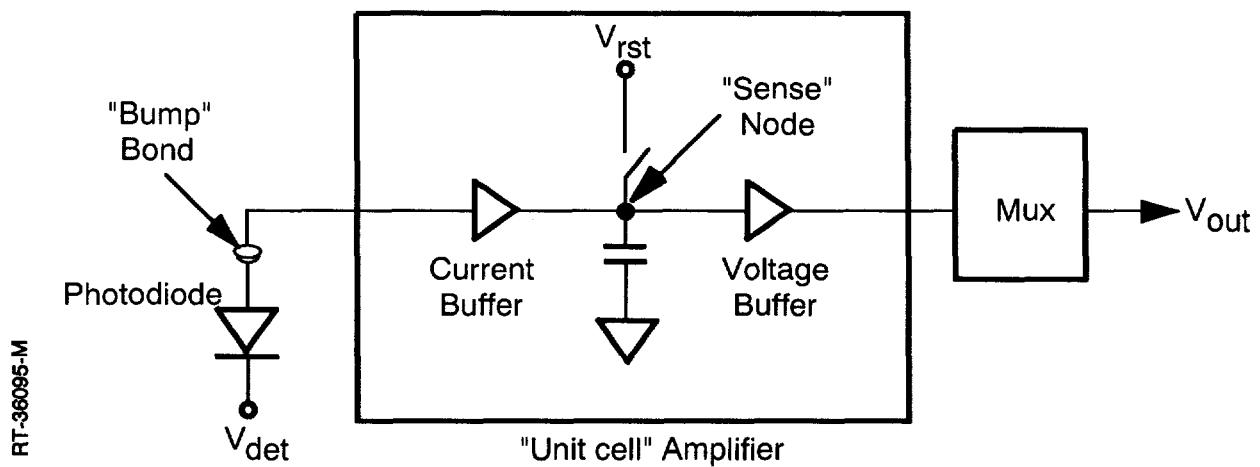


Figure 21. Functional diagram for unit cell amplifier on a hybrid FPA.

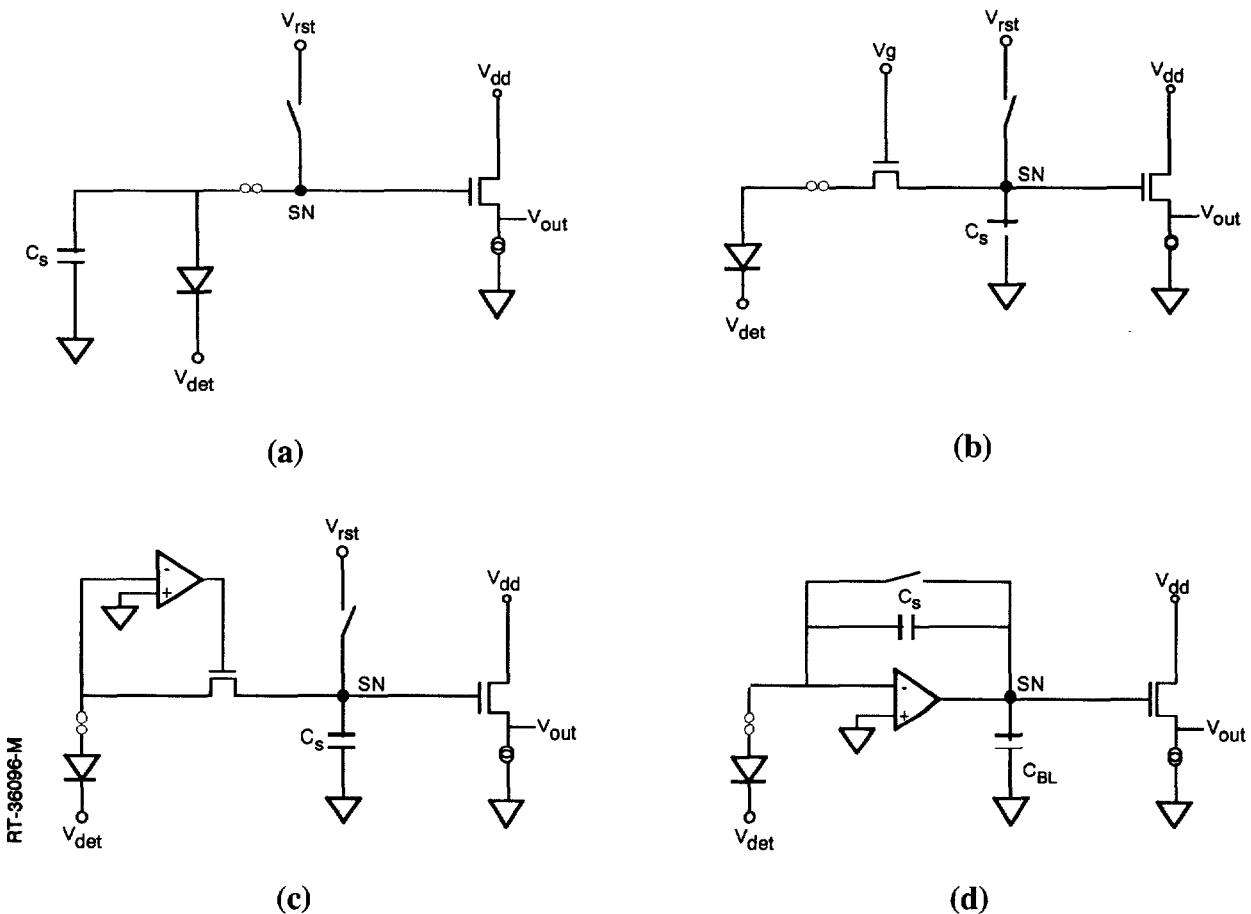


Figure 22. Functional diagram for common readout circuit technologies. (a) Source Follower per Detector. (b) Direct Injection. (c) Buffered Direct Injection. (d) Capacitive Trans-Impedance Amplifier.

On-Focal-Plane Signal Processing

Signal processing on the focal plane is typically limited because of the lack of room in the unit cell to implement complex electronic circuits. Processing functions that are usually included on the focal plane following preamplification are filtering and multiplexing the analog data.

Another processing function that is often performed on the focal plane is Correlated-Double-Sampling (CDS). In CDS, correlated measurements of the signal voltage and the reset voltage at each pixel are made and the output is the signal minus the reset voltage. This scheme eliminates the reset noise.

A processing function that is sometimes performed on the focal plane is Time-Delay-and-Integrate (TDI). The basis of TDI is to combine several signal packets, which serves to increase the signal to noise ratio since approximately the same noise level is present in each sample. This signal processing approach is applicable to line-scanned arrays. Such approaches are often used to mitigate the effects of gamma noise spikes.

Post-Focal-Plane-Signal-Processing

The post focal plane signal processing functions consist of analog to digital (A/D) conversion, Time Dependent Processing, Object Dependent Processing and Mission Dependent Processing. The names are descriptive of the general functions. Figure 23 illustrates the overall signal processing chain from the detector to the communications link. We have already discussed the on-focal-plane processing and now turn to the post-focal-plane processing functions.

Analog to Digital Conversion: For noise immunity and processing efficiency, it is desirable to convert the analog signals to the digital domain as soon in the signal processing chain as possible. At present, ADC is normally accomplished off the FPA in the warm region, but technology for performing the ADC function on the FPA is under development. There can be a considerable advantage to performing ADC on focal plane since the number of interconnects between the cold and warm regions can be reduced. However, the difficulties lie in achieving the functions in low power, high density formats.

Time Dependent Processing: TDI, correction for responsivity nonuniformity, gamma suppression, dynamic range compression, and digital multiplexing are examples of signal processing functions that have temporal dependence.

Object Dependent Processing: Examples of signal processing functions that are dependent on the object (target) being analyzed include pulse matching, thresholding, closely spaced object (CSO) resolution, and color coincidence.

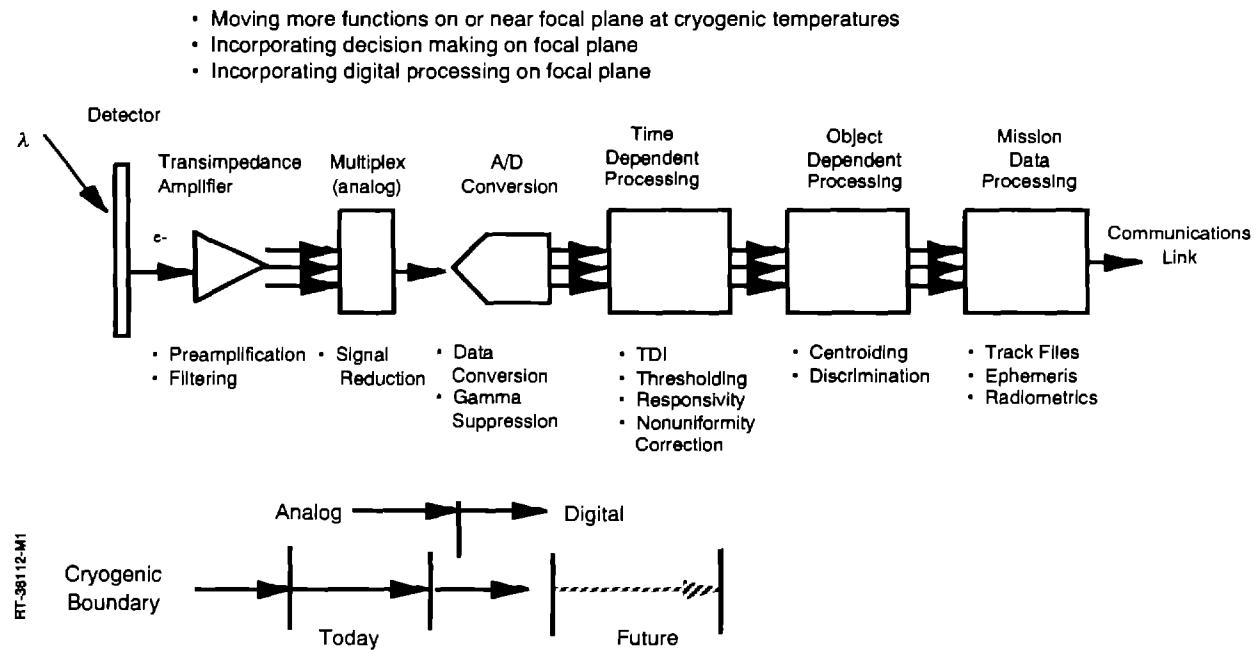


Figure 23. Major technology thrusts for signal processing.

Mission Data Processing: Examples of mission-dependent functions include target classification, generation of track files, ephemeris files and radiometrics.

4.2.3 Subsystem Interconnects and Data Bus

In the internal workings of a satellite, it is no simple matter to interconnect all the subsystems. With optical imaging sensors in particular, very large amounts of data must be handled at high speed.

Future generation satellites will include more complex sensors, high speed on-board data processing and high speed communication subsystems. Application of Fiber-Optic-Data-Bus (FODB) as an approach to high speed data transfer between satellite subsystems is presently under investigation.

The concept is use of a light emitting diode (LED) transmitter and PIN photodetector coupled with a fiber optic cable (Figure 24). Performance requirements for FODBs in satellite applications are driven by latency, bit error rate (BER) and optical signal loss. Latency is the time between appearance of a signal on the data bus and the time the signal is received at the destination. This is a design parameter that depends on the speed of the transmitter and receiver components. Bit error rate is the ratio of the number of incorrect bits to the total number of bits transmitted. Proton-induced upset is a source of error from the radiation environment. Optical signal loss can occur due to

misalignment of the optical fibers in connectors, absorption of signal light in the fiber cladding and increased absorption in the optical fiber due to total-dose-induced darkening of the fiber.

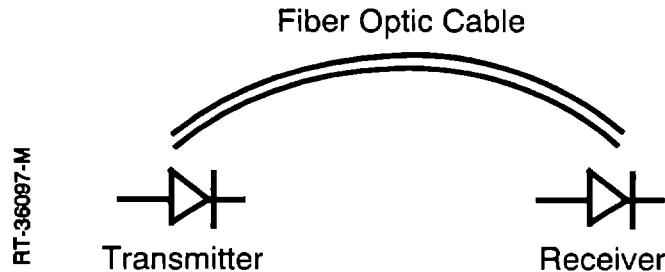


Figure 24. Fiber Optic Data Bus for sensor application.

Advantages of FODBs are high data rate, low weight and less complexity compared to conventional wiring harnesses. Initial space experiments in the Long Duration Space Experiment (LDEF) satellite [25] and recent component tests [26] that simulate space radiation effects indicate the feasibility of fiber optic data busses for satellite applications.

4.2.4 Cryo-Cooling Technology

There are four general categories of cryo-cooler technology [27]: stored cryogens, passive radiators, thermoelectric and closed cycle (Stirling) mechanical refrigerators.

Stored cryogens can be either liquid, solid or gas. Gases are used for cooling by Joule-Thompson (J-T) expansion through a nozzle. Solids and liquids are stored as a cold material such as liquid helium or solid carbon dioxide, and cool the focal plane by conduction. Liquids generally require supercritical storage for space. Solids are the most suited for long term missions. Temperatures in the range from 4.2 to 150 K can be achieved with stored cryogens.

Passive radiators are based on a large surface area to radiate energy to the low temperature sink of space. These are effective in the 80 to 200 K range. The lower temperatures only apply to a small load and require very large areas for the radiators.

Thermoelectric coolers use electric energy to achieve cooling. They are effective in the temperature range from 145-230 K. The disadvantages are that they are useful only for small loads and they require high power.

The closed cycle mechanical refrigerator operates on the principal of the Stirling cycle. They can operate at temperatures down to 10 K, however, the lower temperatures require high power. The practical technology that has been developed to date is limited to about 60 K. Cooling to 65 K with 0.25 watt load is about the present state-of-the-art. Reliability, vibration and efficiency are practical

concerns. Commercially available coolers claim lifetime specifications in the 4000 to 5000 hour range. Ultimately a lifetime of 10,000 hours is anticipated with the present Twin Opposed Piston (TOP) linear Stirling coolers [28].

5. UNIQUE RADIATION EFFECTS

One of the major considerations that must go into the design of sensors and associated electronics for use in space is the hostile radiation environments. These environments includes trapped electrons and trapped protons, solar flare protons, and galactic cosmic ray protons and heavy ions. In addition hardened sensors must design for potential nuclear weapon effects. This section discusses the unique aspects of radiation effects on detectors and sensor electronics. Many of the basic mechanisms are the same as the familiar radiation effects on room temperature electronics, however there are some unique considerations for the sensor applications and some important differences in the mechanisms.

Optical signal currents and charges in detectors are very small and frequently near the noise level. Detectors are large charge collection volumes for ionizing radiations such as electrons and protons. As a consequence, even a single electron can easily produce enough charge to mask the signal charge when it passes through a detector. This leads to the unique sensitivity of detectors to ionizing noise. This is often referred to as gamma noise, since this is one of the major environments of concern in hardened sensors. However, ionizing electrons and protons from the natural environment also cause ionization noise that can be a problem in some applications. The electrons can readily be shielded but the protons cannot. This can be a particular concern for operation within the trapped proton belts or through the South Atlantic Anomaly, as was discussed in the second paper.

Another unique aspect of IR sensor electronics is the fact that the on-FPA and near-FPA electronics are operated at cryogenic temperature. There is enhanced charge trapping in oxides at cryogenic temperature because holes are nearly immobile while electrons have high mobility. The result is that electron-hole pairs that are generated by ionizing radiation can become separated by electric fields within the dielectric and the electrons are swept out but holes are not. The holes move slowly through the oxide by stochastic hopping but, for all practical purposes, are immobile on time scales of interest. This results in a positive charge buildup in the oxide which alters the surface potential in the silicon beneath. The consequence of the charge buildup in gate dielectrics of MOSFETs is threshold voltage shift; the consequence in field dielectrics is leakage currents. In addition there are deleterious effects in detectors as discussed below.

Because detectors are sensitive to optical radiation, any source that produces a high background of in-band radiation can mask the signal that is being detected. This is a particular concern in the nuclear environments and requires careful design with band-pass filters and signal processing.

Radiation effects on space-based sensors are viewed differently depending on perspective. Figure 25 shows three different ways of looking at radiation effects from the perspective of the radiation effects analysts, the sensor designer and the program manager. However, radiation effects must be given consideration in the design because there is no way to avoid the natural radiation environments of space. Much of the knowledge and data base concerning radiation effects mechanisms and hardening approaches in microelectronics is applicable to sensors. There are some unique considerations for sensors, as discussed above, but sensor hardening has been going on for decades and much progress has been made.

5.1 TRANSIENT EFFECTS IN DETECTORS

Optical detectors are such sensitive detectors of ionization that each ionizing particle which interacts with the detector deposits sufficient charge to be measured. The effect is analogous to single event upset (SEU) in analog circuits. Electrons, protons and heavy ions interact directly, generating

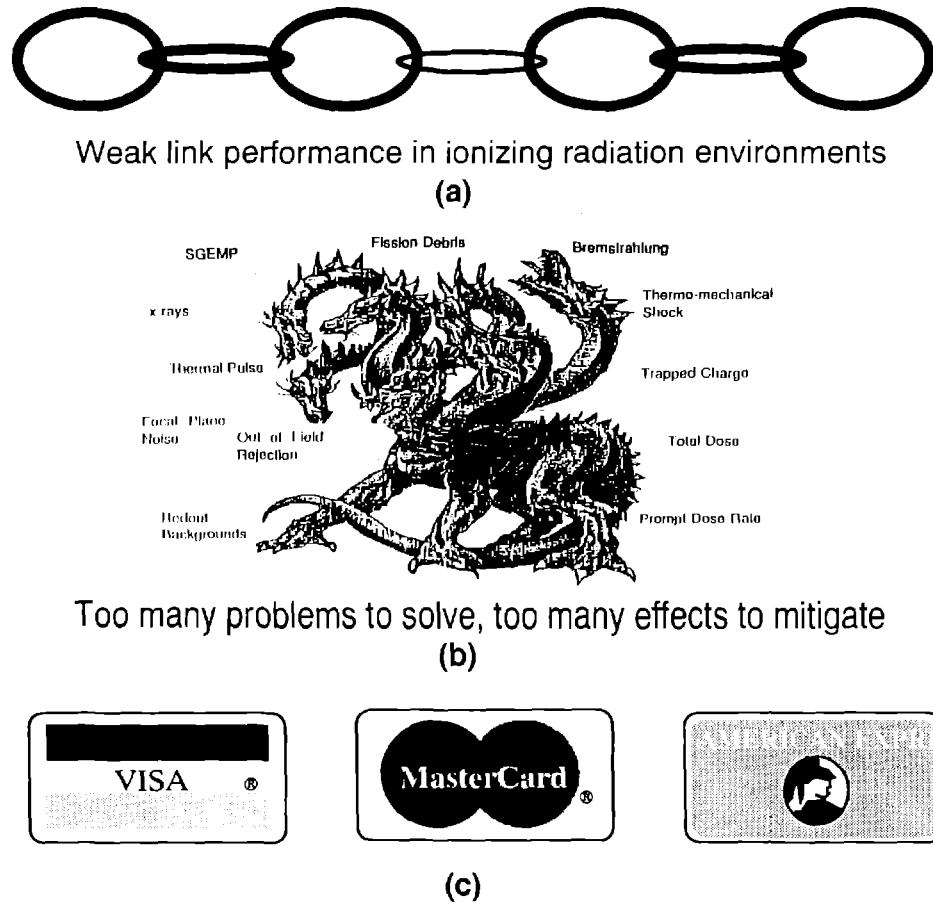


Figure 25. Different perspectives on radiation effects in optical sensors. (a) The radiation effects analyst. (b) The sensor designer. (c) The program manager.

hole-electron pairs along their path through the collection volume of the detector. Gamma rays interact indirectly by first generating secondary electrons by Compton scattering, photoelectric effect or pair production.

The event rate for the transients depends on the particle flux and the effective interaction area. The events have a distribution in amplitude that is determined by the pathlength distribution in the detector sensitive volume and the intrinsic bandgap of the detector material. Analytical models have been developed and verified for predicting noise pulse event rates and amplitudes [29].

The result of the transients is increased noise and decreased signal-to-noise ratio. False events can contaminate data that goes to the sensor signal processor. Figure 26 illustrates the noise introduced by ionizing radiation in a typical detector signal channel. In Figure 26-a the detector current is shown without any optical signal or ionization flux. Figure 26-b shows the output with an optical signal. In Figures 26-c and 26-d we see the same outputs in the presence of ionizing particles — these could be either secondary electrons from gamma interactions in and around the detector, primary electrons from the trapped electron belts, protons from the trapped proton belts, particles from solar flares, or primary cosmic rays. Each particle interaction is a single event with a fast rise time and an exponential decay.

Transient ionizing effects in CCDs, due to protons or electrons, appear as unwanted signals in the video data. The major problem is usually protons since electrons are easily shielded. When

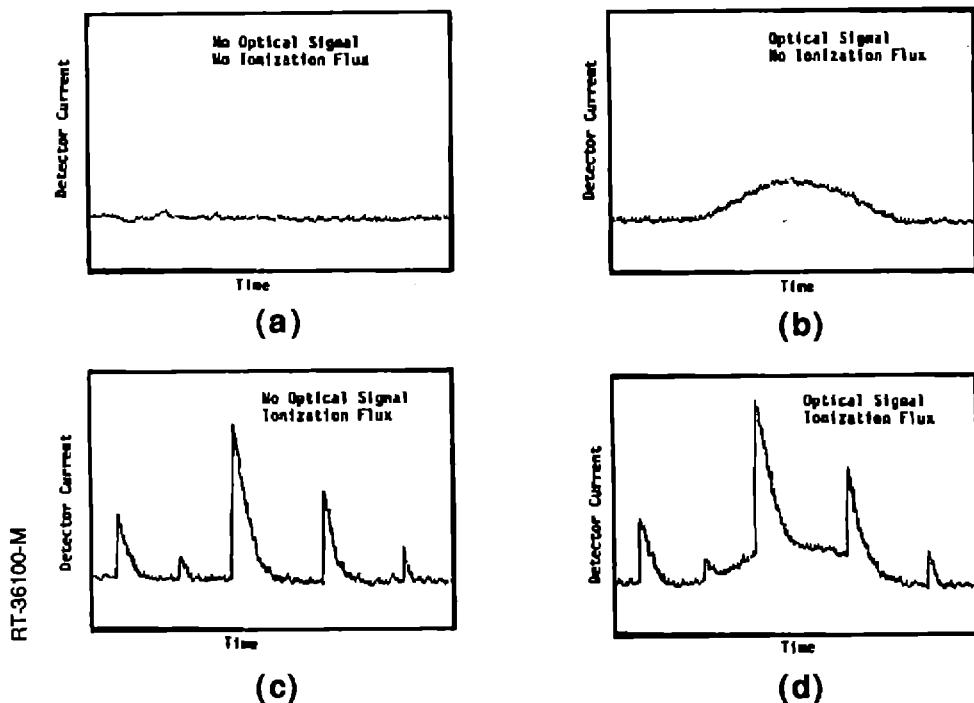


Figure 26. Ionization-induced pulses in an infrared detector.

the proton-induced events compete with the CCD signals, detectivity is degraded due to an increase in the false alarm rate. Mitigation approaches might include: reducing the system sensitivity by turning it off during periods of high proton flux such as traversal of the South Atlantic Anomaly or during solar flares, applying signal processing techniques to reject transient events, or designing the CCD structure to be less susceptible to ionizing radiation. One signal processing technique which works at relatively low proton flux is to differentiate the targets, which persist from frame to frame from the proton events which occupy only a single frame. At higher fluxes, more sophisticated processing is required based on differences between the spatial and amplitude characteristics of the signal and the proton events. In principle, the inherent sensitivity of the CCD to ionizing radiation can be reduced by reducing the diffusion and depletion depth, but the penalty is reduction of detector responsivity at longer wavelengths [30].

The general approach for hardening detectors is to make them thin and as small in surface area as practical. Thinning reduces the average noise pulse amplitude. Reducing the area reduces the event rate. As discussed earlier, the IBC mode for Si:As provides a hardening method by making the active charge collection volume thinner. Signal processing schemes are also used to mitigate noise effects in detectors.

We will now compare the relative performance of the major IR detector technologies regarding ionizing noise. Figure 27 shows the relative performance (in terms of D^*) of the principal MWIR detector technologies as a function of relative gamma flux. A similar relation holds if the y-axis is protons or electrons. The scales are logarithmic so each tic mark represents an order of magnitude. In terms of hardness to transients, we see that Superlattice detectors are the hardest, however the performance without radiation is not as good as some other technologies. PbSe is also fairly hard to transients and has reasonably good clear environment performance; however this technology is not amenable to large arrays and is not a major contender for most modern applications. PtSi is also quite hard to transient noise effects, but it lacks good clear environment performance compared to its competition due to a relatively low quantum efficiency. The best overall performers for most applications in the MWIR are PV HgCdTe and InSb. For operation in a high flux of ionizing particles, other technologies can be considered and signal processing can mitigate some of the noise effects.

Figure 28 shows the relative performance for principal LWIR technologies. Again, we see that PV HgCdTe has the best performance in the clear environment but degrades rapidly with ionizing particle flux. PC HgCdTe is somewhat harder to transients but has lower clear environment performance. IBC Si also has better gamma noise performance than PV HgCdTe, but is not a real contender against the competition in the LWIR regime because of the requirements for cooling to ~ 10 K for Si:As. Superlattice detectors are quite hard to gamma noise but lack sufficient optical sensitivity in the present state of technology to be given serious consideration.

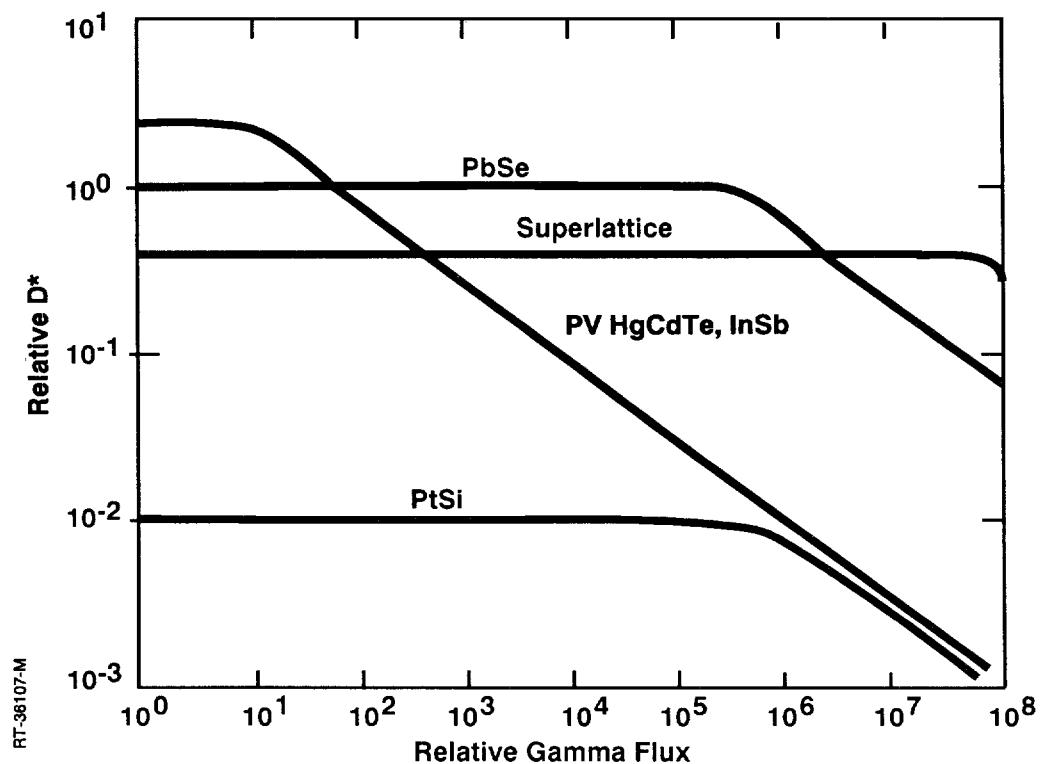


Figure 27. Relative performance of principal MWIR detector technologies in ionizing pulse environment.

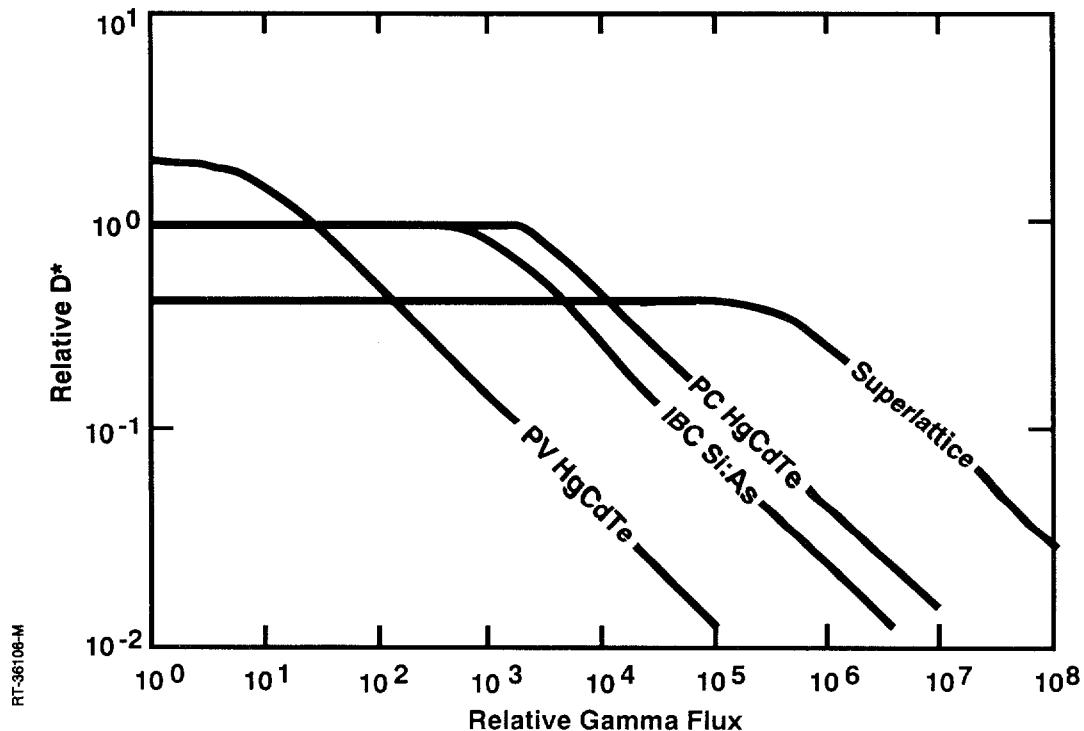


Figure 28. Relative performance of principal LWIR detector technologies in ionizing pulse environment.

The relative performance of the VLWIR technologies is shown in Figure 29. Here the dominant technology is IBC Si:As simply because there is no competing technology that has been developed. PC HgCdTe theoretically has a higher clear environment performance, but this technology has not been fully developed.

5.2 TOTAL DOSE EFFECTS IN CRYOGENIC MOS ELECTRONICS

Total dose mechanisms in MOSFET gate dielectrics have been studied for many years and are fairly well understood. A good tutorial discussion of the subject is given in Reference 31. The mechanisms are illustrated in Figure 30 for room-temperature electronics. Electron-hole pairs are generated by ionizing radiation by intrinsic excitation across the Si bandgap. The electrons readily escape the dielectric; the holes are transported by hopping between localized states in the oxide bulk. For positive bias, the holes are transported to the Si/oxide interface where they can become trapped in deep traps. The difference between a “hardened” and “unhardened” gate oxide (for room-temperature) is the concentration of the deep hole traps near the interface. The occupation of the hole traps can also lead to interface traps within the Si bandgap, resulting in radiation-induced interface states. The effect of the net charge buildup and annealing in the oxide is a time dependent threshold voltage.

At cryogenic temperatures ($\sim <120K$) ionization generated holes in oxides have much reduced hopping probability and are effectively immobile, while electrons still have high mobility and readily escape. The result is a uniform distribution of trapped holes throughout the oxide after total dose irradiation with considerably higher density than in the room temperature case. There is a slow annealing of the trapped holes by stochastic hopping and by tunneling at the oxide-silicon interface, but this process is very slow except for thin ($<20\text{ nm}$) oxides. One bit of good news resulting from the fact that holes are immobile is that interface state formation following dose is eliminated until the device is warmed [32].

The result of this trapped positive charge is a change in surface potential for the silicon underneath, which leads to threshold voltage shift, leakage current and loss of isolation on the integrated circuit. The standard approaches for hardening the oxides by decreasing the density of deep hole trap levels do not work for the cryogenic temperature because the holes are effectively trapped in shallow traps that are part of the nature of oxides [31].

There are hardening approaches for cryogenic oxides, however. The most effective approach to harden the gate oxide is to make the oxide thin [33]. Thin oxides are inherently total-dose-hard for two reasons. First because there is less oxide to trap charge there is less charge trapped. Second, there is always tunnel annealing at the interfaces and if the oxides are thin, the two interfaces are sufficiently close that the tunneling fronts from the two sides rapidly converge to anneal out the trapped charge [34]. The convergence of tunneling fronts is illustrated in Figure 31. If the charge anneals out on time scales that are fast relative to times of interest, the device is total-dose hard.

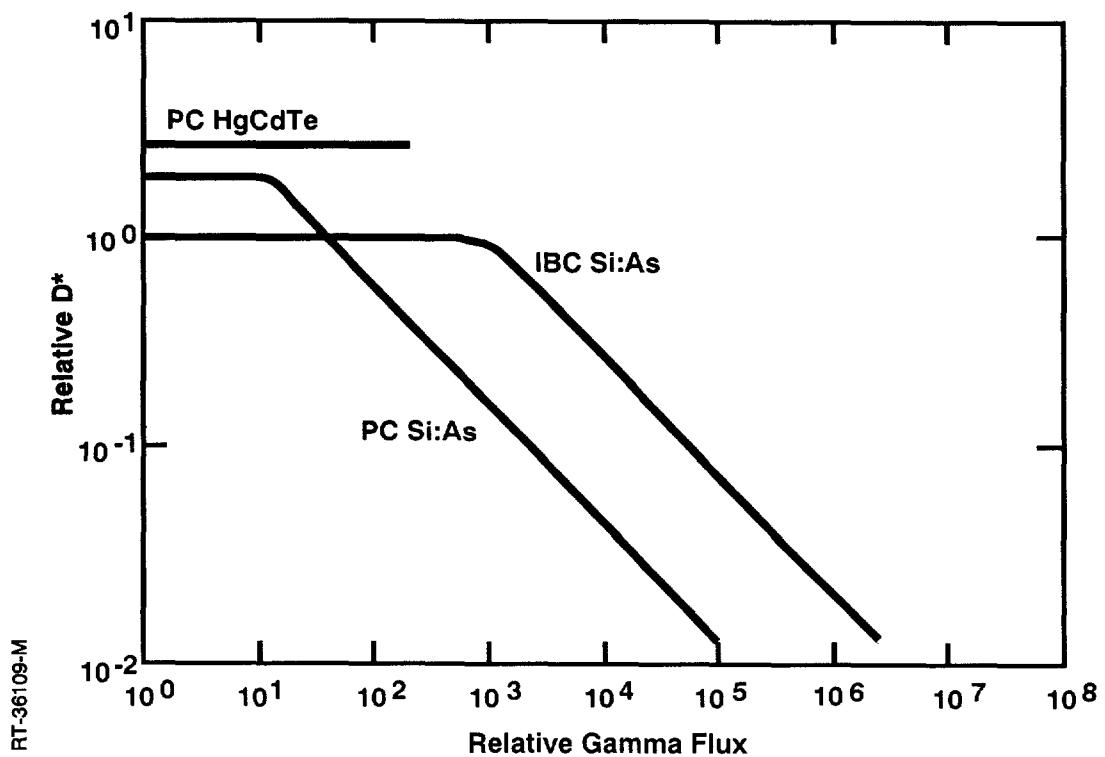


Figure 29. Relative performance of principal VLWIR detector technologies in ionizing pulse environment.

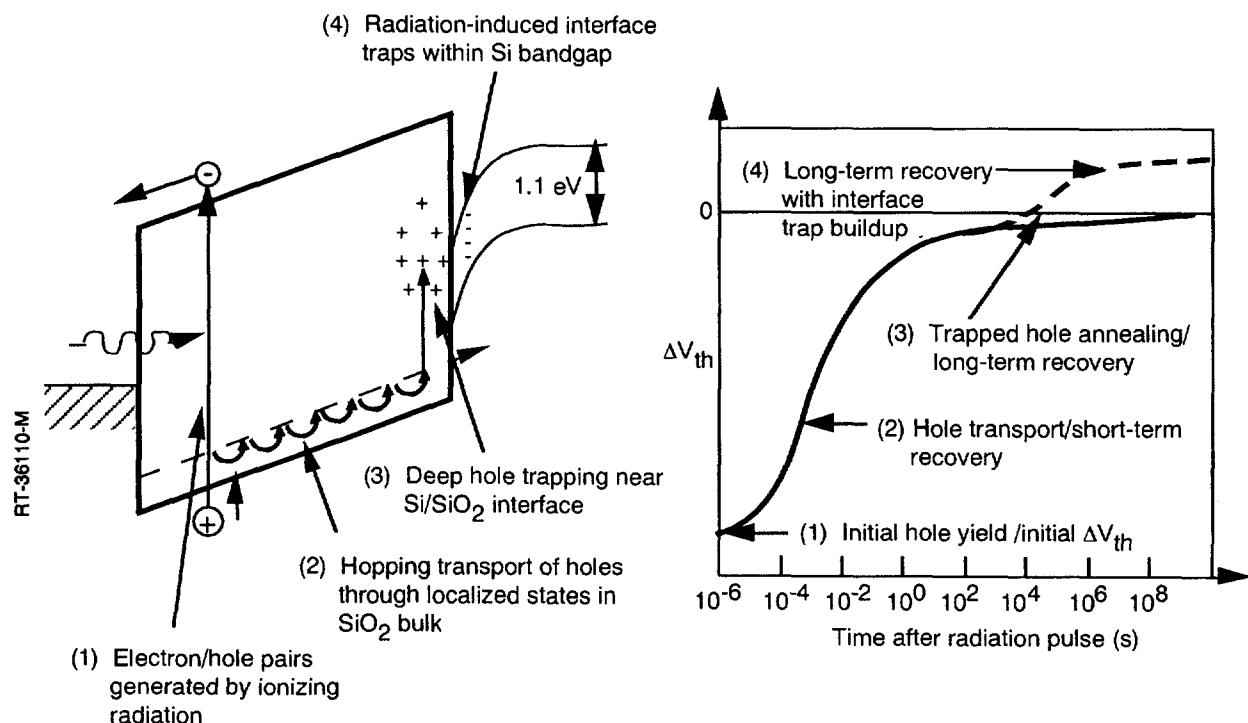
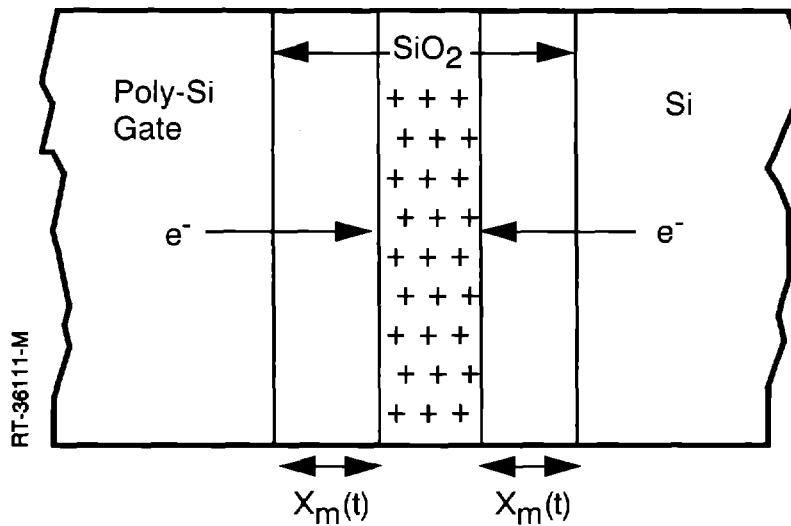


Figure 30. Total-dose mechanisms in MOSFET gate dielectrics.



- Model schematic of trapped hole removal in thin gate-oxide MOS structures by electron tunneling from both Si substrate and polysilicon gate

Figure 31. Thin oxides benefit from tunnel annealing.

Figure 32 illustrates the two principal parameters that affect total-dose-induced flatband shift in cryogenic MOS devices: the electric field in the oxide during irradiation and the thickness of the oxide [34]. Because there are competing mechanisms of electron-hole recombination at low electric field and enhanced hole transport (by stochastic hopping) at high electric field, there is a maximum in the total-dose-induced flatband shift that occurs at oxide fields of ~2 MV/cm. It turns out that since many MOSFETs in a FPA readout circuit are operated in the subthreshold mode, the electric field in the oxide is often near the worst case value.

Thinning is the best approach to harden oxides, for the reasons discussed above. However a practical issue with thin oxides is dielectric integrity. Some improvement in dielectric integrity is seen with oxynitrides [35], particularly reoxidized nitrided oxides (variously referred to in the literature as ONO, RONO, RNO). Additional benefits of ONO gate dielectrics include better immunity to hot electron trapping [36] and somewhat improved total dose hardness [37].

The problem with field inversion due to trapped positive charge in the field oxide is illustrated in Figure 33. For the nMOSFET, the positive charge tends to invert the underlying p-type Si, leading to crosstalk between potential nodes on the integrated circuit and to leakage current from source to drain on individual MOSFETs. This is not a problem for pMOSFETs since the underlying Si is n-type and is accumulated by the positive charge buildup.

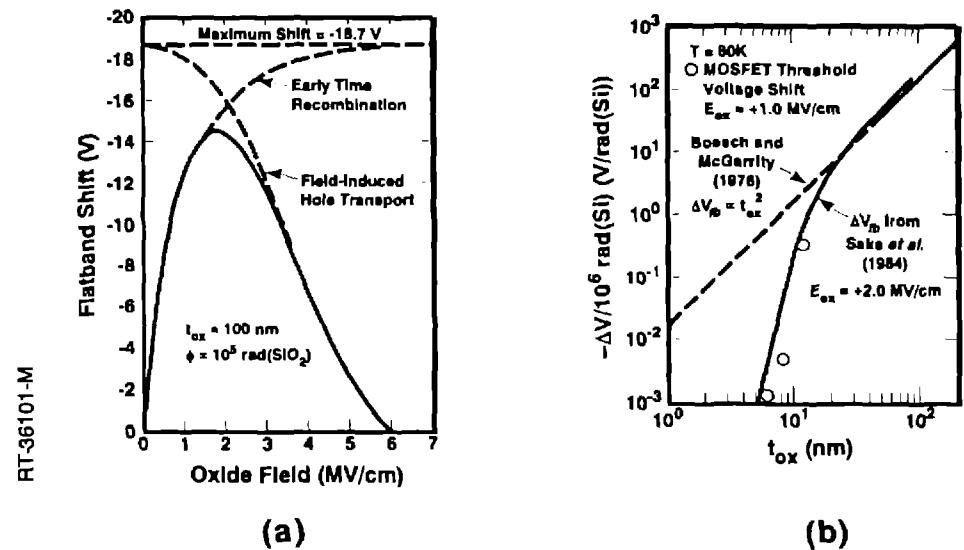


Figure 32. Total-dose-induced flatband voltage shift in MOS devices. (a) Competing mechanisms of recombination at low fields and hole transport at high fields give a worst-case field at 1-2 MV/cm. (b) Thinning benefits from less charge trapped and faster tunnel annealing.

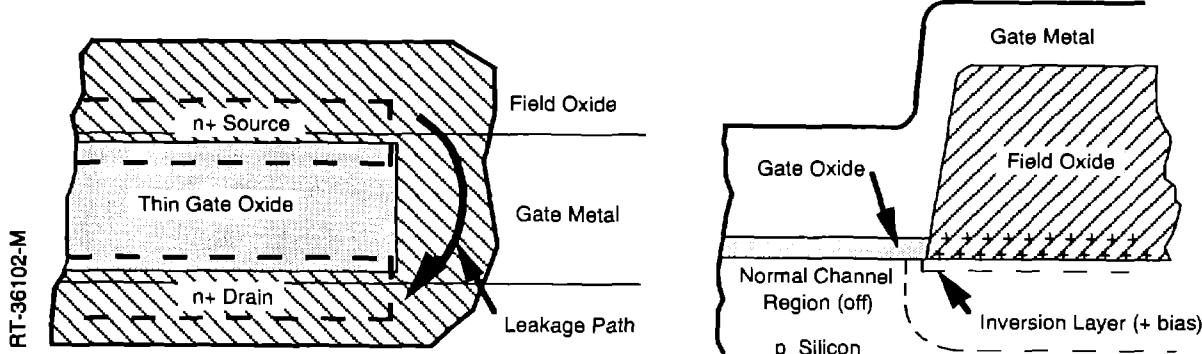


Figure 33. Positive charge in field oxide causes leakage current in nMOSFET.

The hardening approach for field leakage is to place a heavily doped p+ guard band around the nMOSFET which does not invert with the charge buildup above it in the field oxide. An example of one possible implementation of a guard ring is shown in Figure 34. This approach is very effective and is commonly used for hardened cryogenic electronics. The down-side is that there is a density penalty.

5.3 TOTAL-DOSE EFFECTS IN DETECTORS

Total dose also leads to charge buildup in passivation layers over detectors. In addition, interface states can be generated and / or populated by ionizing radiation. For PV detectors, increased

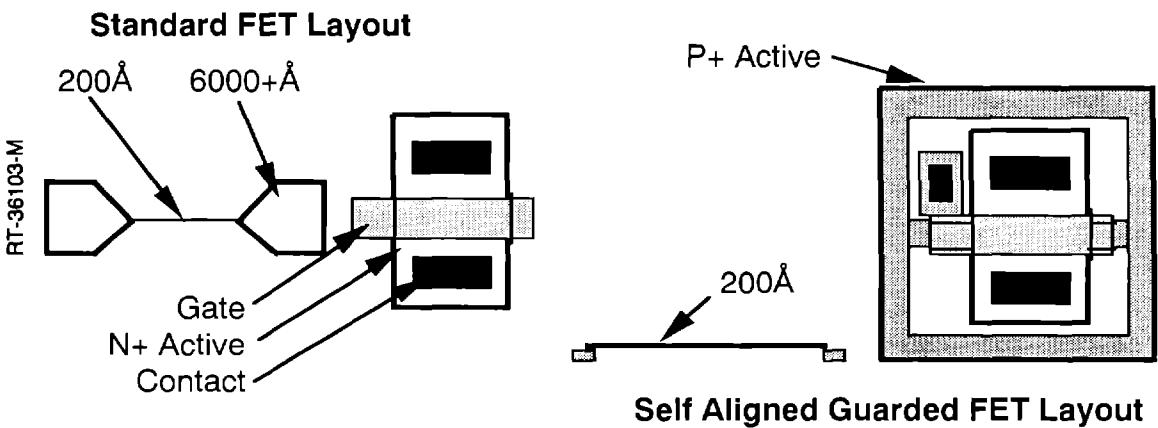


Figure 34. P+ guard-band is effective hardening for field leakage.

surface tunneling lowers RoA, and therefore the D^* of the detector . Changes in surface recombination cause change in effective optical area and can lead to cross-talk or reduced optical signal. Surface inversion or accumulation can cause cross-talk [38,39].

For CCDs, charge buildup in the surface oxide can cause interface traps which interfere with charge transfer efficiency (CTE). For this reason, CCDs are often fabricated to operate in a buried channel mode so that the signal charge is not affected by the radiation-induced surface traps.

Figure 35 shows the degradation in RoA versus relative dose for LWIR HgCdTe PV detectors operating in tunneling limited, high performance mode, i.e., 40 K, low background. Total-dose hardening approaches for detectors involve choice of passivating insulator and treatment of the interface between the semiconductor and the passivation layer [40]. Relative data is shown in the figure for different passivation technologies; major hardening advances have been made.

Figure 36 illustrates the total-dose-induced crosstalk mechanisms in detectors. For IBC Si:As, the n-type Si beneath the oxide passivation layer can be accumulated for total-dose-induced positive charge buildup, leading to a lower resistivity between the active n+ detector regions (Figure 36-a). Because the detectors are very high impedance ($>1E10$ ohms) only a slight decrease in resistivity can lead to crosstalk. Figure 36-b shows the crosstalk mechanism in PV HgCdTe. In this case it is surface inversion that causes the problem. The p-type semiconductor is inverted by positive charge in the ZnS passivation layer.

Another dose effect that is restricted to extrinsic Si PC detectors is an anomalous increase in responsivity after very low doses, on the order of one rad [41,42]. There is no detectivity gain because the noise is also increased by the same amount. The effect is termed the gamma response anomaly and the mechanism is related to an increased carrier lifetime because of total-dose-induced changes to the charge state of donor impurities. Such changes compromise calibrations and make precision radiometric measurements difficult. The IBC mode for extrinsic Si is free of the gamma response anomaly and this was one of the factors that drove extrinsic detectors to IBC technology.

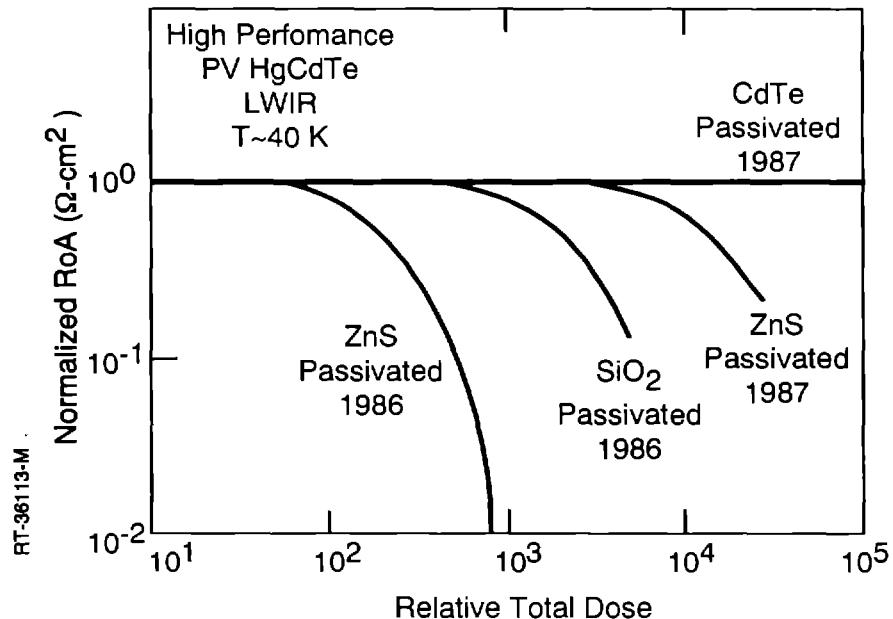


Figure 35. Trapped charge degrades zero-bias resistance in PV HgCdTe detectors. Major total-dose hardening advances have been made by use of improved passivation technology.

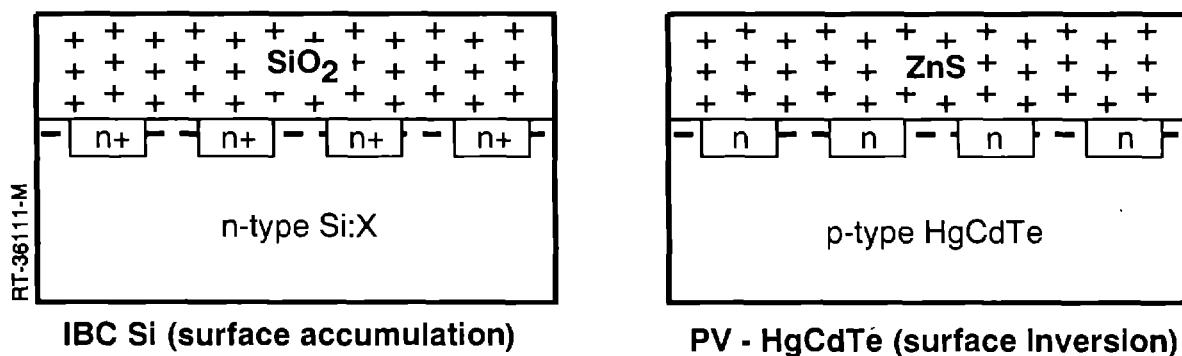


Figure 36. Total-dose-induced inversion or accumulation can produce cross-talk in detector arrays. (a) Surface accumulation in IBC Si causes crosstalk. (b) Surface inversion in PV n on p HgCdTe causes crosstalk.

5.4 DISPLACEMENT DAMAGE EFFECTS IN DETECTORS

Displacement damage is an important consideration for sensors that are operated within the trapped proton belts; even for sensors operated above the belts, such as geosynchronous orbit, displacement damage from solar flares could be an issue. For low earth orbits that are below the belts, traversal of the South Atlantic Anomaly (SAA) can lead to cumulative proton fluences that could be of concern. For detectors that have minimal shielding against the trapped electrons, this environment can also be of concern for displacement damage for sufficiently high total dose.

The effect is displacement of atoms from lattice sites and it is a permanent damage problem, like total-dose, rather than a transient. The displaced atoms often can form stable complexes with the impurities or with other displaced atoms or vacancies. There is an initial rearrangement of the complexes (short term anneal) that can occur but this effect is less at cryogenic temperature.

The impact of the defects on semiconductor materials is lifetime decrease, mobility decrease, and carrier removal. The device impacts are varied but the decrease in minority carrier lifetime is usually the most readily observed effect. Higher quality (less defects in the crystal structure) devices are affected the most.

Figure 37 illustrates the proton displacement damage mechanism in a Si-CCD. The potential well for a buried channel CCD is shown in the top trace. CCDs commonly use buried channels for operation in radiation environments in order to avoid interface traps at the oxide - silicon interface. However, protons or other energetic heavy ions that pass through the buried channel region can displace atoms from their lattice sites, leading to a vacancy (the site in the lattice from where the displaced atom was removed) and an interstitial atom. Permanent traps can be formed that result in new energy levels in the bandgap that interfere with the signal charge. A common displacement-induced trap in Si-CCDs is a stable complex between a phosphorous impurity atom and a vacancy (the P-V complex) with a trap energy of 0.4 eV. These traps can rob charge from the signal and release it later, with the result that the CTE is degraded. They also serve as sources of dark current.

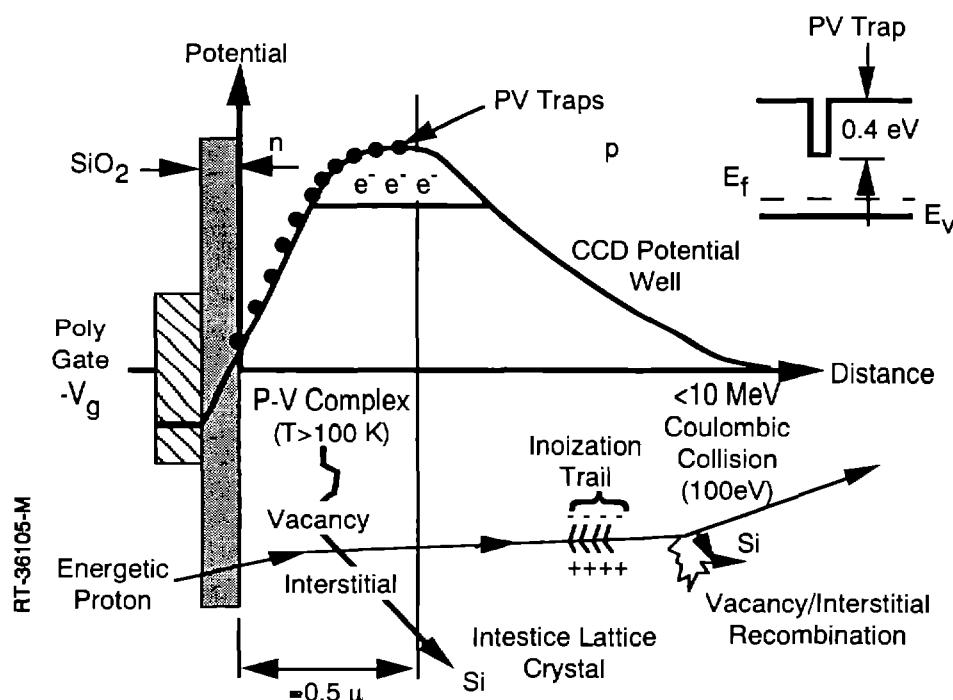


Figure 37. Proton displacement effects in Si-CCD.

CCD operation requires physical transport of photoelectrons across macroscopic distances within the bulk of the semiconductor to an output amplifier. For a large format (e.g., 1024x1024) sensor to be used in a low light level application, the CTE must be very high to transport the low photosignal to the output amplifier with fidelity. If the CTE is as high as 0.999^c, then after 1000 transfers the net transfer efficiency is 0.9999¹⁰⁰⁰ or only 0.90, an unacceptable low value in many cases. For a low light level of 100 photoelectrons, preservation of contrast implies that only a few electrons can be lost in the 1000 transfers. Since a single silicon defect can and usually will trap a signal electron, the silicon must be nearly defect free. Only a few proton hits to any of the 1024 CCD cells along the transfer column or the cells along the output serial register could result in unacceptable degradation.

At present, approaches for proton damage mitigation in CCDs are somewhat limited. Two approaches that have been used are cooling and flood-illumination; these are illustrated in Figure 38. When the CCD is cooled the displacement-induced traps can be frozen-out, so when filled with electrons they will remain filled for very long times and not influence CTE performance. For example, P-V traps will freeze-out at -90 C. Uniform flood-illumination with a low level light signal keeps the radiation-induced traps filled before an exposure is taken; the signal offset is removed in signal processing. This technique is used by Galileo and CRAF/Cassini [43]. Cooling also has the

RT-36106-M

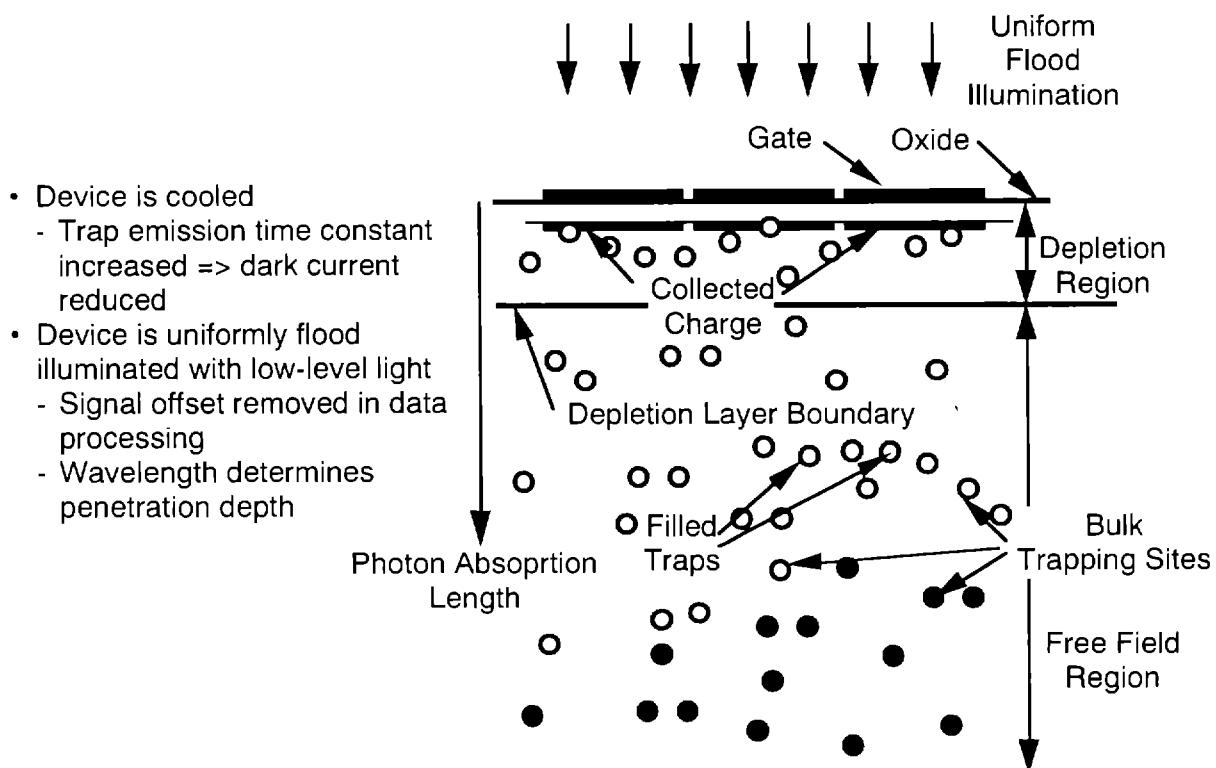


Figure 38. Proton damage mitigation in CCDs.

advantage of freezing out displacement-induced dark spikes, pixels that generate abnormally high levels of dark current because of radiation induced damage [44].

For operation in radiation environments with a large amount of displacement damage, such as in the heart of the proton belts, a technology that does not require charge transfer is required.

The effect of displacement damage on other detector technologies, such as PV HgCdTe and Si:As is under investigation. The general rule is that the higher performance detectors are the most sensitive to displacement damage.

6. ADDITIONAL CONSIDERATIONS

This section summarizes additional practical considerations and tradeoffs facing the sensor designer.

6.1 NOISE

Noise is one of the major performance drivers for any sensor system design. The noise level directly determines the sensitivity for detection, the quality of imaging and the fidelity of radiometric data. Noise sources have two general forms, temporal and spatial.

Temporal noise is usually described in terms of noise equivalent irradiance (NEI) in photons/cm²-s. Temporal noise sources are effectively uncorrelated in both time and space. Several temporal noise source affect focal plane arrays such as fluctuations in the background, thermal fluctuations in the target and noise in the FPA detector and readout electronics. Examples of the latter include detector thermal noise, flicker (1/f) noise, Johnson (resistive thermal) noise, shot noise, readout noise and clock feedthrough. For operation in radiation environment, ionization noise must be included.

FPA used for imaging typically require calibration to mitigate variations in responsivity and dc offset among detector channels. Periodic recalibration is sometimes required to incrementally update some drifting characteristic of the response. Pattern noise arises because the calibration and equalization models cannot accurately model actual response or the drift is beyond the capability of the calibration scheme. In general, pattern noise in scanning arrays appears as streaks or bands, depending on how many detectors cluster in the mechanism. In a staring array, pattern noise will appear as artificial “hot” and “cold” areas.

Noise for a sensor is often quantified in terms of noise electrons associated with the detection and signal read process. Typical noise levels for IR FPA imaging applications range from a few thousand down to less than one hundred electrons. The lowest noise processes are usually associated with CCDs with levels as low as 1.5 electron reported [43].

A good discussion of general noise processes in hybrid infrared focal plane arrays and an example of noise modeling is given in References 13 and 14.

6.2 READOUT OF A LARGE ARRAY

As array sizes increase, the issue of a larger quantity of data and the speed with which it can be retrieved from the focal plane becomes more critical. While this issue has not been a primary concern for present-generation IR focal plane technologies, it will become increasingly important for future IR technologies, and is a necessary consideration for present CCD visible imagers.

For example, the largest FPAs at present are 4096x4096 CCDs. Assuming 25 kilo pixels/s pixel rate, typical of slow scan astronomical cameras, the 4096x4096 sensor requires 11.3 minutes for complete frame readout [3].

There are three basic approaches to readout of an area array: Charge Coupled Device (CCD), Charge Injection Device (CID) and Amplifier per Detector. In the CCD, charge is moved from the detection site to the edge of the array by sequentially manipulating the surface potential in closely spaced capacitors. The advantages of this approach are simplicity and the ability to quickly read out a large array. The disadvantages of the CCD readout is that signal charge is subject to contamination from ionization transients or to charge loss from pre-existing defects or displacement-induced defects, i.e., degradation of charge transfer efficiency.

A CID collects charge at the detector site and then the pixels are read out by x-y address busses, similar to a random access memory (RAM) chip. Because of parasitic capacitance on the buss lines, this readout approach is more subject to noise and has lower sensitivity.

An alternate approach is to put a dedicated preamplifier readout circuit, such as a Source Follower per Detector (SFD) or more complex circuit such as a CTIA, within each unit cell. Then the signals that go to the readout multiplexer buss lines are less susceptible to noise contamination because they are larger charges. In the case of Si photodetectors, such as photodiodes or PIN diodes or PtSi Schottky diodes, the readout circuit can be integrated within the same chip as the detectors. For the case of hybrid detectors, the readout array is connected to the detector array through bump bonding. The dedicated amplifier per unit cell approach is more appropriate for radiation environments since there is no long series of charge transfers that could be subject to compromise by the radiation environments.

6.3 RELIABILITY OF HYBRIDS

Hybridization brings up a reliability issue. The indium bump bonds are required to maintain electrical and mechanical integrity over time and through multiple thermal cycles from room temperature to the cryogenic temperatures. Some thermal cycling is inevitable during the assembly and testing of the FPA and sensor system. Planarity in the two chips being joined is critical. Also, differences in thermal expansion / contraction are a concern. The Coefficient of Thermal Expansion

(CTE) mismatch between Si (the readout chip) and other popular detector materials such as HgCdTe and InSb is a definite reliability issue and becomes increasingly important as array size is increased. Various approaches including mechanical shims, thinning and general “engineering” of the mechanical interface between the two dissimilar materials are under development. Largest hybrid array sizes demonstrated to date are 512x512 and 640x480.

6.4 CRYO-COOLING

Application of cryo-coolers to space requires concern for reliability, power, weight and cooling load capacity. Passive radiative coolers are generally limited to applications above 100 K. Thermoelectric coolers generally require too much power for space application. Stored cryogen coolers are the most reliable, but have a limited lifetime — on the order of one year is about as long as is practical.

Long-term space applications are limited to mechanical coolers. Practical operation is possible to temperatures as low as ~60 K for present technology. However, reliability is still a concern. Technology for mechanical cryo-coolers for operation under 60 K is still under development.

7. CONCLUSIONS

In conclusion, we summarize the following issues as the principal ones concerning microelectronics that the sensor designer must consider and make tradeoff decisions:

Choice of Detector Technology

The choice of detector material depends on the needed sensitivity in the wavelength bands of interest. Operating temperature must be carefully considered because this is a major factor that drives system power, weight and cost. Depending on the application, uniformity may be more or less of a concern. Uniformity can be specified for the detector array or can be corrected, to a degree, by signal processing. Specifying a high degree of uniformity at the detector array level may impose a producibility penalty. Producibility of the detector arrays is often a significant consideration, since most of the IR technologies are fabricated in small volume, and producibility is an issue.

Focal Plane Array Format

The choices are scanning or staring, conventional hybrid or z-plane technology, and the array size; the selections are driven by the mission of the sensor. A choice might be made between a single large FPA or several smaller buttable modules combined to effect a larger FPA. Considerations may involve tradeoffs with the amount of on-focal-plane signal processing that can be placed in the unit cell. Larger arrays require smaller unit cells because there is a limit to practical chip sizes for both the detector array and the readout integrated circuit. The z-plane technology provides more real-estate for the signal processing but may have producibility and reliability tradeoffs.

Design of Readout Circuit

The tradeoffs involve noise requirements, dynamic range and available real estate to implement the circuits in a unit cell.

Data Processing

The driving issue is the data rate. Moving more of the signal processing functions onto the focal plane, particularly analog to digital conversion, will produce major benefits.

Cryo-Coolers

Active mechanical coolers are usually required. The major issues are reliability and the power burden on the system.

Radiation Effects

Radiation effects must be considered in the design of the sensor. There is no way to avoid the natural environment; it is not practical to shield it all out.

Space-based sensors are an example of a highly sophisticated use of microelectronics for performing imaging, surveillance, detection, astronomy and other scientific measurements from a unique vantage point. Most all of the practical considerations for use of microelectronics in space for non-sensor applications, as well as a set of unique considerations specifically related to optical sensor components, apply to successful sensor system designs. Microelectronic technologists have provided the sensor designers with wide choice of component technology and an understanding of degradation mechanisms to allow for informed design tradeoffs.

This brief overview has only skimmed the surface of optical sensor technology. There is a mountain of detail held in the collective knowledge base of the sensor community.

ACKNOWLEDGEMENTS

I wish to thank my colleagues at S-Cubed, including Joe Boisvert, Jim Duffey, Arne Kalma, and Jason Wilkenfeld for their helpful suggestions and review of this material. Special thanks go to Paul Morse of Loral for providing details on a real sensor design. I also wish to acknowledge the numerous researchers in the sensor technology community with whom I have had the pleasure to work with and learn from over the course of many years.

REFERENCES

1. Proceedings of the Annual Meeting of the IRIS Specialty Group on Infrared Detectors, published annually by The Infrared Information and Analysis Center, Environmental Research Institute of Michigan.
2. Paul Morse, Loral, private communication.
3. J.Janesick, T.Elliott, A.Dingizian, R.Bredthauer, C.Chandler, J.Westphal and J.Gunn, "New Advancements in Charge-Coupled Device Technology - Sub Electron Noise and 4096x4096 Pixel CCDs," Astronomical Society of Pacific Conference Series, 1989.
4. S.Selberherr, "MOS Device Modeling at 77 K," IEEE Trans. Elect. Dev., ED-36, pp. 1464-1474 (Aug. 1989).
5. K.S.Kalonia and F.C.Jain, "An Analytical Two-Dimensional Model for CMOS Devices at Room and Cryogenic Temperatures," Solid State Electronics, Vol. 33, No. 7, pp. 947-951 (July 1990).
6. C-L.Huang and G.S.Gildenblat, "Engineering Model of MOS Transistors for the 60-300 K Temperature Range," Electronics Letters, Vol. 26, No. 6, pp. 365-367 (15 March 1990).
7. D.S.Jeon and D.E.Burk, "MOSFET Electron Inversion Layer Mobilities - A Physically Based Semi-Empirical Model for a Wide Temperature Range," IEEE Trans. Elect. Dev., ED-36, pp. 1456-1463 (Aug. 1989).
8. R.M.Fox and R.C.Jaeger, "MOSFET Behavior and Circuit Considerations for Analog Applications at 77 K," IEEE Trans. Elect. Dev., ED-34, pp. 114-123, (Jan. 1987).
9. R.M.Glidden, S.C.Lizotte, and J.S.Cable, "Optimization of Cryogenic CMOS Processes for Sub 10 K Applications," presentation to SPIE Conference, Orlando FL (1992).
10. M.Song, J.S.Cable, J.C.S.Woo and K.P.MacWilliams, "Optimization of LDD Devices for Cryogenic Operation," IEEE Elect. Dev. Lett., EDL-12, pp.375-378 (July 1991).
11. R.B.Hammond, G.L.Hey-Shipton and G.L.Matthaei, "Designing With Superconductors," IEEE Spectrum, pp. 34-39, April, 1993.
12. James Ritter, Naval Research Laboratory, private communication.
13. M.D.Nelson, J.F.Johnson and T.S.Lomheim, "General Noise Processes in Hybrid Infrared Focal Plane Arrays," Optical Engineering, Vol.30, No.11 (Nov. 1991).
14. J.F.Johnson and T.S.Lomheim, "Hybrid Infrared Focal Plane Signal and Noise Modeling," paper presented at SPIE's 1991 Symposium on Optical Applied Science and Engineering, San Diego, July 1991.

15. G.Orias, A.W.Hoffman and M.F.Casselman, "58x62 InSb Focal Plane Array for Infrared Astronomy," in Instrumentation in Astronomy VI, Proc. SPIE 627, pp. 408-417 (1986).
16. A.M.Fowler, R.G.Probst, J.P.Britt, R.R.Joyce, and F.C.Gillett, "Evaluation of Indium Antimonide Focal Plane Array for Ground-Based Infrared Astronomy," Opt. Eng. 26(3), pp. 232-240 (1987).
17. A.W.Hoffman, "Operation and Calibration of Self-Integrating Multiplexed Arrays," in Infrared Astronomy with Arrays, C.G.Wynn-Williams and E.E.Beklin, Eds., Univ. of Hawaii (1987).
18. N.Bluzer and A.S.Jensen, "Current Readout of Infrared Detectors," Opt. Eng. 26(3) (1987).
19. K.Chow, J.P.Rode, D.H.Seib and and J.D.Blackwell, "Hybrid Infrared Focal Plane Arrays," IEEE Trans. Elect. Dev., ED-29(1) (1982).
20. J.T.Longo, D.T.Cheung, A.M.Andrews, C.C.Wang and J.M.Tracy, "Infrared Focal Planes in Intrinsic Semiconductors," IEEE Trans. Elect. Dev., ED-25(2) (1978).
21. A.H.Lockwood and W.J.Parrish, "Predicted Performance of Indium Antimonide Focal Plane Arrays," Opt.Eng. 26(3), pp. 228-231 (1987).
22. J.Stobie, A.Hairston, N.Hartle, P.Zimmerman, R.Capps and D.Hall, "Short Wavelength HgCdTe Focal Plane Technology for Low Background Astronomy Applications," IRIS Detector Specialty Group (Aug 1987).
23. L.K.Kozlowski, V.H.Johnson, J.K.Chen, D.D.Edwall, A.M.Blume, I.S.Gerghis, K.Vural, W.J.Parrish and J.D.Blackwell, "Low Noise SWIR HgCdTe FPA with Advanced Signal Processing in a CMOS Readout," IRIS Detectors Specialty Group (Aug. 1988).
24. K.L.Pettijohn, G.Orias, M.J.Hewitt, D.A.Newman, P.J.Love, I.Kasai, B.R.Pagel, M.C.Kelley and C.J.Fischer, "Producibility and Performance of MWIR HgCdTe SCAs Utilizing CTIA Readout Circuits," IRIS Detector Specialty Group (Aug. 1989).
25. A.R.Johnston and E.W.Taylor, "A Survey of the LDEF Fiber Optic Experiments," JPL Report JPL D-10069, November 1992.
26. D.C.Meshel and G.K.Lum, "Radiation Hardness Assurance Characterization Testing of InGaAsP Fiber Optic Transmitter and Receiver Modules," to be presented at 1993 IEEE Nuclear and Space Radiation Effects Conference, Snowbird, Utah.
27. The Infrared Handbook, edited by W.I Wolfe and G.J.Zeiss, The Infrared Information and Analysis Center, Environmental Research Institute of Michigan, Revised Edition, 1985.
28. J.R.McCoy, "Industrial-Grade Cryogenics Test the IR Market," Photonics Spectra, pp 89-90, Jan. 1993.

29. J.C.Pickel and M.D.Petroff, "Nuclear Radiation-Induced Noise in Infrared Detectors," IEEE Trans. Nucl. Sci., NS-22, p. 2456 (Dec. 1975).
30. T.S.Lomheim, R.M.Shima, J.R.Angione, W.F.Woodward, D.J.Asman, R.A.Keller and L.W.Schumann, "Imaging Charge-Coupled Device (CCD) Transient Response to 17 and 50 MeV Proton and Heavy-Ion Irradiation," IEEE Trans. Nucl. Sci., NS-37, pp. 1876-1885 (Dec. 1990).
31. F.B.McLean and T.R.Oldham, "Basic Mechanisms of Radiation Effects in Electronics Materials and Devices," Harry Diamond Laboratory Report HDL-TR-2129 (Sept 1987).
32. N.S.Saks and M.G.Ancona, "Generation of Interface States by Ionizing Radiation at 80 K Measured by Charge Pumping and Subthreshold Slope Techniques," IEEE Trans. Nucl. Sci., NS-34, pp. 1348-1354 (Dec. 1987).
33. N.S.Saks, M.G.Ancona and J.A.Modolo, "Radiation Effects in MOS Capacitors with Very Thin Oxides at 80 K," IEEE Trans. Nucl. Sci., NS-31, pp. 1249-1255 (Dec. 1984).
34. J.M.Benedetto, H.E.Boesch, Jr., F.B.McLean and J.P.Mize, "Hole Removal in Thin-Gate MOSFETs by Tunneling," IEEE Trans. Nucl. Sci., NS-32, pp. 3916-3920 (Dec. 1985).
35. T.Ito, T.Nakamura and H.Ishikawa, "Advantages of Thermal Nitride and Nitroxide Gate Films in VLSI Processes," IEEE Trans. Elect. Dev., ED-29, p. 498 (1982).
36. T.Hori and H.Iwasaki, "Improved Hot Carrier Immunity in Submicrometer MOSFETs With Reoxidized Nitride Oxides Prepared by Rapid Thermal Processing," IEEE Elect. Dev. Lett., EDL-10, p. 195 (1989).
37. G.J.Dunn and P.W.Wyatt, "Reoxidized Nitrided Oxides for Radiation Hardened MOS Devices," IEEE Trans. Nucl. Sci., NS-26, p. 2161 (1989).
38. G.M.Williams, A.H.B.Vanderwyck, E.R.Blazejewski, R.P.Ginn, C.C.Li, and S.J.Nelson, "Gamma Radiation Response of MWIR and LWIR HgCdTe Photodiodes," IEEE Trans. Nucl. Sci., NS-34, pp. 1592-1596 (Dec. 1987).
39. D.W.Domkowski, D.G.Feller, L.R.Johnson, C.I.Westmark, C.B.Norris, C.T.Fuller, and J.Bajaj, "Effects of 6 MeV Electron Irradiation on the Electrical Characteristics of LPE HgCdTe/CdTe Mesa Photodiodes," IEEE Trans. Nucl. Sci., NS-33, pp. 1471-1473 (Dec. 1986).
40. J.R.Waterman and R.A.Schiebel, "Ionizing Radiation effects in N-Channel HgCdTe MISFETs with Anodic Sulfide Passivation," IEEE Trans. Nucl. Sci., NS-34, pp. 1597-1601 (Dec. 1987).
41. M.D.Petroff, J.C.Pickel and M.D.Curry, "Low-Level Radiation Effects in Extrinsic Infrared Detectors," IEEE Trans. Nucl. Sci., NS-26, p. 4848 (Dec. 1979).

42. L.Varnell and D.E.Langford, "Radiation Effects in IRAS Extrinsic Infrared Detectors," IEEE Trans. Nucl. Sci., NS-29, p.p. 1551-1554 (Dec. 1982).
43. J.Janesick and T.Elliott, 'History and Advancement of Large Area Array Scientific CCD Imagers," Astronomical Society of Pacific Conference Series, 1991.
44. C.J.Dale, P.W.Marshall and E.A.Burke, "Particle-Induced Spatial Dark Current Fluctuations in Focal Plane Arrays," IEEE Trans. Nucl. Sci., NS-37, p.p. 1784-1791 (Dec. 1990).