Landmarks in Terrestrial Single-Event Effects

2013 NSREC Short Course

Robert C. Baumann

IEEE and TI Fellow, Technology Office
Medical/High Reliability Group
High Performance Analog
Texas Instruments, Dallas, Texas

I. Single Events Primer

1. Motivation: Real-life Failure
2. Ions, Junctions, and Types of Single Event Effects (SEEs)
3. Q_{crit}, Q_{coll}, DRAM, and SRAM Single Event Upset (SEU)
4. The Terrestrial Radiation Mechanisms
    a. High-Energy Neutrons
    b. Alpha Particles
    c. Low-Energy Neutrons and Boron
    d. Cosmic Ray Muons?

II. Historical Retrospective

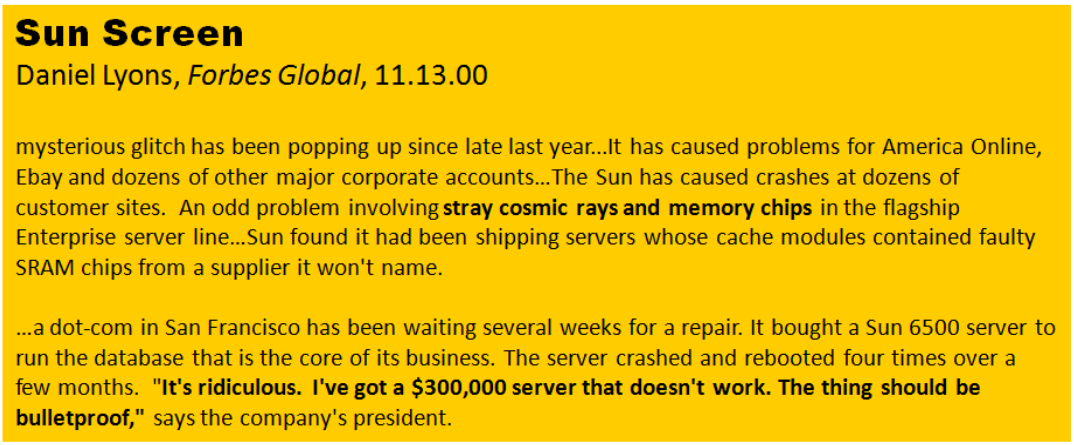
1. Thomas Kuhn: A "New" Theory of Scientific Revolutions
2. From X-rays to Natural Radioactivity and Cosmic Rays
3. A Brief History of Electronics – from Vacuum to Solid State
4. Two Early Views on Scaling: The Pessimist vs. Optimist
5. The First Hints of Fragility - Upsets in Space
6. Upsets from Alpha Particles
7. Upsets from High-Energy Neutrons
8. Upsets from ^{10}B and Thermal Neutrons
9. Summary

I-1. Motivation: Real-life Failure

In homage to the auspicious 50th anniversary of the NSREC, this short course is focused on giving a historic retrospective of breakthroughs in the understanding of single-event effects (SEEs) in commercial electronics in the terrestrial environment. We will cover the discoveries and growth of knowledge that accompanied the semiconductor revolution as it fulfilled Moore's Law [1] of ever-increasing density, functionality, and cost-reduction as semiconductor technology has been inexorably scaled from micrometer, sub-micron, deep-sub-micron, and, more recently, to nanometer-scale dimensions. In an

effort to make the impact and context of each “landmark” more easily understood, we will spend the first half of the short course reviewing the current understanding of SEEs in commercial digital devices. Even in the relatively benign terrestrial environment (as compared with the more hostile space, medical, nuclear reactor, or weapons environments), radiation effects have a profound impact on the reliability performance of microelectronics. Our discussion on how SEEs in the terrestrial environment affect electronics is focused on digital CMOS (Complementary Metal Oxide Semiconductor) technology as this represents a large fraction of the commercial market. Indeed, if you open virtually any smart-phone, tablet, or personal computer, you will find some form of central processor to handle computation, display, and various interface tasks. These processors consist of static random access memory (SRAM) partitioned into several distinct areas (data, program, cache, display, etc.) surrounded by “glue” logic to provide functionality and move data between the various blocks. The system will also frequently be augmented with a large amount of dynamic random access memory (DRAM) to allow operations of large programs and a hard disk or flash disk for permanent storage. Studying the impact of terrestrial SEEs on these ubiquitous memory and logic components provides a good, general understanding that can be directly applied to a large majority of commercial devices. Whether induced by a SEE or other mechanisms, a high failure rate predicted by simulation and modeling or extrapolated by accelerated testing is meaningless if it does not ultimately affect the end customer in some detectable, visceral way. The primary reason that the commercial industry worries about SEEs is their dominant impact on observed failure rates in the field. In mature qualified technologies, SEEs currently limit the ultimate reliability performance of most microelectronic products (usually even if significant mitigation strategies such as error correction circuits have been employed). Stated another way SEEs currently induce the highest failure rate of all other failure mechanisms combined. Whether or not this ultimately impacts a customer is simply a matter of the criticality of the application. In non-critical consumer applications (such as a cell

phone) that do not have a life-sustaining role (some might argue with this categorization ☺) a relatively high soft failure rate is often tolerable. In contrast, a similar failure rate in a safety application (such as automotive dynamic stability control, air bag deployment, etc.) would be deemed unacceptably high, particularly where human life or safety is at risk if the circuit fails. In such high-reliability applications, unmitigated SEEs can lead to such high-profile product failures that the manufacturer loses significant market share and revenue. Understanding the impact of SEEs on commercial electronic systems is a crucial task to enable manufacturers to build reliable products. As a "poster-child" for this theme, we present in fig. I-1 an article discussing a real-world SEE problem having a huge commercial impact. This story appeared in a number of publications, but when nuclear physics problems are highlighted in business magazines [2], one can imagine that the problem had significant monetary impact. In this case,



Sun Screen

Daniel Lyons, *Forbes Global*, 11.13.00

mysterious glitch has been popping up since late last year...It has caused problems for America Online, Ebay and dozens of other major corporate accounts...The Sun has caused crashes at dozens of customer sites. An odd problem involving **stray cosmic rays and memory chips** in the flagship Enterprise server line...Sun found it had been shipping servers whose cache modules contained faulty SRAM chips from a supplier it won't name.

...a dot-com in San Francisco has been waiting several weeks for a repair. It bought a Sun 6500 server to run the database that is the core of its business. The server crashed and rebooted four times over a few months. **"It's ridiculous. I've got a $300,000 server that doesn't work. The thing should be bulletproof,"** says the company's president.

Fig. I-1 Excerpt from Forbes Global Magazine article highlighting reliability failures caused by neutron cosmic ray effects in an Enterprise server line from SUN Microsystems. This problem led to significant revenue loss when major SUN customers shifted to other vendors because of a perceived lack of reliability and serviceability.

the flagship product of a top-tier server manufacturer (SUN Microsystems) suffered from a high soft failure rate (~1000% higher than predicted) in its SRAM cache modules leading to unacceptably poor reliability performance for many end customers. The cause of this infamous problem was dominated by low-energy cosmic neutron activation of ^{10}B-doped glass in the SRAM, which led to unacceptably high soft failures in the product. Since the component soft error rate had been seriously under-scoped, the system's response to errors had been configured to simply lock the system into a "safe mode" until service engineers could restore the machine. Had the SRAM vendor been aware of the risk, they could have removed the ^{10}B from its process and avoided the high failure rate. Alternatively, if the server system designers had been aware that the failure rate was going to be so high, they could have implemented mitigation that did not require a stoppage of service. The key point of this example is that SEEs are currently the dominant reliability failure mode in advanced semiconductor products and that

reliability failures that affect end users degrade customer confidence and ultimately lead to significant revenue loss. The revenue loss was estimated to be in excess of a billion dollars [3] due to SUN's loss of credibility in an arena where system reliability, accessibility, and serviceability (up-time) are of paramount importance.

Though not strictly a "terrestrial" occurrence, this incident is a good example of what a SEE can do to complex control systems. Modern aircraft rely heavily on electrical flight control systems (commonly referred to as 'fly-by-wire' systems) where there is no direct mechanical linkage between the pilot's manual controls and the actual flight control surfaces. Flight control computers send movement commands as electrical signals via wires to hydraulic actuators driving the control surfaces. The computers (for reliability the systems usually have triple redundancy) sense the response of the control surfaces and the attitude and kinematics of the aircraft and adjust the commands as required. In December 2011, the Australian Transportation Safety Board (ATSB) released its final report on investigations surrounding a fly-by-wire control system failure leading to a dangerous pitch-down event (nose-dive) [4]. In October 2008, an Airbus A330-303 (Qantas flight 72) on route from Singapore to Perth, Western Australia (flight path shown in fig. I-2) was cruising on autopilot at an altitude of 37,000 ft, when one of its three inertial reference units (monitoring the aircraft's pitch, roll, bank, etc.) started outputting incorrect values. Specifically, in response to erroneous "angle-of-attack" data, suggesting incorrectly that the aircraft was about to enter a stall, the aircraft's computers commanded the aircraft to pitch the nose down (to reduce the attack angle as illustrated in the diagram at the bottom in fig. I-2). This dramatic and sudden change in pitch injured (in some cases severely) more than one-third of the passengers and three-quarters of the crew members. The failure mode has only been observed three times in over 128 million hours of unit operation, and the unit met the aircraft manufacturer's specifications for reliability and undetected failure rates (the aircraft manufacturer subsequently redesigned the algorithm to prevent the same type of accident). The failure

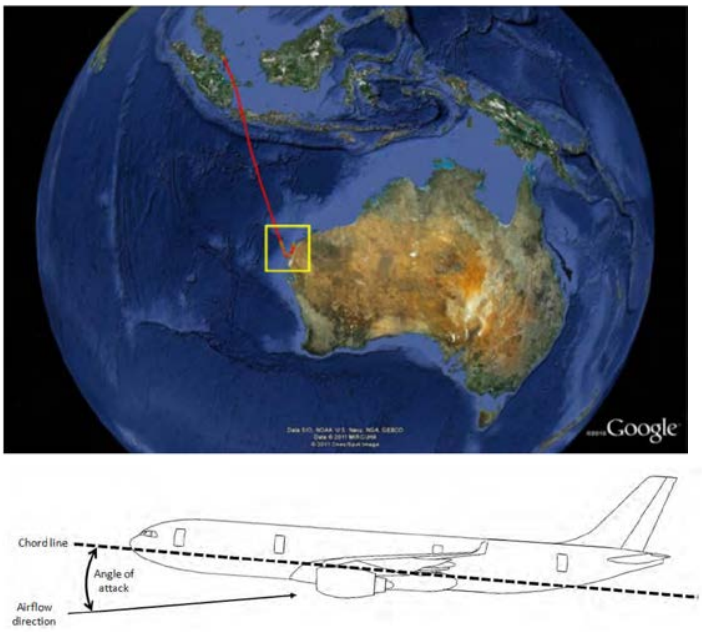


Fig. I-2 Flight path of the aircraft and area where upsets occurred (yellow box). In response to a glitch erroneously telling flight computers that the plane was entering a stall (nose too high for given airspeed = angle of attack too large), when in fact the plane was in level flight, the computer pitched the nose down aggressively causing significant injuries to crew and passengers. Images from [4].

rate 3/0.128 = 23.44 FIT is deceiving since the SEU rate at 37,000 feet in unprotected memory will be thousands of times higher – thus the rate of induced "glitches" will typically be much higher. This implies that the system has a high robustness against glitches (triple redundancy can do that ☺). However, in this particular case, and two others like it, the confluence of a glitch in a specific subsystem and system sensitivity under certain conditions led to a serious incident. Investigators concluded that the failure mode could have been initiated by a single, rare type of internal or external trigger event (since the CPU memory did not have error correction) combined with a susceptibility to that event within a hardware component. One potential triggering event is considered to be a neutron-induced SEE in one of the circuits in the CPU. However, as with many SEE-induced field failures, investigators felt there was insufficient evidence available to determine conclusively if a SEE could have triggered the failure. It should be noted that a proper accelerated experiment using a white neutron source (spallation neutron beam facilities such as Los Alamos, TSL, TRIUMF, etc. provide a spectrum of neutron energies spanning from < 1 MeV to several hundred MeV that properly simulate the avionics environment) would potentially be able to reveal this rare, but obviously consequential, failure mode.

In most digital electronic devices, SEEs induce "soft" errors. A soft error occurs when a radiation event, typically an energetic ion, causes enough of a charge disturbance to reverse or flip the data state of a memory cell, register, latch, or flip-flop. The error is considered "soft" in the sense that the circuit/device, itself, is not permanently damaged by the radiation event (if new data is written to the bit, the device will store it correctly). In contrast, a "hard" error is manifested when the device is physically damaged so that a malfunction occurs, data is lost, AND the damaged state is permanent (power devices can be susceptible to hard failures from SEEs). There is another intermediate failure mode known as an "intermittent" failure that can be sometimes mistaken for a soft error, but intermittent fails are more closely related to hard failures in the sense that they occur in a marginal device or one with a defect or latent damage that, only under certain operating conditions, ceases to function properly. An intermittent failure leads to a location that passes some test but fails others. The distinguishing feature between a true radiation-induced soft error

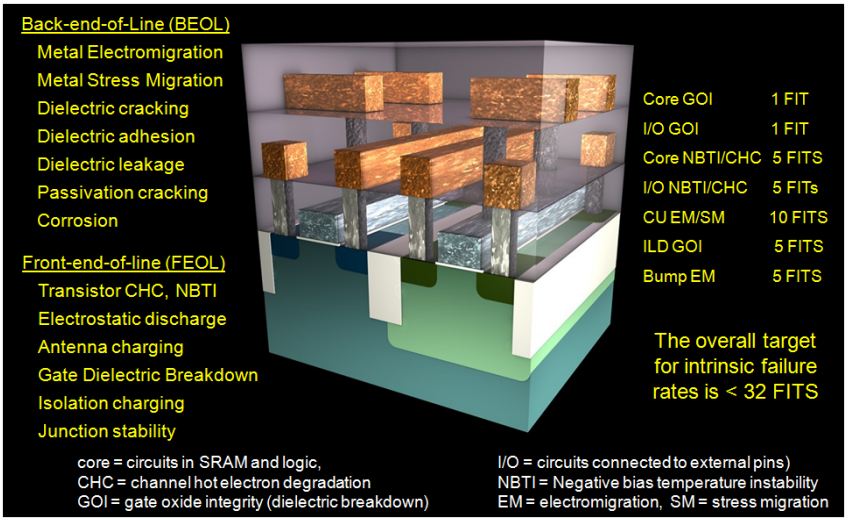


Fig. I-3 List of some of the major reliability mechanisms and a typical "hard" failure budget allowed for each in a qualified semiconductor process. Note that the failure rates listed are budgets (a conservative upper bound) and are largely defined by the temperature and voltage of operation. Thus, for typical product the actual hard failure rate will actually be many times lower than this.

and an intermittent failure that looks like a soft error is that the location of the failing device will be random for the former while the same location will fail in the latter case. For completeness it is important to note that there are several other potential sources of soft-errors not related to particle radiation effects. These include errors induced by electromagnetic interference, noise, and cross-talk (from high-frequency signals and various parasitic glitches occurring in improperly shielded or designed PC boards) [5]-[9]. Like the intermittent failures these will usually have some location dependence and thus can be separated from true particle-induced soft errors. In accelerated experiments these do not pose a problem but when observed as a few field failures it can be hard or impossible to determine that failing locations are repeated. I once had a customer claim of a large number of "soft" field failures; however, the field failure rate was 25x what the predicted soft error failure rate should have been AND was limited to two boards out of twenty (not random in location). I asked them to examine their PC boards and they found that the board manufacturer had failed to install decoupling capacitors on these boards. After the boards were repaired the soft error problem disappeared. In properly designed systems that follow industry standard decoupling, ground loop, and EMI standards, the only soft errors that occur will be due to particle radiation.

The main concern regarding soft errors is that if corrupted data is used in a down-stream process that alters a machine state of the system, it will lead to a system failure (e.g. an invalid instruction, an erroneous calculation or operation, a system reboot, etc.). The rate at which soft errors occur is called the soft error rate (SER). SEEs and soft errors have been used interchangeably, but SEEs include several mechanisms that can induce hard circuit failures. The unit of measure commonly used with SER and indeed all other reliability mechanisms is the FIT (Failure-In-Time). One FIT is equivalent to one failure in one-billion device hours (10^9 device-hours) or 114,155 years! The FIT is thus a VERY small number, as it should be, since we want a very low failure rate in our devices. There are more than a dozen critical reliability mechanisms that can potentially degrade integrated circuit performance [10, 11] if not addressed, but the overall failure rate from all these mechanisms is typically in the 5 - 40 FIT range for qualified products (as shown in fig I-3).

Soft errors have become a huge concern in advanced computer chips because, uncorrected, they produce a failure rate that is higher than all the other reliability mechanisms combined! In stark contrast to the hard failure rates, without mitigation, SER can easily run 10,000 FIT/chip! While 10,000 FIT sounds huge in comparison with typical hard failure rates, it represents less than one failure in 10 years assuming 24 hours/day operation. However, in the context of high-reliability applications, critical systems, or large systems with thousands of chips, this failure rate can become an overwhelming problem. Thus the once-ephemeral, radiation-induced soft error has become a key threat to advanced commercial electronic components and systems in a large variety of customer applications.

I-2. Ions, Junctions, and Types of SEE and their Effects

The primary cause of SEEs in the terrestrial environment is from the deposition of energy caused by the passage of positive ions through, or in, sensitive device layers. These ions, are the nuclear fragments emitted from various neutron reactions [12] or alpha particles (doubly-ionized helium ion) emitted by the natural radioactive decay of uranium, thorium, and daughter impurities in materials [13]. The net effect of the energy deposited by the ion is the creation of a high-concentration of electron-hole pairs (orders-of-magnitude higher than the nominal doping concentrations in devices) along the trajectory [15], the creation of highly energetic electrons (delta-rays) [16], and physical lattice damage in the form of recoil nuclei (each forming an interstitial and vacancy defect as well as additional follow-on secondary damage by the resultant recoils) [17]. The higher the kinetic energy of the ion, the farther it travels before being "stopped" by the material. The distance required to stop an ion (its range) is both a function of its energy and the properties of the material (primarily the material's density) in which it is traveling. A simulation of thousands of 5 MeV alpha particles incident normal to a silicon surface is shown in fig. I-4. Note that the range is fairly well-defined (in this case the alpha particles are stopped at a depth of 27 μm of silicon) even though many of the ions exhibit lateral straggling from the incident axis. The stopping power or linear energy transfer (LET) is a function of the material through which a charged particle is traveling and refers to the energy loss of the particle per unit length in the material (dE/dx). The LET (usually reported in units of MeV-cm^2/mg) is a non-linear function of the ion's mass and energy AND a function of the density of the material (electronic structure and physical structure are also variables but to a lesser degree) through which the particle is traveling. Higher LET implies more energy deposited within a given volume. For example the more energetic cosmic-ray-induced heavy recoils generate larger charge disturbances over shorter ranges as compared to alpha particles emitted from uranium and thorium impurities. This trend is shown in fig. I-5 which compares the LET of various ions in silicon. In silicon a ^{28}Si ion (recoil) with a kinetic energy of 10 MeV has an LET of 13 MeV-cm^2/mg and produces about 130 fC/μm while lighter ions such as alpha particles

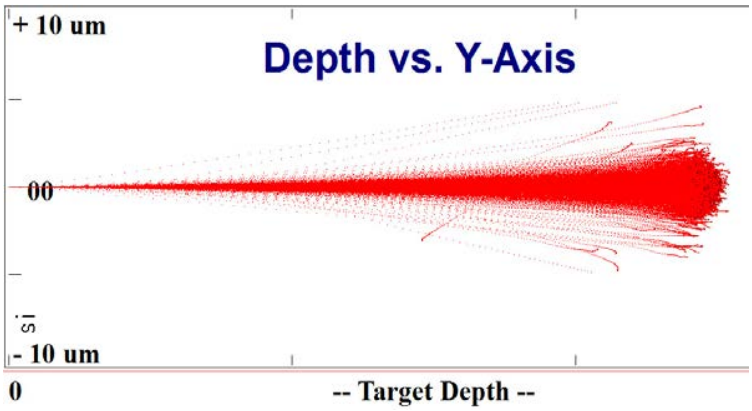


Fig. I-4 A plot from a simulation of 10,000 5 MeV alpha particles normally incident on a silicon target. Note that, on average, the ions are stopped at a depth of 27 μm. Generated with SRIM software [14].

(He) at 10MeV have an LET of 0.4 MeV-cm^2/mg. Thus, heavier ions will deposit more energy over the same path length. The LET in the terrestrial environment is usually limited to < 14 MeV-cm^2/mg (the term "LET" is often used interchangeably to mean effective LET. LET$_{eff}$ accounts for the longer path lengths within a given volume for ions impinging at non-normal angles – the cosine of the incident angle renders LET$_{eff}$ much higher than the normal LET, particularly for glancing angles). For SEE work it can be convenient to convert LET curves to units of charge generated per unit length or femto-coulombs/μm since this unit is far easier to deal with in a device/circuit context (note that the right axis in fig. I-5 shows the LET converted to these units). The LET is converted to charge generated in silicon based on the observation that one electron-hole pair is produced for every 3.6 eV of energy lost by the ion in ionization processes (a fraction of the energy is dissipated through non-ionizing energy-loss or NIEL, and thus does not contribute to the generation of charge) [18]. One can estimate the total generated charge along a specific length of the ion trajectory simply by integrating the LET over the path defined by the ion's instantaneous energy at each point along the path. The result of this estimate is the maximum charge generated or Q_{gen} for a given event assuming some device volume. Later we shall consider how a fraction of this generated charge is actually collected by sensitive device structures.

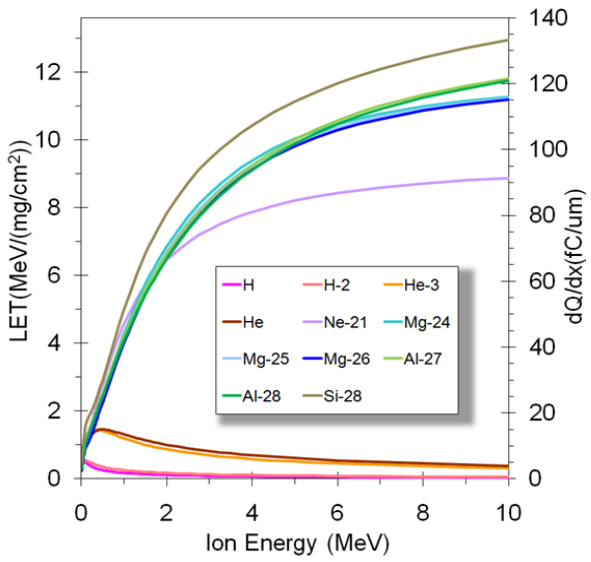
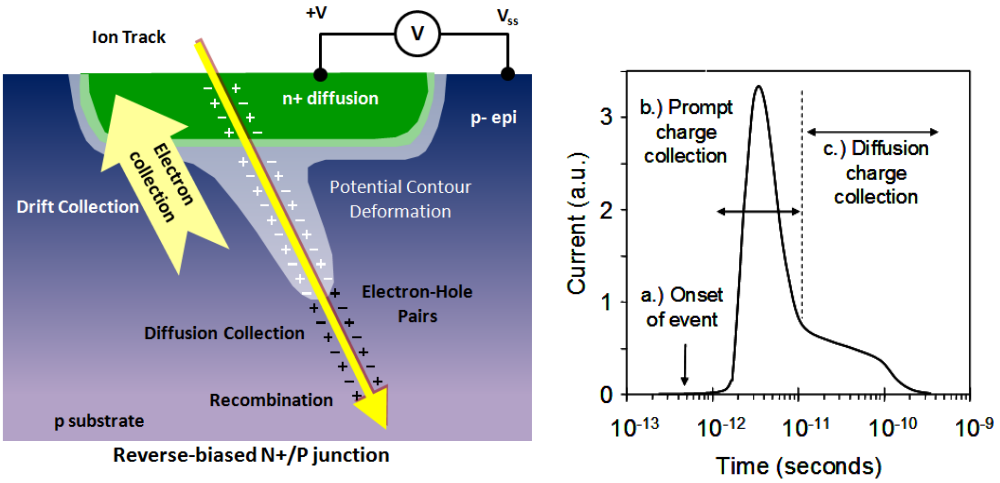


Fig. I-5 Graph of linear energy transfer (LET) of various ions in silicon as a function of incident ion energy. Note that right-hand axis represents charge generated per distance travelled in silicon. Generated with the SRIM software [14].

The most sensitive device structure is the ubiquitous reverse-biased junction. This is fundamentally the "radiation detector" of modern electronic circuits. In worst-case the junction is floating (as in DRAMs, CCDs, dynamic logic circuits, and some analog designs) rendering it extremely sensitive to any charge collected from the passage of an energetic ion. Since no external circuit is restoring the state of the struck node any event will decrease the stored charge. In circuits with feedback (SRAMs, Flip-flops, etc.) the reverse-biased junctions are still as sensitive, but the external circuit may be able to provide enough restoring current (charge) to stop a change in state (soft error) from occurring. In general the n+/p node is worse than its complement because the separation and collection of electrons is more efficient due to the greater mobility of electrons as compared with holes. Additionally, since most of today's manufacturers use p-type substrates, collection can occur deeper within the bulk as compared with p+/n junctions formed in n-wells. Many manufactures use epitaxial layers (epi) with the same doping as the substrate,

with the epi being used solely to reduce surface defects, so unlike older technologies where the epi was actually more-heavily doped, these modern "epi" substrates offer no advantage in mitigating the charge collected during an event. A reverse biased n+/p junction with a positive voltage on the n+ node is shown in fig. I-6a. At the onset of an ionizing radiation event, a cylindrical track of electron-hole pairs (quasi-



a.) b.)

Fig. I-6 a.) Reverse-biased n+/p junction struck by an ion. Note the corresponding deformation of the potential (funnel). Collection proceeds very rapidly as electrons are first collected by drift. The depleted n+ is compensated by the collected electrons and the electric field is weakened, thereby collapsing the funnel as the system relaxes. On a longer time scale collection is dominated by diffusion. b.) The transient current collected as a function of time for this event.

neutral) with a sub-micrometer radius and very high carrier concentration is formed in the wake of the energetic (positively charged) ion's passage (< 0.1 ps) illustrated by the yellow arrow. If the resultant ionization track traverses the junction or its depletion region, carriers are rapidly collected (prompt collection) by the local electric field creating a large current/voltage transient at that node. A notable feature of this type of event is the concurrent distortion of the local electric field into a funnel shape (the so-called field-funneling effect) [19] by the high non-equilibrium carrier concentration. The funnel greatly enhances the efficiency of the drift collection by extending the high field depletion region deeper into the substrate and thus increases the amount of charge collected by the drift process. The size of the funnel is a function of substrate doping (for a fixed doping of the heavily-doped side) – the funnel distortion increasing for lower substrate doping (higher substrate resistivity) [20]. This "prompt" collection phase is typically completed within a few tens of picoseconds and is followed by a phase where diffusion dominates the collection process, as illustrated in the current transient shown in fig. I-6b. Additional charge is collected as electrons diffuse into the depletion region on a longer time scale (hundreds or thousands of picoseconds) until all excess carriers have been collected, recombined, or diffused away from the junction area. The diffusion process is much slower and, typically, the total

charge collected from diffusion is significantly less than that collected initially by prompt collection in the case of advanced technologies [21, 22]. Ions typically produce bigger responses in lightly-doped substrates, since funneling and, hence, charge collection efficiency increases with decreased doping. For events that occur some distance from the sensitive node (e.g. the ion trajectory is far from a junction or depletion region), diffusion will dominate, but ultimately these types of events will result in much less collected charge [23]. Particularly in advanced technologies, the events occur in a sea of nodes in close proximity; thus, it is important to note that charge sharing among nodes is a common feature of collection in real devices. It should be noted that the collected charge will always be less than the charge generated by the ion event.

There are a variety of SEE mechanisms in semiconductor devices that vary in magnitude from "soft" single-bit data disruptions to permanent parametric shifts and complete device destruction. Some of the key acronyms are illustrated in fig. I-7. Of primary concern for commercial terrestrial applications are the "soft" SEEs as opposed to the "hard" SEEs [24, 25] - in the terrestrial environment dose and dose-rate related radiation damage usually is not large enough to create detectable changes in device performance. As the name implies, SEEs are device failures induced by a single radiation event. We use the commercial term "soft error" for these types of events interchangeably to encompass the key "soft" SEEs that effect semiconductor device technologies. However, it is useful to be aware of the characteristics of the different failure modes. Soft errors are also referred to as a single event upset (SEU). If the radiation event is of a very high energy and glancing incidence angle, more than a single bit may be affected, creating a multi-bit upset (MBU) or multi-cell upset (MCU) as opposed to the, more probable, single bit upset (SBU). MBUs/MCUs are usually a small fraction of the total observed SEU rate, their occurrence has implications for memory architecture in systems utilizing error correction [26]-[29]. The distinction between MCU and MBU is that the MCU is the total number of bits upset by an event, while the MBU is the total number of bit upsets in a single logical word. Thus an MCU is composed of at least two logical words afflicted with either an SBU or MBU. Another type of soft error occurs when the bit that is flipped is in a critical system control register, such as that found in field-programmable gate arrays

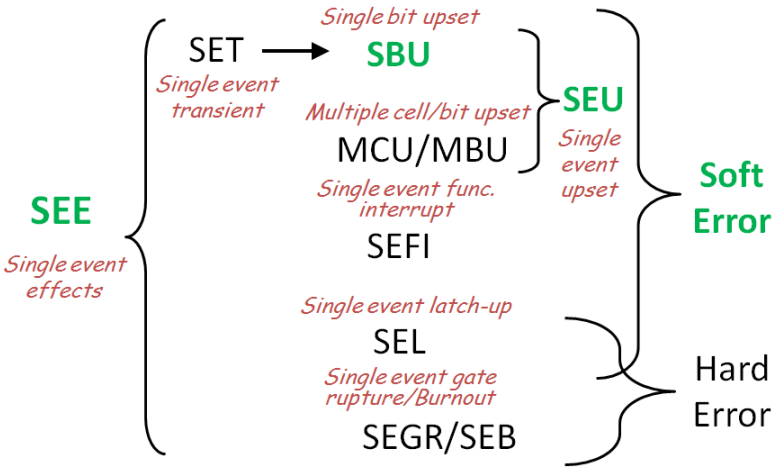


Fig. I-7 Some common acronyms associated with single-event radiation mechanisms in semiconductor devices (dose effect excluded).

(FPGAs) or memory control circuitry, so that the error causes the product to malfunction [30,31]. This type of soft error, called a single event functional interrupt (SEFI), impacts the product reliability since each one leads to a product malfunction as opposed to typical memory soft errors that may or may not affect the final product operation depending on the algorithm, data sensitivity, etc. Radiation events occurring in combinational logic result in the generation of single event transients (SET) which, if propagated and latched into a memory element, will lead to a soft error [32] (in the literature single event transients in analog electronics are often referred to as ASET while those in digital combinatorial logic are referred to as DSET). A SEE can also cause disruption of electrical systems indirectly by turning on the two parasitic bipolar transistors between CMOS well and substrate – inducing a high-current latch-up condition [33, 34]. The only difference between single event latch-up (SEL) and electrical latch-up is that the current injection that turns on the parasitic bipolar elements is provided by the radiation event instead of an electrical over-voltage. SEL can also be debilitating since its occurrence will necessitate a full chip power down to remove the condition. In general SEL is non-destructive since current is limited by external resistances, however, in extreme cases if enough current is drawn, latent or catastrophic damage to metallization and junctions can lead to hard errors. The last failure mode related to SEEs is the single event gate rupture (SEGR) and the single event breakdown (SEB) [25]. Both mechanisms are destructive and lead to hard failures. In SEGR the gate oxide is rendered to a high conduction state (breakdown) while in the SEB the junction is broken-down when the event causes avalanche and thermal run-away. Both these mechanisms are not usually associated with low power digital circuits but have been observed in power MOSFET devices.

The system impact of a SEE depends on the type of error(s) that is induced and its location. As illustrated in fig. I-8, a single-event upset (SEU) in a memory array can affect a single bit (single bit upset or SBU) or several bits (a multiple-cell upset or MCU). To determine the reliability impact MCU is divided into cell upsets within logical words. A multi-cell upset in a single word is called a multi bit upset or MBU. Note that a single MCU can create multiple SBUs, a combination of SBUs and MBUs, or multiple MBUs. MBUs are much more likely to be produced by high LET events and, thus, in the terrestrial environment, higher energy neutron

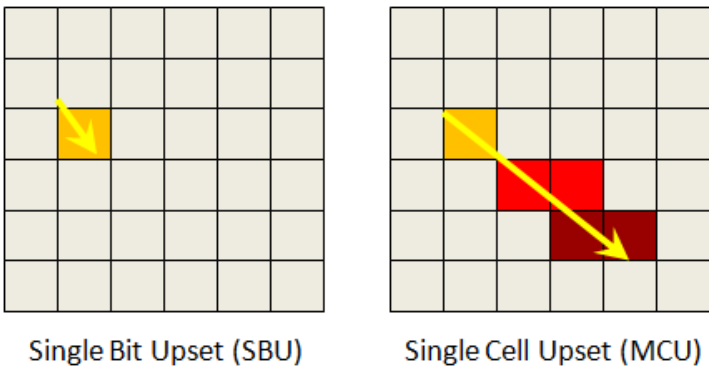


Fig. I-8 Diagram of a single bit upset (left) and a multiple cell upset (right) in a memory array. Note that this MCU contains one single bit error (orange) and two different 2-bit multiple bit upsets (red and brown). In most ECC implementations that correct single bit errors and detect double bit errors, this particular MCU would lead to one correction and two uncorrectable, but, detectable, error events.

reactions are responsible for the majority of larger bit MCUs/MBUs [35]. Since a single event induces an MCU, the MCU fail pattern will typically be contiguous and follow a trajectory. However in systems with different data state sensitivity, some non-contiguous fail patterns may be seen. If caused by multiple separate events in a highly accelerated experiment, adjacent bit errors should not be considered MBUs since under unaccelerated conditions these types of failures would never occur. Statistical methods can be applied to sort out adjacent SBU that appear to be MBU along with fast bitmapping. Another type of soft error occurs when the bit that is flipped is in a critical system control register, such as that found in memory control circuitry, so that the error causes the product to change its operating mode (e.g. a read operation suddenly and erroneously becomes a write operation). SEFI's obviously impact the product reliability more dramatically than a SBU/MCU since each SEFI leads to a direct product malfunction and, typically, many errors are introduced. SEFI's can often be cleared from the system by proper mode selection (one that induces a full reset in all control registers). Though often recoverable, the impact of the reset will depend on the product application.

In a combinatorial circuit, where the output is based on a logical relation to the inputs (with no capability for retention), if enough radiation-induced charge is collected, a short-lived transient in the output will be generated (a single-event transient or SET, as shown in fig. I-9). If this radiation-induced "glitch" is actually propagated to the input of a latch or flip-flop during a latching clock signal, the erroneous input will be "latched" and will be stored. For older technologies the SET could not propagate since it usually could not produce a full output swing and/or it was quickly attenuated due to large load capacitances and large propagation delays. In advanced technologies where the propagation delay is reduced and the clock frequency is high, the SET can more easily traverse many logic gates, and the probability that it is latched increases since the operating frequency is much higher. Once a SET can propagate freely, synchronous, and especially asynchronous (self-clocked), circuits will be extremely sensitive to such events. In technology nodes beyond 90nm and at high product operating frequencies, there is increased risk that a large fraction of observed soft failures will be related to latched SET events.

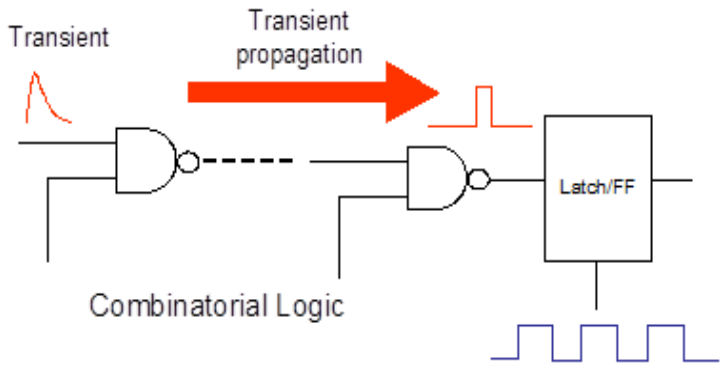
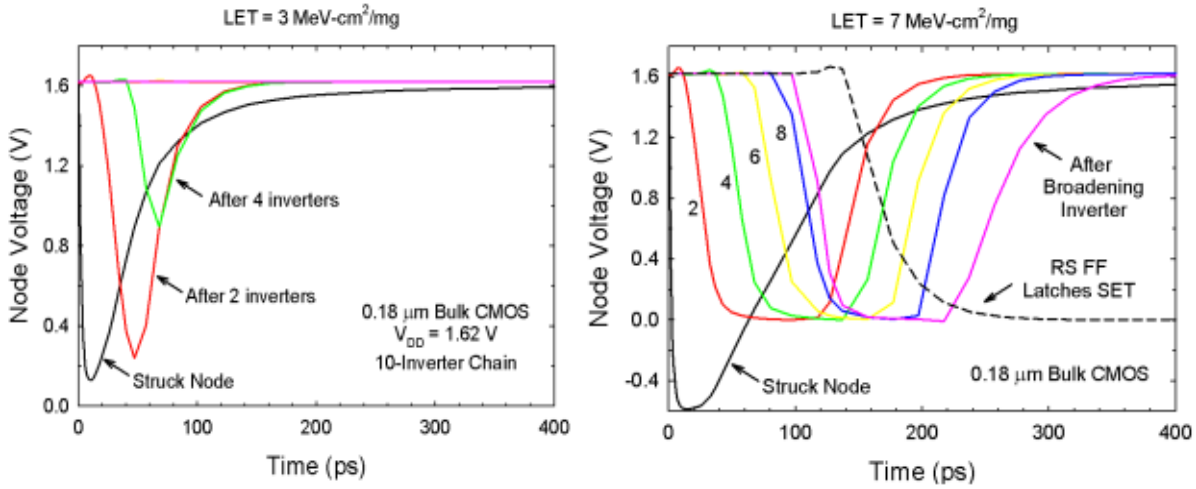


Fig. I-9 Representation of a single event transient (SET) in combinatorial logic. An SET only becomes a problem if it is latched or captured in a memory element if it arrives during a clocking event.

The key feature that must be considered for SET risk assessment is; will it propagate? This depends on the circuit response and the LET of the event. Higher LET events produce wider pulses (since the charge density is higher) with larger voltage disturbances, thus, increasing the probability that the SET will be

propagated to a memory element as illustrated in fig. I-10. If the pulse does propagate to the input of a



a.) b.)

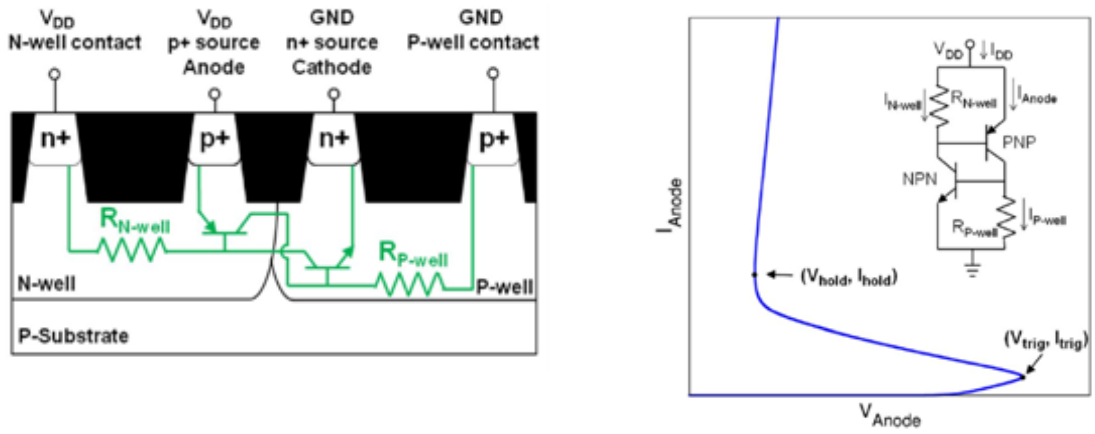
Fig. I-10 a.) Simulations of 180 nm node voltage as a function of the number of inverter stages (distance traveled) for a 3 MeV-cm^2/mg event, and b.) for a 7 MeV-cm^2/mg event. Note that for the lower LET event the pulse is very narrow and not full swing and thus, after four stages it is attenuated, while the higher LET event causes a rail-to-rail pulse that is wider and propagates easily – this type of event would likely cause an SEU if captured. After [36].

sequential or memory element, then the next question is; will it be latched? Even if the pulse propagates through many gates, if it is not latched into a memory or sequential component, it will NOT affect the system since its occurrence will not have been "captured". The probability of a SET being captured depends on when the pulse arrives, the duration of the set-up and hold time, the clock frequency, and the SET pulse width. Only if the SET pulse changes the input value of the sequential or memory element as it is clocked during the set-up and hold time will it be captured and produce an SEU. One aspect of recent digital design (optimized to limit switching power losses) that may limit the impact of SET is the extensive use of clock-gated sequential logic. This means that while the flip-flops can operate at higher clock frequencies enabled by newer technologies, clock-gating turns off large numbers of flip-flops that are not needed for particular operations, thus greatly reducing their effective frequency and, thus, their ability to catch a transient erroneous value from an incoming SET. Recent reports on the impact of terrestrial SET-induced soft errors, even at advanced nodes (32nm), suggest that the contribution of latched SET is less than 5% at frequency [37].

The potential for latch-up exists in virtually all junction-isolated or bulk CMOS processes (an exception would be an isolated-well CMOS). Latch-up is the formation of and sustained parasitic *PNPN* (in CMOS based on p-substrates) paths creating a low impedance connection between power and ground. A single event depositing enough charge to turn on the parasitic vertical and lateral bipolar transistors in the CMOS well structure can induce a single event latch-up (SEL). Although not exact (things are a bit more complicated), a rough approximation of the requirement for latch-up to be possible is that the loop

gain of relevant *PNPN* must exceed unity ($\beta_{pnp}\beta_{npn} \geq 1$). A triggering stimulus must allow the loop to achieve the current level required to turn itself on. For steady-state latch-up, the circuit must be able to supply the holding current and voltage required to sustain the loop. The "typical" parasitic structure is illustrated in fig. I-11a. If sufficient charge is injected so that the N-well voltage rises one diode drop (~ 0.7V) above the anode voltage, the N-well/anode junction becomes forward biased, turning on the parasitic PNP transistor. The increased anode-current then forward biases the emitter/base junction of the parasitic NPN, turning it on as well (V_{trig}, I_{trig} on the plot in fig. I-11b). With both of the BJTs turned on, there is a low impedance path between power and ground (in modern circuits the latch-up current is limited by external resistance so that burnout of the metallization seldom occurs, BUT latent damage can be generated, inducing a hard failure later in time). The parasitic PNPN structure will remain in its "latched" high-current state until the anode voltage is reduced to a point where the voltage drop across both parasitic bipolar transistors is reduced below the turn-on voltage (effectively turning them off). From a process point of view, as R_{sub} and R_{nwell} are increased (in processes using lower doping levels), the likelihood that transient charge will turn on the parasitic bipolar transistors increases (since the required forward diode voltage drop is more easily generated). Similarly, in technologies with a larger distance between contacts (effectively increasing the R_{sub} and R_{nwell}), a higher sensitivity to SEL has been reported [40]. Heavy-ion studies from this work demonstrated that, as expected, the SRAM with the 8x contact spacing was significantly more sensitive to SEL than the baseline SRAM with 1x contact spacing. Heavy ions are only present in the space environment but are useful to indicate differences in sensitivity that would not generally be seen in with neutron testing or terrestrial environment. In fact, neither of the SRAM devices exhibited SEL at high temperature (125°C) and high voltage (1.4V) conditions during extensive neutron testing at Los Alamos, thus, from a terrestrial point-of-view, the 1x baseline and 8x

SRAMs were identical in terms of SEL. As operating voltage is increased, sustained latch-up becomes possible when the operating voltage exceeds the holding voltage. Hence, technologies with an operating voltage lower than the holding voltage are robust against the creation of sustained latch-up. However, transient SEL is still a potential failure mode where bits or circuits can be corrupted even when the latch-up condition is not sustained. Operating CMOS circuits at higher temperature increases the susceptibility to SEL since the small-signal forward current gain increases with increasing temperature (most SEL



a.) b.)

Fig. I-11 Cross-section of (a.) of twin-well CMOS showing parasitic bipolar transistor network responsible for latchup and I-V curve (b.) showing the trigger point and snap-back that occurs during latchup. From [38, 39].

experiments are conducted as 85° to 125°C). The ion energy, type of ion, and its trajectory through the device determine the ion's ability to create an SEL event. Due to the more invasive nature of the bipolar injection that occurs during an ion event (as compared with the typical unipolar injection of an electrical latch-up), SEL can be observed in CMOS devices that pass all electrical latch-up testing. So a good, general rule-of-thumb is that good electrical latch-up performance is "necessary but not sufficient" to ensure good SEL performance. Ultimately, accelerated particle testing is the only accurate way to assess SEL susceptibility of a particular circuit. In general, fully-depleted CMOS SOI with body-ties is potentially immune to SEL, although some BiCMOS SOI can, and do, exhibit SEL in some cases [41].

I-3. Q_{crit}, Q_{coll}, DRAM, and SRAM Single Event Upset (SEU)

Collected and Critical Charge

For understanding if a particular event causes an upset in a specific device, the generated charge is not an accurate measure of the disruption it will cause because the charge is usually distributed across multiple device nodes, and recombination will eliminate some of the original charge generated. Thus, generated charge will be a conservative over-estimate of the impact on the device nodes since they will certainly collect less than the total generated charge. The magnitude of the collected charge, Q_{coll} for each node depends on a complex combination of factors including the size (volume) of the node (usually a reverse-biased junction) and its depletion region, the biasing of that node, and the size of the collection-enhancing field-funnel as determined by the local doping parameters, the substrate structure (buried layers, buried insulators, etc.), the state of the device during the event, and, of course, all the factors that define the generated charge. You have probably heard of the "signal-to-noise" response, or ratio, of various circuits - the Q_{coll} is the "noise".

Fig. I-12 Simplistic weight scale analogy of the "battle" between Q_{crit} and Q_{coll} every time a single event occurs. If $Q_{coll} < Q_{crit}$ (top) then the circuit "passes" while if $Q_{coll} > Q_{crit}$ (bottom) then the circuit "fails".

The second half of the "equation" is to determine the device's sensitivity to the excess charge to determine if an upset occurs for a given ion event. The device node sensitivity is largely defined by the node capacitance, operating voltage, and other factors that define the amount of "signal" charge, or critical charge (Q_{crit}). Simplistically, Q_{crit} is often defined as the minimum amount of charge collected at a sensitive node that causes the circuit to suffer a soft error. This definition is simplistic because Q_{crit} may not be a constant, varying with the dynamic sensitivity (clock-gated circuits, circuits with feedback, etc.) of the circuit and the

time evolution of the charge injected by the event. That being said, the simple concept of Q_{crit} is helpful in setting the foundation for more complex behaviors. To this end, let us image a simple, old-style weight scale, where weight of a sample placed on one side is determined by a known reference on the other. Since Q_{crit} is defined by fixed circuit parameters, layout, and operating voltage, for a given state of the circuit, this value can be viewed as a constant (reference). In stark contrast, the Q_{coll} varies over a large range based on the ion type, energy, trajectory, etc., as we have already mentioned, so Q_{coll} is the unknown quantity to be "weighed". For proper circuit operation, unlike a real scale, we always want the signal (Q_{crit}) to be larger than the noise induced by Q_{coll}. If $Q_{coll} < Q_{crit}$ then as the illustration at the top of fig I-12 demonstrates, the Q_{crit} "out-weighs" Q_{coll}, and the circuit does not suffer a soft error. In such a case the signal is greater than the noise induced by the event and thus the circuit operates normally, albeit with a reduced signal margin, since some of the charge was compensated by the Q_{coll} (remember we are assuming no feedback in this simple example). If, on the other hand, the radiation event induces enough charge such that $Q_{coll} > Q_{crit}$ then noise swamps out the signal. The Q_{coll} "out-weighs" the Q_{crit} and a soft error occurs as illustrated in the bottom image of fig. I-12. With this simple interpretation in mind, we now move to SEEs in "real" devices.

Dynamic Random Access Memory (DRAM)

The advantage of DRAM is the simplicity of its bit-cell - only requiring a single transistor (and some area for the storage node junction) compared to the six transistors typically used in SRAMs. DRAM, therefore, allows much higher levels of memory density (bits/area). On the negative side, the charge storage on the capacitor is not stable, and the device must be refreshed over short intervals to ensure that the storage node maintains its data. This refresh requirement leads to increased circuit complexity and relatively high power dissipation even when the memory is not being accessed. Let us define the charged state (full voltage on the cell node) as a "1" and the absence of charge, or fully discharged node capacitance, as the "0" state. Since the capacitor will naturally discharge as a consequence of parasitic leakage current across the plates of the capacitor and the junction leakage, the "1" data state will decay

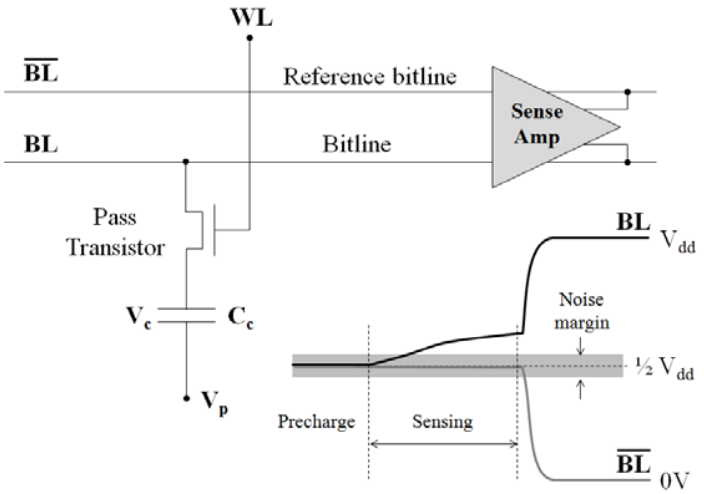


Fig. I-13 A typical 1T-1C DRAM cell (top) and the voltage waveforms (bottom) of bitline and bitline-bar during a sense operation (with a "1" data state). The vertical scale is exaggerated for clarity the actual ΔV between the bit lines is typically < 100mV.

into the “0” data state if not refreshed. Hence the “1” data state is unstable while the “0” data state is stable since no charge or voltage is required for the “0”. The most common DRAM structures employ a single transistor connected between the storage cell capacitor and the bitline, which is connected to the input of a differential amplifier used during sensing (the sense amplifier) in the so-called 1T-1C configuration shown in fig. I-13. When the cell is refreshed or read (the refresh is essentially an automated read operation), the pass transistor is turned on by the word line (WL) connected to the gate, and the stored signal charge is transferred from the bit-cell to the bitline (BL). The bitline is connected to one input of a differential amplifier (the sense amplifier) with the other input of the amplifier connected to a reference bitline (BL-bar) with no cell capacitor. Prior to the sensing operation, both the bitlines are pre-charged to the same voltage, usually to one-half of the supply voltage. If a “0” is stored on the bit-cell, then the positive pre-charge voltage on the bitline will be forced to lower value with respect to the reference bitline voltage. For a stored “1” the bitline voltage is forced to a positive value as compared with the bitline reference voltage. In each case, a positive or negative differential voltage is developed between the amplifier inputs. The sense amplifier offset voltage (on the order of tens of mV) defines the minimum voltage difference (between the two bitlines) that can be reliably sensed (gray region in the bottom diagram in fig. I-13).

There are fundamentally three ways in which a single-event can induce a soft error within a DRAM. The most prevalent soft error source is a single-event charge collection by the junction of the bit-cell capacitor. These bit-cell errors are caused by the charge collected during an ion strike in or near the storage node (the drain of the pass transistor and the junction area) directly compensating signal charge stored on the node. At cycle times lower than 100 MHz bit-cell single-events predominate as the primary cause of DRAM errors. If a cell collects more charge than the Q_{crit} for the circuit, the “1” data state will be destroyed. Storage node hits will generally always be a single type of data error. For a simple reverse-biased junction with no feedback like a DRAM bit-cell in retention mode (isolated for the bitline), a soft error will be induced if a single event generates enough charge so that $Q_{coll} > Q_{crit}$ where Q_{crit} has the form:

$$Q_{crit} = C_{node} \cdot V_{data} = C_{node} \cdot \tfrac{1}{2} V_{DD} \qquad \text{(eq. I-1)}$$

Where C_{node} is the bit-cell capacitance and V_{data} actually half of the supply voltage, V_{DD} (assuming the “1” and “0” data state are symmetrical and equal amount of voltage is assigned to both data states). As we mentioned previously, due to the sense amplifiers non-zero offset voltage, each data state actually represents a little less than one-half the full array voltage. Since most bit-cells are in retention mode (not being accessed) nearly all of the time, the bit-cell is sensitive to single events practically at all times. Even during sensing, a charge injection from a single-event in the bit-cell can get transferred to the bitline and

cause an error. The only time the circuit is potentially immune is in the short period after sensing when the sense amplifier is fully driving the bitlines to the correct data state – most single events will not produce enough Q_{coll} to compensate the sense amplifier drive when it is forcing the bitlines. The rate of bit-cell events stays constant or decreases slightly with decreasing cycle-time. The second mechanism for soft errors in DRAMs occurs when the single-event injects charge to the bit-line during the sensing operation (while the sense-amp is trying to "read" the data state of the bit-cell). While the vulnerability time is short for any given bit-line, ion-induced Q_{coll} can occur from any of the access-transistor drains along the bit-line (usually hundreds) or from a direct strike to the sense amplifier itself. As the DRAM cycle-time is reduced below ten nanoseconds, bitline hits will start contributing to the overall failure rate. Since the bitline is only sensitive (floating) during the sensing portion of the timing, and the sense timing is largely fixed by the parasitic capacitances/resistances of a particular DRAM technology, decreasing the cycle-time increases the percentage of time the sense amplifier is sensitive to a radiation event. The rate of bitline-induced soft failures increases linearly with increases in frequency. For modern high-speed DRAMs a significant portion of the observed soft fail rate is due to bitline or sense-amplifier hits. The third, albeit, less common source of soft errors occurs when a glancing event temporarily shunts the source-drain of the pass-transistor [42] thus electrically connecting the bit-cell the bit-line. This transient "connection" can upset the stored data state by charge injection from the bitline (this can occur for either data state). In stack or stack-fin designs (Cell-Over-Bitline) this type of event is very rare since only events with a glancing trajectory that is aligned with the transistor channel and that has a high enough LET can create a conductive filament to form between the source-drain regions. In basic trench structures (with the charge stored in the silicon) this effect is more pronounced and gets worse with scaling. In either case the "target" area is larger than a bit-cell strike so the probability of a hit in these areas is higher - the only saving grace here is that the event must occur during the narrow sensing window for that particular bitline so probability can still be quite small).

To address pause-refresh and soft error problems while increasing packing density, DRAM manufacturers have developed three-dimensional capacitor designs that significantly increase the Q_{crit} while greatly reducing junction collection efficiency by eliminating the large two-dimensional storage junction in silicon [43]. Collection efficiency decreases with the decreasing volume of the junction (junction/well doping also play a role) while the cell capacitance remains relatively constant with scaling since it is defined by external three-dimensional capacitor cell structures (trench, collared-trench, stack, stack-fin, etc.). Of course, concurrent voltage scaling of the power supply reduces Q_{crit}. However, in the last few generations, voltage scaling has saturated in the 1V regime while junction volume scaling decreases the charge collection efficiency with each successive technology node. Consequently, DRAM SER sensitivity to soft errors is shrinking about 3-5x per generation as illustrated in fig. I-14. Contrary to

the popular misconception that DRAM SER is problematic (undoubtedly left over from the days when DRAM designs utilized planar cells and were basically arrays of radiation detectors), DRAM bit-cell sensitivity has been reduced by more than three orders-of-magnitude over the last seven generations. DRAM is currently one of the more robust devices in terms of soft error immunity in the terrestrial environment. It is important to note that bit-cell sensitivity reductions, however aggressive, will not lead to reliability improvements of similar magnitude if the amount of memory is increased concurrently at each new technology node. Indeed, as application and system requirements have increased, the memory size has grown almost as fast as the SER reduction provided by the technology scaling. Thus, over-all DRAM system reliability has remained roughly constant over many generations.

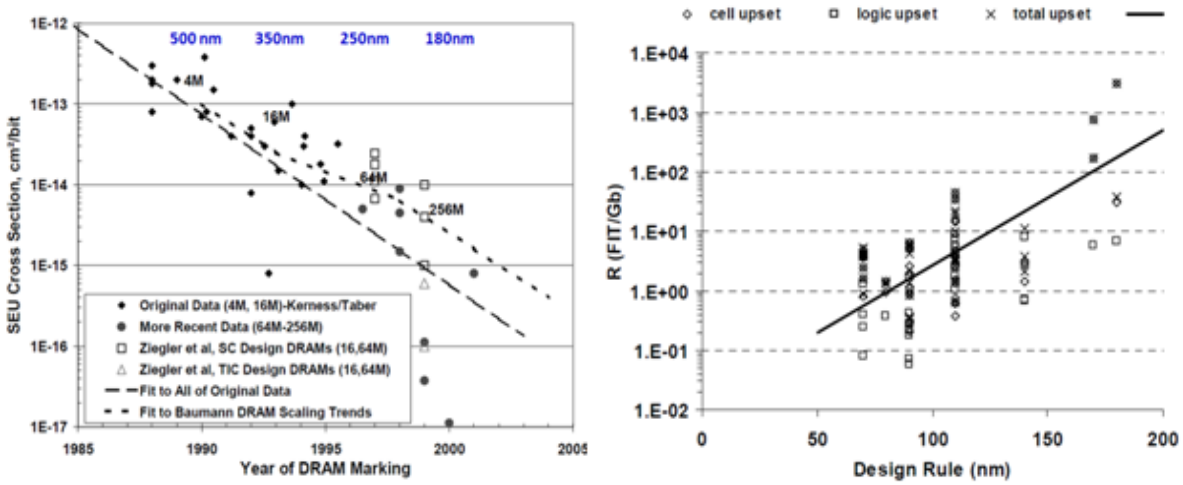


Fig I-14 SEU/SER reduction due to DRAM technology scaling (from 500 to 65nm). Once the bit-cell capacitance was decoupled from junction area (with the use of 3D structures) DRAM bit-cell sensitivity dropped from 3-5x per generation [44, 45]. The variation from vendor-to-vendor is largely related to the cell design being used.

Static Random Access Memory (SRAM)

The standard 6T SRAM cell is composed of a storage cell composed of two PMOS and two NMOS transistors forming two cross-coupled inverters as illustrated in fig. I-15. Two additional pass transistors (usually NMOS) allow access to the bit-cell nodes. Beyond this 6T cells, additional transistors are sometimes added to create 8T or 10T SRAM memory cells that allow memory features such as multiple output ports for register files. Unlike the DRAM with one stable state (capacitor discharged) and one dynamic state (capacitor fully-charged), the SRAM bit-cell has two symmetrical stable states (it is a bistable device). Access to the SRAM memory cell is enabled by the wordline (WL). When a particular bit is being read or written to, the wordline goes high, turning “on” both pass transistors simultaneously, which connects the two internal storage nodes (between P1/N1 and P2/N2) to the two bitlines (note that

one node always stores the complement of the other node's data state). In a similar fashion to the DRAM sense operation, the voltage on the storage node drives the bitline voltage higher or lower than its pre-charged value. Unlike the DRAM for which there is only one driven bitline, each bitline is driven in the opposite direction. While it is, theoretically, unnecessary to drive the two bit lines independently, using the signal and its inverse doubles the noise margin and greatly improves the data integrity. In retention or storage mode, the wordline is brought low, effectively isolating the memory cell from the bitlines. The output of one inverter drives the input of the other, and regenerative feedback maintains the data state of the cell as long as power is applied. Note that with a "1" data state stored on the left, by definition, the opposite, a "0" data state is stored on the right. Thus the left PMOS (P1) is "on" pulling up on the left storage node to keep it "high" while ensuring that the NMOS on the right (N2) is "on" and thus pulling down the right storage node to ground. This, in turn, ensures that the left PMOS (P1) is "on", ensuring the left node is kept high. Now suppose that an ion traverses near the N+ drain node of N1 storing the "1" data state on the left side. As described previously a large "cloud" of electron-hole pairs is produced along the wake of the particle's trajectory, and a large portion of these charge carriers are separated by the local electrical field. For this case, electrons will be collected causing a rapid drop in the stored voltage of the left node. As the node voltage drops, current from the left PMOS (P1) will start to compensate. Note that because of the feedback, as the node voltage drops on the left side, there are consequences for the right side as the right PMOS (P2) starts to turn on while the right NMOS (N2) is turning off. This further aggravates the situation since this will tend to turn off the left PMOS (P1) while turning on the left NMOS (N1), forcing the left node voltage lower. If the current from the left PMOS (P1) cannot quench the excess charge collected by the "1" left storage node and its voltage drops below critical trip value, the SRAM bit-cell will switch states producing an error. Latches and flip-flops (usually in a two-stage master and slave configuration) operate in a similar fashion to the SRAM cell in retention mode, and they are sensitive to radiation events in the same manner as SRAM. Sequential logic does not

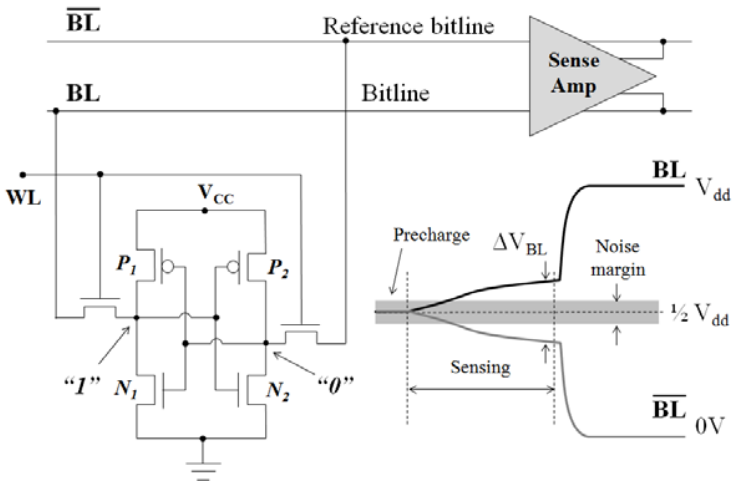


Fig. I-15 Schematic of common 6T SRAM bit-cell and the voltage waveforms of bitline and bitline-bar during sensing. Note the vertical voltage scale is exaggerated; the actual ΔV between bitlines during sensing is typically less than 100mV.

employ sense amplifiers, but the drive of the transistors is usually larger to ensure stable switching and data state retention performance.

In contrast to early DRAMs (one famous electronics writer/hobbyist even suggested using 64k DRAMs with glass lids as an inexpensive imager due to their high light sensitivity [46]), early SRAM technologies were more robust against terrestrial single events because of their high operating voltages and the fact that data in an SRAM is stored as an active state of a bi-stable circuit made up of two cross-coupled inverters, each driving the other to keep the SRAM bit in its programmed state. The Q_{crit} for the SRAM cell is largely defined by the charge on the node capacitance as with the DRAM bit-cell. However the SRAM has a dynamic second term related to the drive capability of the transistors ($I_{restore}$) keeping the node voltages at their proper values, and τ_{switch} which is related to how fast the opposite data state is latched by the SRAM. The Q_{crit} for the SRAM can is expressed as:

$$Q_{crit} = C_{node} \cdot V_{data} + \tau_{switch} \cdot I_{restore} \qquad \text{(eq. I-2)}$$

Where C_{node} and V_{data} are the node capacitance and voltage margin of the data state (in the case of SRAM this is the voltage difference between the point at which the cell switches and the voltage when a full "0" or full "1" is written to the cell). The dynamic second term increases the Q_{crit} of an SRAM. The stronger the transistors providing restoring current to each storage node, the larger Q_{coll} from a single-event must be to push the node voltage over the switching threshold. Thus, using higher-drive transistors (be aware that using larger transistors reduces bit density while potentially increasing collection efficiency) increases the amount of current available to compensate for the extraneous charge injected on a storage node. The τ_{switch} term corresponds to the response time of the circuit when perturbed. It is a measure of how fast the circuit reinforces a data state change based on a the change in storage node voltage. Additional charge margin is provided by this term increases if τ_{switch} is longer since the restoring currents have a longer window in which to provide charge [47]. This is the basis for the success of resistive hardening where resistors are added to deliberately slow down the SRAM and, hence, increase the dynamic Q_{crit}. With technology scaling, the SRAM junction area has been deliberately minimized to reduce capacitance, leakage, and cell area while, simultaneously, the SRAM operating voltage has been aggressively scaled down to minimize power. These commercial SRAM device scaling trends, having occurred over the last decade, are shown in fig. I-16. Initially with each successive SRAM generation, reductions in cell collection efficiency, due to shrinking cell depletion volume, were cancelled out by big reductions in operating voltage and reductions in node capacitance. It can be seen that SRAM bit SEU initially was increasing with each successive generation (this is particularly true in technologies using boron-doped glass layers such as BPSG, where thermal neutrons create soft failure rates that are 3-8x times higher than other mechanisms. Since BPSG has been removed from most processes at or beyond

180nm we do not include its effects in fig. I-16). More recently, as feature sizes have been reduced into the deep submicron regime (< 130 nm), the SRAM bit SEU is now decreasing at every node. This decrease is primarily due to the saturation in voltage scaling (around 1V operation), reductions in junction collection efficiency, and increased charge sharing due to short-channel effects with neighboring nodes.

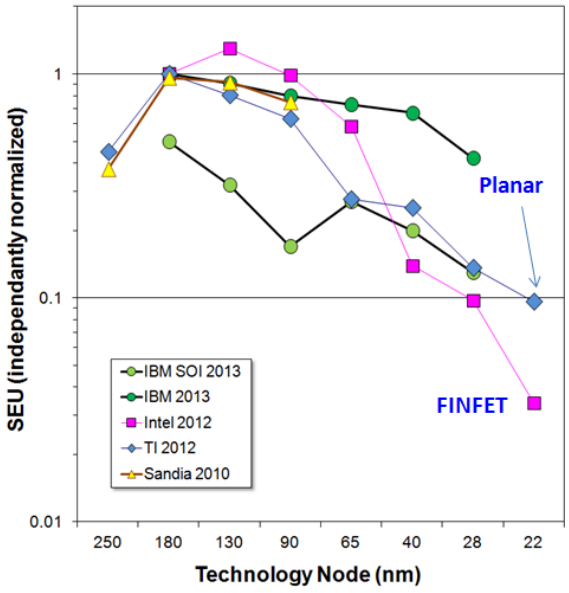


Fig. I-16 Normalized SRAM bit SEU from cosmic ray neutrons as a function of feature size (technology node). 3D SRAM bit-cells such as FinFETs promise even greater reduction in SEU as illustrated at the 22nm node.

Ultimately, just like the DRAM case, scaling also implies increased memory density (number of bits per application), so the reduction in SRAM bit SER does not translate to a reduction in the SRAM system SER. The exponential growth in the amount of SRAM in microprocessors and digital signal processors has led the system/chip SER to increase with each generation with no end in sight. This trend is of great concern to chip manufacturers since SRAM constitutes a large part of all advanced integrated circuits today. Although we focused only on SRAM in this section, sequential logic such as flip-flops and latches function in a similar way and are currently exhibiting similar SER trends on a per-flop/latch basis as compared with SRAM bit-cells in the same technology. In current technologies, the failure rate of a single flip-flop instance is similar to that of an SRAM bit. However, since the instances of sequential logic represent only a small fraction of the memory instances (most products integrate large amounts of memory with a much smaller amount of sequential and combinatorial logic), soft errors in logic have a much less significant impact on actual product failure rates. Additionally, soft errors occurring in sequential logic only lead to machine-state failures less than 10% of the time in most applications (e.g. 90% of the soft errors occurring in sequential elements do not cause any system malfunction), therefore SER in logic is typically a small fraction of the total chip SER. SER in sequential logic is usually only a concern in products where much, or all, of the critical memory is protected with error correction circuits (ECC) [48]-[50]. In such cases the product failure rate will be much lower, since memory errors are corrected when they occur (assuming they are all single-bit upsets), leaving soft failures in the sequential logic the dominant product failure mode.

I-4. The Terrestrial Radiation Mechanisms

a. High-Energy Neutrons

We have considered, in detail, how ions lose energy in matter and how this energy creates spurious charge transients that can upset semiconductors in a variety of ways. Now we shall focus on the different mechanisms that create these ions in the terrestrial environment. While we are focused on terrestrial radiation effects, it is important to appreciate that a sizable fraction of the ions that cause SEE are cosmic in origin. The terrestrial neutron spectrum is the direct result of high-energy cosmic radiation interacting with the Earth's magnetic field and filtered by the atmosphere. Primary cosmic rays are composed of galactic particles (E >> 1 GeV) that are theorized to be left over from the Big Bang or remnants of Super Novae – the cosmic rays reaching Earth are mostly protons with a small fraction of energetic light and heavy ions. There is an additional component of protons emitted by solar activity (usually E < 1 GeV). The energetic cosmic particles produce complex cascades of secondary particles via nuclear reactions with the nitrogen and oxygen nuclei composing the majority of the Earth's atmosphere, these secondary particles continue on deeper into the atmosphere creating tertiary particle cascades and so on (as shown in fig. I-17). At terrestrial altitudes (as opposed to flight or satellite altitudes) less than 1% of the primary flux reaches sea-level where the particle flux is composed predominantly of electrons, muons, neutrons, and a smaller fraction of protons (there are an equal number of protons as neutrons produced in the upper atmosphere, but as the protons are charged, they lose energy through coulombic interactions with atmospheric electrons and, thus, are removed from the sea-level terrestrial flux) and pions (can actually cause significant SEE [51] but occur at significantly lower probabilities) as shown in fig. I-18a. Due to the relatively high LET of the resulting secondary particles from neutron reactions and their abundance as compared with most of the other particle types, neutrons are considered the dominant source of SEE from cosmic sources. Before looking at the terrestrial neutron spectrum, it is helpful to consider how neutrons interact with matter as a function of their energy. To this

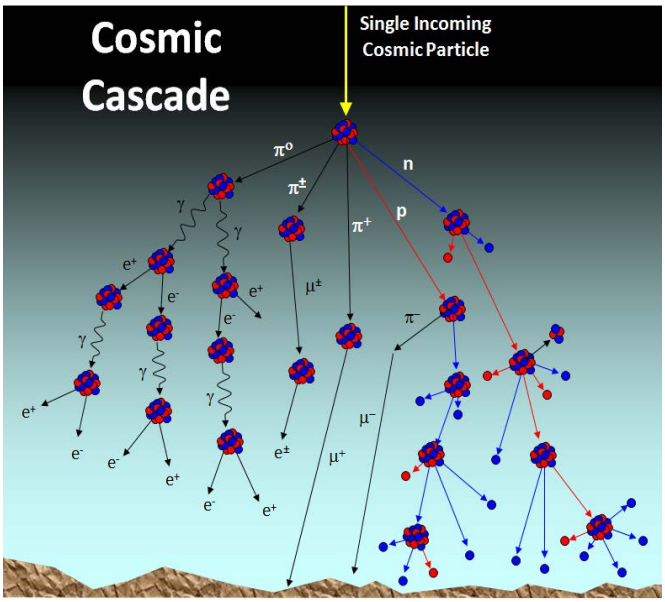
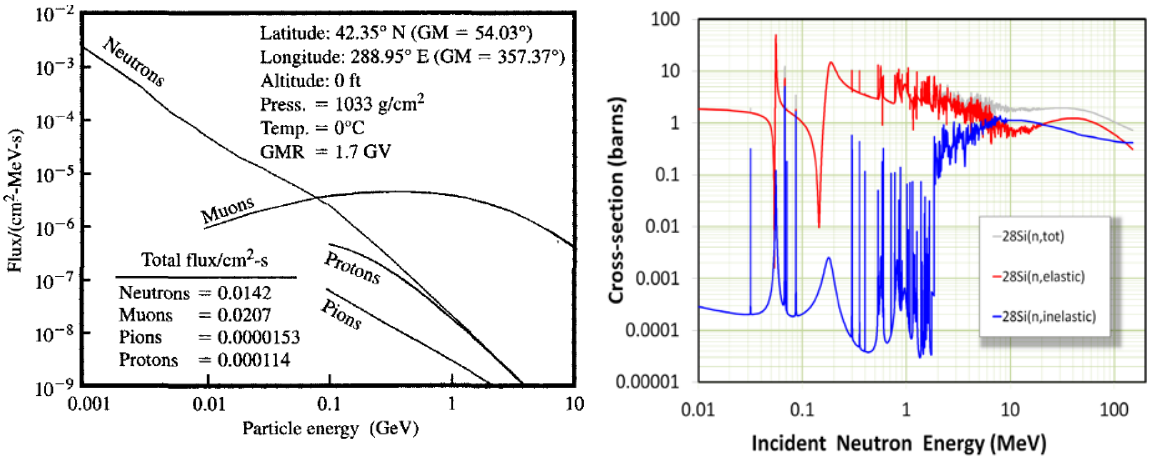


Fig. I-17 Energetic cosmic ray particle (usually a proton with E >> GeV) incident on the Earth's atmosphere and the resultant particle cascade that ensues. Note that the terrestrial flux is composed of numerous 2nd, 3rd, etc. order particles.



a.) b.)

Fig. I-18 a.) Terrestrial particle flux from cosmic sources as a function of particle energy. Note that neutrons and protons tend to induce larger charge bursts in silicon as compared with electrons and muons – and since they are abundant, neutrons are the leading source of SEEs from cosmic sources [52]. b.) The total neutron (light gray) cross-section as a function of neutron energy. At neutron energies above 2 MeV, the inelastic reactions (blue) become a significant contributor. Neutron cross-section data obtained from [53].

end, let us consider energetic terrestrial neutrons incident on a volume of silicon representing a semiconductor device. Since the active device layers are composed of silicon, we limit our discussion to interactions with silicon (neutrons obviously interact with all chip materials, and if those reactions are within a few micrometers of the active silicon device layers, they can contribute to the production of soft errors [54]). From the neutron cross-sections curves in silicon illustrated in fig. I-18b, we can see that elastic reactions (red curve) dominate the interactions up to neutron energies of about 2MeV. At 2MeV the neutron can be treated as non-relativistic (with less than 0.3% error). Elastic scattering is analogous to billiard ball physics - both energy and momentum conserved. The kinematics of a collision between the incident neutron and the target ^{28}Si nucleus is described by the following relationship:

$$\frac{1}{2}m_n v_n^2 = \frac{1}{2}m_n v'^2_n + \frac{1}{2}M_{Si}V_{Si}^2 \qquad \text{(eq. I-3)}$$

The kinetic energy of the incident neutron prior to the collision is on the left-hand side where m_n is the neutron mass and v_n is its velocity. The neutron's kinetic energy after the collision is the first term on the right-hand side where v'_n is the neutron's velocity after the collision and the second term on the right-hand side is the energy imparted to the ^{28}Si nucleus after the collision, where M_{si} and v_{Si} are the mass of the silicon nucleus and its velocity, respectively. For this discussion we just want to determine the maximum amount of energy that can be transferred from the incident neutron to the silicon nucleus during the collision process, because this will determine the minimum neutron energy required to produce a silicon recoil from the elastic collision. The maximum energy will be transferred for head-on collisions where the neutron is scattered such that it reverses its direction (scattered 180°). This assumption also greatly

simplifies the calculation since the cosine term is eliminated. Solving for the neutron velocity after the collision we get:

$$v'_n = -\frac{M_{Si}-m_n}{M_{Si}+m_n} v_n \qquad \text{(eq. I-4)}$$

This result can be substituted to calculate the energy imparted to the silicon nucleus during the head-on elastic collision as follows:

$$E_{kSi} = \frac{1}{2} M_{Si} V_{Si}^2 = \frac{1}{2} m_n v_n^2 - \frac{1}{2} m_n v'^2_n = \frac{4\, m_n\, M_{Si}}{(M_{Si}+m_n)^2} E_n \qquad \text{(eq. I-5)}$$

Where E_n is the initial neutron energy and E_{kSi} is the maximum energy imparted to the silicon nucleus. Plugging in the mass of a neutron and a ^{28}Si nucleus, we get an $E_{kSi} \sim 0.13E_n$. Assuming that the energy of formation of a silicon vacancy is about 4 eV [55], an incident neutron must have, at least, 30 eV to be able to generate a silicon recoil (to impart an energy of 4 eV to the silicon nucleus). Below 30 eV the elastic collision will have insufficient energy to create a recoil, and the net result will be that the silicon nucleus vibrates (creating phonons in the silicon lattice). As neutron energy increases above 30 eV, silicon nuclear recoils will be generated with increasing kinetic energy (as in the image at the top of fig. I-19). Any excess kinetic energy imparted to the silicon nucleus will be lost through electronic processes (creation of electron-hole pairs as with any ion in a material) as it moves away from its original lattice site. As the silicon recoil energy increases, so does the probability that it will be capable of inducing enough charge to upset a device. If we assume that at least 1 fC of generated charge is required to upset a device node (Q_{crit} is in this range for 45nm SRAM), we can use the LET of the silicon recoil in silicon to determine the threshold energy above which recoils have enough energy to cause SEEs. Using the SRIM software referenced previously, it can be shown that a silicon recoil with 60 keV of kinetic energy can generate 1 fQ of charge (at this energy the recoil only travels a few nanometers and thus the recoil would have to occur within a junction or depletion zone for the charge to be collected). Each silicon recoil creates two defects, leaving a vacancy

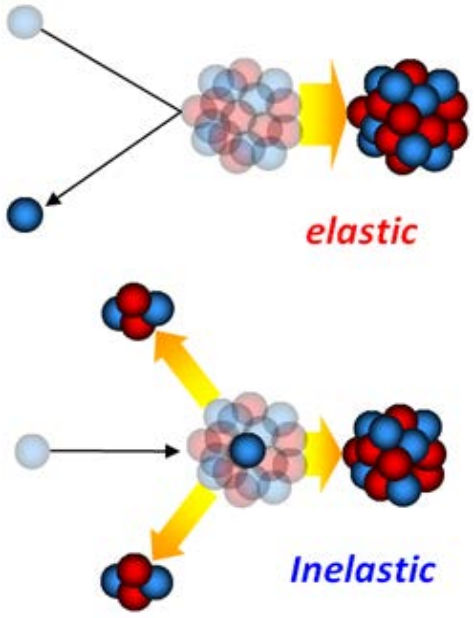


Fig. I-19 Diagram of the two major reaction processes between neutrons and silicon nuclei. The onset of elastic recoils (top) occurs at energies ≥ 30 eV but need > 60keV neutrons to create events > 1fC. Inelastic reactions (bottom) start to contribute significantly at neutron energies above 2MeV.

where the recoil originated and an interstitial where the recoil stops. In addition, if energetic enough, the recoil, itself, induces additional defect formation if it collides with other nuclei. All of these defects can be electrically active and potentially contribute to dose-related marginalities. For majority-carrier devices like CMOS field-effect transistors, the accumulation of neutron-dose damage in terrestrial and avionic environments is far too low to create significant damage-related marginalities since the flux is many orders of magnitude too low.

Table I-1 Neutron wavelengths as a function of neutron energy (1 fm = 10^{-15} m). As energy increases the wavelength of the neutron decreases and the types of interactions it has with nuclei change from elastic to inelastic reactions.

Neutron Energy (eV)	Notes	Wavelength (fm)	Scale
2.5E-02	Thermal energy, elastic reactions	18,090	Interatomic spacing
3.0E+01	Creation of recoils possible	5222	
6.0E+04	Onset of recoils capable of causing upsets	117	
2.0E+06	Onset of Inelastic reactions	20	
1.5E+07	Compound nucleus formation	7.4	Silicon nucleus
6.0E+07	Direct reactions (spallation)	3.6	Individual nucleons
1.0E+08	Direct reactions (spallation)	2.7	Individual nucleons
1.0E+09	Direct reactions (spallation)	0.7	Individual nucleons

Inelastic reactions (shown in the cartoon at the bottom of fig. I-19) occur when some of the neutron energy creates nuclear excitation. Above 2 MeV the wavelength of the neutron approaches nuclear dimensions (at 16 MeV it is the same as the estimated diameter of the ^{28}Si nucleus as illustrated in table I-1), and begins to be able to interact with multiple nucleons (protons and neutrons in the nucleus). In this energy range, inelastic interactions occur when the neutron interacts with multiple nucleons, creating a "new" nucleus that is in a highly excited state. This new excited nucleus, or compound nucleus (made up of the incident particle and the original nucleons), decays a short time later (10^{-16} – 10^{-18} seconds) by casting off nucleons and larger nuclear fragments with significant kinetic energies (in the MeV range). There is a threshold energy for inelastic reactions and, as neutron energy is increased, the number of reaction pathways or "channels" (unique reactions) increases. When the neutron energy exceeds 20 MeV, the neutron wavelength is smaller than the nuclear diameter and actually approaches that of individual nucleons. At these energies and above, the nucleus, as a whole, becomes exceedingly less "visible", and the incident neutron interacts by direct collisions with single nucleons or valence nucleons – these direct reactions occur over very short timescales on the order of 10^{-22} seconds [56]. These "knock-out" reactions become dominant at higher energies.

The terrestrial high-energy (loosely defined here as E > 1 MeV for historical reasons) neutron flux, estimated at ~ 20 n/cm^2-hr [57], is capable of producing secondary products that are highly ionizing and very disruptive to electronic devices. A plot of the terrestrial neutron flux measured at various locations is shown in figure I-20. At sea-level, neutron energies range from thermal to GeV. The three peaks coincide with the different nuclear reaction types that produced them. The first peak at low energy represents the population of neutrons that have been thermalized by numerous elastic interactions with oxygen and nitrogen nuclei in the atmosphere – thermal neutrons can be captured by nuclei with the usual result of a

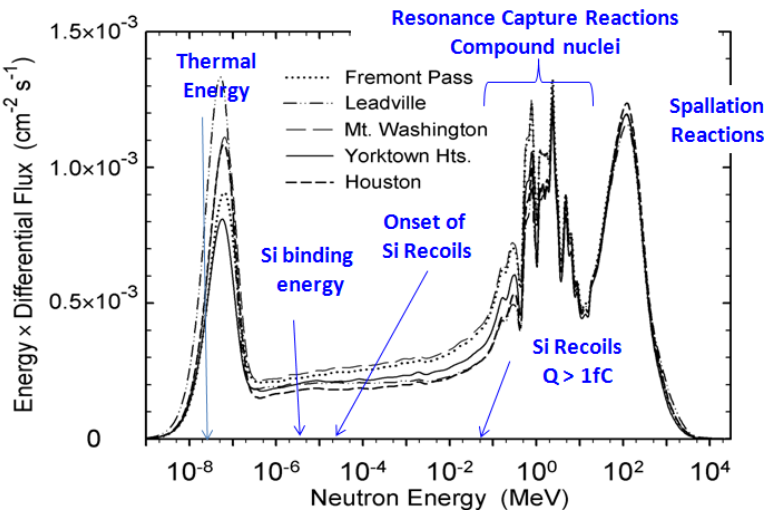


Fig. I-20 Measured neutron spectra for five geographical sites [58]. Each spectrum has been scaled for sea-level at the cutoff of New York City and solar modulation for Nov. 2002 (20% above minimum).

gamma photon emission. If they are not captured, then their interaction is predominantly elastic. The peak in the middle is due to inelastic compound and resonance reactions (reactions at specific energies where the incoming particle gets captured into a bound state of the nucleus) with the actual resonance energies defined largely by quantum nuclear states in the nitrogen and oxygen nuclei. The high-energy peak to the right is predominantly the result of direct or knock-out reactions where nucleons are emitted with a large share of the kinetic energy of the incident neutron. The thermal peak does not usually pose a problem for electronic devices since thermal neutrons do not create reactions with ionizing secondaries (gamma photons can create ionization, but the interaction cross-section within active silicon is rather low). However, as we shall see shortly, thermal neutrons can cause SEEs when they interact with ^{10}B, which is prevalent in semiconductor devices. The other two higher energy peaks are the leading source of soft errors since both neutron-induced silicon recoils and reaction products create hundreds of fC of charge (as we saw in the LET curves shown in fig. I-5), especially considering that most advanced CMOS devices operate with a Q_{crit} in the range of 0.5 - 5 fC. For neutron reactions occurring in close proximity to active device silicon, the likelihood of a SEE being generated is quite high. Additionally, certain SEEs (described earlier) like MBU and SEL cannot, generally, be induced by alpha particles because the LET threshold for these types of events is above the maximum LET of alpha particles. MBU and SEL are typically due to higher energy neutron events.

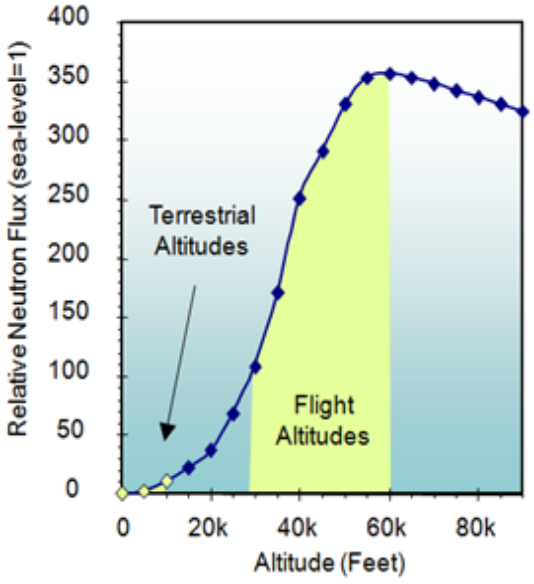


Fig. I-21 Neutron flux variation with altitude for 1 – 10 MeV neutrons. Over terrestrial altitude the neutron flux varies about 20x from sea-level to 10,000 feet (3000m). Adopted from [59].

As mentioned previously, the Earth's atmosphere and magnetic field convert and divert the energetic cosmic rays (mostly protons) incident on the upper

atmosphere. Through nuclear interactions, the predominant proton flux in space is converted into the neutrons and other particles in the terrestrial environment at sea-level. The neutron flux as a function of altitude is shown in fig. I-21. Over terrestrial altitudes (those that can be reached without aircraft) from sea-level to 10k feet (~ 3000m) the neutron flux increases nearly 20x (neutrons are absorbed more quickly deeper in the atmosphere, hence the population of neutrons increases with altitude at terrestrial altitudes). At flight altitudes 28k-60kft (10-20km), the neutron flux is many hundreds-of-times higher than the sea-level flux. However, as altitude increases further, the complex balance between shielding primaries and creating and absorbing more secondaries changes. The interplay of these complex processes results in the neutron flux having a peak at ~ 60k feet, called "Pfotzer maximum" (of course as altitude increases, the proton flux increases. At about 100k feet the proton and neutron flux are equal). The bottom line is that in the terrestrial environment, altitude has the biggest effect on the neutron flux and, consequently, on the neutron-induced upset rates.

The Earth's magnetic field, emanating from the poles, causes charged cosmic particles (mostly protons) to become trapped within specific regions, leading to areas with high proton densities and others with low proton densities (the so-called Van Allen belts) - this effect translates to variations in the terrestrial neutron flux as a function of geographical location as illustrated in fig. I-22. The concentration of particle fluxes incident on the Earth vary over time and produce ionization in the upper atmosphere (a fraction of this flux also contributes to sea-level radiation), as evidenced by the dramatic Auroras that can be seen at higher Northern and Southern latitudes. The Auroras are usually limited to high latitudes because magnetic field lines enter the North and South poles nearly perpendicularly to the Earth's surface, thus the magnetic shielding effect, or geomagnetic rigidity, is minimized at the poles, concentrating the ionization in these areas. Thus the particle flux at the poles is significantly higher than that at the equatorial regions where the magnetic field lines are parallel to the Earth's surface and more easily deflect or trap the protons. At sea-level, the neutron flux increases by ~ 2x, going from equatorial to polar latitudes. Note that the effect is magnified at avionics altitudes to about 5x. Changes in longitude do not, in general, affect the neutron flux since the geomagnetic rigidity does not vary significantly

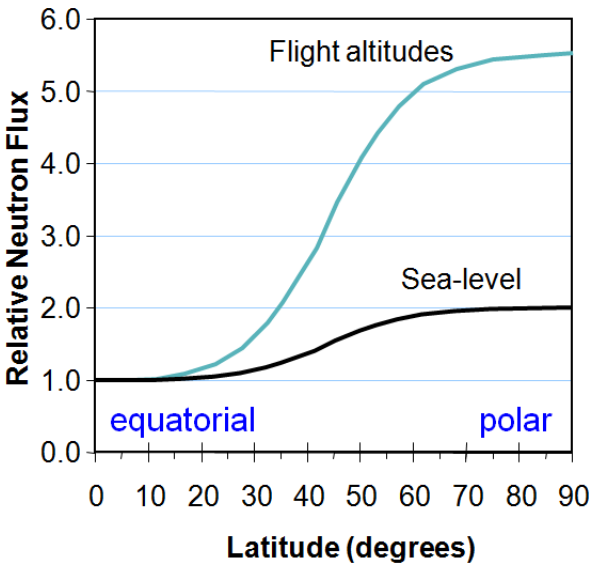


Fig. I-22 Neutron flux variation with latitude. Since geomagnetic rigidity or "shielding" is minimized at the poles, flux is maximized for high latitudes, particularly at flight altitudes. After [58].

for different longitudes at any given latitude. The only exception to this is the South Atlantic Anomaly. The effect occurs because the Earth's magnetic field is unusually weak in this region, departing from the rotational symmetry (about the polar axis) of the field. The anomaly's reduced magnetic field strength increases the flux of cosmic ray protons, exposing orbiting satellites to unusually high levels of radiation. Many satellites whose orbits intersect the anomaly have their on-board electronics shut down during their passage through this region to limit dose and SEE effects. As might be expected, at sea-level the neutron flux in this region is somewhat higher than normal, as a result of the increased proton impingement.

In addition to the filtering effects of the Earth's atmosphere and magnetic field, there is an external variable that also modulates the flux of terrestrial neutrons – the solar activity cycle. This is an 11-year cycle where the sun shifts gradually between active and quiescent modes. High activity is usually accompanied by an increase in the number of sun spots and a corresponding increase in solar flare and coronal mass ejection activity. A solar flare is a large energy release via the ejection of extremely high-temperature "plasma clouds" of electrons, ions, and atoms into space during magnetic disturbances around sun-spots on the sun's surface. Solar flares occur when the magnetic field twists and accelerates plasma so that bursts of x-rays, light, and radio-waves are emitted. The x-rays and ultraviolet radiation emitted by solar flares can affect Earth's ionosphere and disrupt long-range radio communications several minutes after they are emitted from the surface of the sun. This is actually a boon for Earth-based systems - since the upper atmosphere becomes more highly ionized by the electromagnetic radiation, creating electrostatic repulsion that actually increases the deflection of solar particles (arriving several days later). So as one might expect, since this "shielding effect" increases during high solar activity, the number of protons that can penetrate deeper into the atmosphere is reduced, thus decreasing the terrestrial neutron flux. In extreme cases the solar magnetic fields become so warped that they "snap" and break, freeing large amounts of surface plasma that explodes into space as a coronal mass ejection. The terrestrial neutron flux measured in Moscow over seven years of the solar cycle is shown in fig. I-23. In some cases the dynamics of a solar flare or coronal mass ejection occurs over a shorter time-

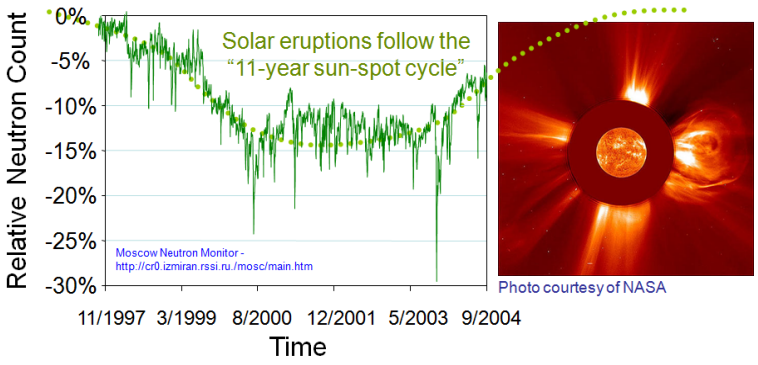


Fig I-23 Plot of terrestrial neutron activity as a function of time (left). Photograph (right) of solar flare activity [courtesy NASA homepage]. Solar flares and coronal mass ejections send huge bursts of energetic particles into space. Note that while punctuated by occasional "big" events, the neutron flux only varies by ~ 15% during the solar cycle.

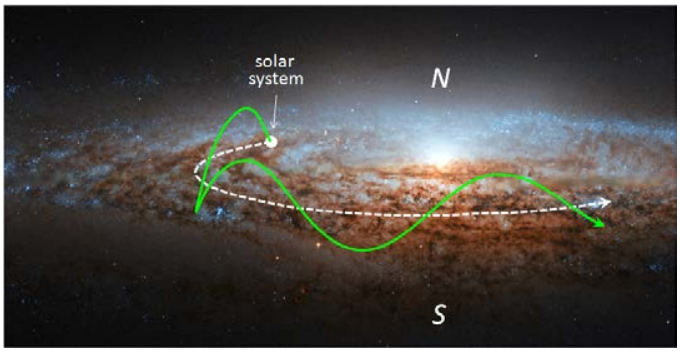
scale so that the ionosphere does not have time to become fully ionized. During such events the ionosphere's repulsive effect is reduced and the solar particle flux impinging the Earth is not turned away. As a result, for these more sudden events the neutron flux actually increases, but these are generally rare isolated events. In any case the impact of solar events on the terrestrial neutron flux is tame in comparison to altitude and latitude effects typically only changing the neutron flux by 15%.

To summarize there are three primary factors that define the neutron flux at any terrestrial location and time. The dominant factor is altitude, effecting a variation in the neutron flux of nearly 20x going from sea-level to 10,000ft (3000m). Latitude, or more specifically, geomagnetic rigidity is another factor that can modulate the neutron flux, increasing it by about 2x (at sea-level) going from equatorial to polar latitudes. The final variable defining the neutron flux is the solar activity cycle - the approximately 11-year variation in the frequency of sunspots and solar flares. Solar effects usually account for < 2x variations in neutron flux, but during very sudden and intense events, the neutron flux can change significantly. For SEE effects from neutrons in commercial electronics, we generally simplify extrapolating and reporting of SEE by de-rating for changes from a standard location. In the JEDEC JESD89A test method [60], the "standard location" has been specified as N.Y.C at sea-level. Every other location can be adequately modeled using a factor generated by the model M.A. Shea and D.F. Smart using the International Geomagnetic Reference Field for 1995 [61]. Depending on the actual use location, one can de-rate for altitude, geographical position, and solar activity. As we have seen, the terrestrial high-energy (E_n > 1 MeV) neutron flux, estimated at ~ 20 n/cm^2-hr, is capable of producing secondary products that are highly ionizing and very disruptive to electronic devices. It should be noted that while the JESD89A specifies a neutron flux 13 n/cm^2-hr and E_n > 10MeV, the author anticipates changes in the specification next time it is updated. There is sufficient evidence now that most advanced technologies with low Q_{crit} are sensitive to neutrons in the 1 – 10 MeV range. Unfortunately, the high-energy neutron flux is not easily shielded [62] (large installations such as main-frames can be located in the concrete basements in office buildings with the concrete and earth foundations offering some level of shielding), and, ultimately, complex devices and systems must be designed to be resilient against such upsets with built-in error checking and correcting.

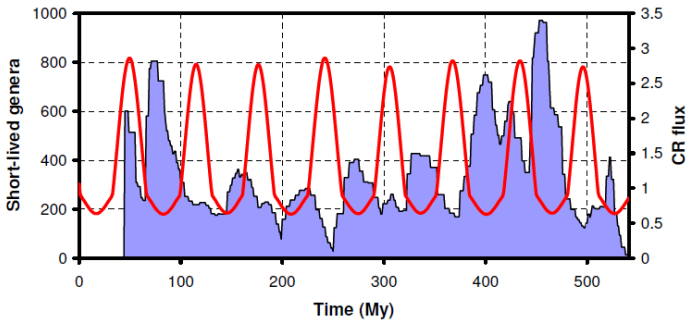
Tangent: Death and Cosmic Rays

Stories of cataclysms born of cosmic origin are as old as human history and religion. Two researchers from the University of Kansas, M. Medvedev and A. Melott claim to have linked mass extinctions in the marine fossil record to the Earth's position in the heavens [63]. Our solar system, as it orbits around the center of the galaxy every 225-250 million years, also oscillates vertically above and below the galactic plane with a period of 62 (±3) million years (as illustrated in fig. I-24a). Apparently, the magnitude of the cosmic ray flux in the galactic halo/wind/termination shock (due to the galactic

motion) is worst when the solar system is at the peak (northern-most point) of its vertical oscillation. Conversely, when the solar system is in or below the galactic plane, it is largely shielded from the harsh cosmic rays by the galactic magnetic fields. Through statistical models of the fossil record and cosmological modeling, the authors maintain that each cosmic ray peak is followed by a drop in biodiversity (see fig. I-24b). Would the approximately 5x change in the terrestrial cosmic ray flux really induce mass extinctions by directly causing genetic damage in organisms? Perhaps the changes in cosmic ray intensity impact global climate and induce drastic changes in the environment that ultimately cause the cyclical boom and bust biodiversity periods. In any case, it is an interesting hypothesis.



a.)



b.)

Fig. I-24 a.) Image of galaxy showing the solar system's vertical oscillation (exaggerated for effect) around the galactic plane [adapted NASA image of a galaxy (not our own!)], and b.) plot of the genera variation and extragalactic cosmic ray flux at the Earth as a function of time. Note that most of the minima in biodiversity coincide with cosmic ray maxima. From [63].

A group in Eastern Europe has studied the effects of cosmic ray variation on death rates from various ailments [64], and concluded that "neutron activity…is linked with the timing of human death...the role of this factor can be seen in the results of timing of cardiac, stroke, and oncology deaths." Members of this same study group also tracked the number and days of discharges from automatic implantable cardioverter defibrillators (ICD) in patients with cardiac arrhythmias and correlated this to neutron activity data and found that "*the mean neutron activity on days of ICD discharges in response to ventricular disturbances was significantly higher than the mean level. Whether this relation is a direct result of low geomagnetic activity or due to an independent role of neutrons in the pathogenesis and timing of cardiac arrhythmias is unknown*" [65]. The author notes that the mean neutron flux during "*cardiac events*" in the report was only 2% above the mean neutron flux between when no cardiac events occurred - it is hard to see it implicated in any mechanism. That the higher neutron flux might actually cause a higher level of soft errors [66] in the ICD leading to false activations does not seem to have been considered. In other words, defibrillator/pacing events were assumed to be based only on actual cardiac arrhythmias events only vs. electronic glitches induced by the higher intensity neutron events.

b. Alpha Particles

Alpha particles emitted by trace uranium and thorium impurities in packaging materials were shown to be the dominant cause of soft errors in dynamic random-access memory (DRAM) devices in the late 1970s [67]. The alpha particle is composed of a tightly bound core of two neutrons and two protons - a doubly-ionized helium atom that is emitted from the disintegration of unstable isotopes. Radioactive decay is a process by which unstable nuclei lose energy by emitting particles and/or electromagnetic radiation. This process occurs spontaneously without any external interaction. The decaying atom is called the parent, and the decay product is called the daughter. Typically the decay results in the emission of an energetic electron (either a beta-minus particle or the ejection of an inner core electron by an internal conversion event [68]) or the emission of an alpha-particle. The kinetic energy of the emitted beta-minus is a smooth function of energy ranging from 0 to a few MeV with the maximum energy determined by the type of nucleus from which it was emitted. In contrast, a conversion electron, having been "knocked" from one of the core atomic orbitals, is emitted with one of a few discrete kinetic energy values (usually in the tens or hundreds of keV). Alpha particles are emitted from decaying nuclei with unique and discrete kinetic energies typically somewhere between 4 to 9 MeV. Additionally, after the radioactive decay event, a gamma-photon is emitted when the newly formed daughter nucleus "relaxes" to a lower energy state [69]. The resultant particle and gamma photon emissions have energies that are unique for a specific decay pathway. Unstable isotopes typically undergo a succession of decays until the nucleus has reached a stable state. The radioactive decay of a large population of identical atoms (nuclides) will occur at a constant rate, and the quantity of the parent isotope N can be determined as a function of time:

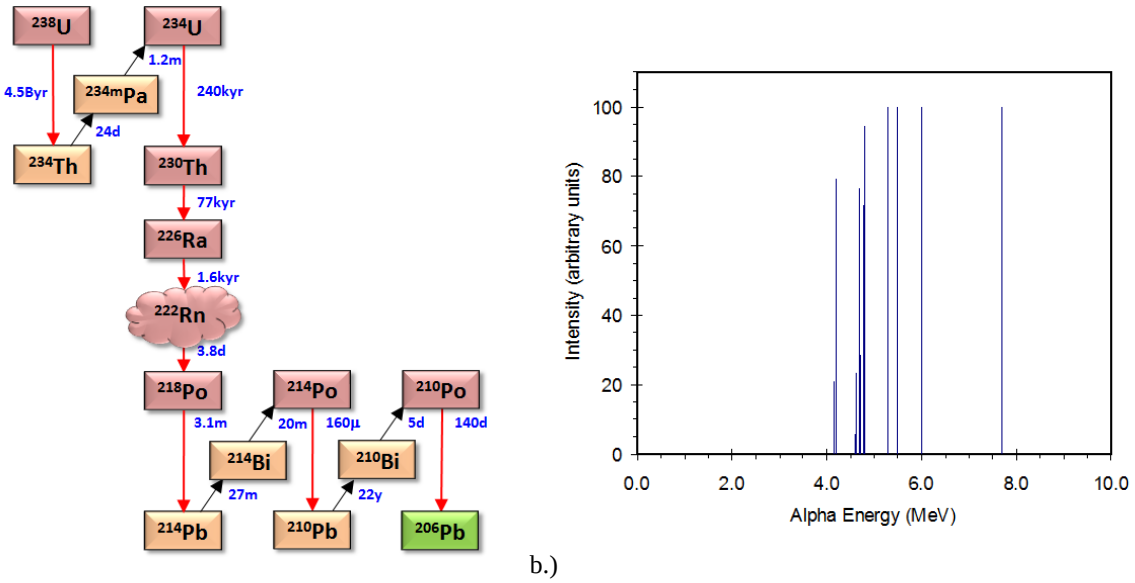
$$N(t) = N_o e^{-\lambda t} \qquad \text{(eq. I-6)}$$

Where N_0 is the initial time-zero concentration (assuming all the parents are created at the same time or within an interval much shorter than the half-life of the parent), $N(t)$ is the concentration at any specific time t, and λ is the decay constant. λ is unique for each nuclide and can be obtained using the half-life, $t_{1/2}$, the time it takes for the initial parent concentration to decay to one-half its original amount as follows:

$$\frac{N(t_{1/2})}{N_o} = 0.5 = e^{-\lambda t_{1/2}} \qquad \text{(eq. I-7)}$$

$$\text{or} \qquad \lambda = \frac{ln2}{t_{1/2}} \qquad \text{and} \qquad A(t) = -\frac{dN(t)}{dt} = \lambda N(t) \qquad \text{(eqs. I-8, I-9)}$$

Where *A(t)* is the activity, a measure of the rate of radioactive decay in a material, defined as the number of disintegrations or decays per time interval. In SI units activity is typically reported in Becquerel (1 Bq = one disintegration per second). Since the activity and half-life are inversely related (eq. I-8), the shorter the half-life of a specific isotope the higher its activity. There are many naturally occurring radioisotopes but the most common source of alpha particles are from the naturally occurring $^{238/235}$U and ^{232}Th due to their relatively high abundance in terrestrial materials and their higher activity. As illustrated in fig. I-25a, a population of ^{238}U atoms in equilibrium emits eight different alpha particles at well-defined energies ranging from 4.149-7.687 MeV (^{232}Th emits six alpha particles from 3.950-8.784 MeV). Note that for



a.) b.)

Fig. I-25 a.) The radioactive decay chain for ^{238}U (vertical red arrows denote decays that produce an alpha particle and black upward arrows denote decays that produce beat particles), and b.) the emission spectrum from an infinitely thin sample (surface emission) of the material.

every transition from the original parent isotope to subsequent daughter transitions, either an alpha particle (vertical red arrow) or a beta particle (diagonal black arrow) is emitted. Each decay event is also accompanied by a unique gamma photon emission. From the terrestrial SEE view-point, the alpha particle is far more damaging than either the beta particle or the gamma photon. The decay from parent to daughter and then to subsequent daughter continues until the nucleus reaches a stable state which in this case is ^{206}Pb. The measured alpha particle spectrum from an infinitely thin of ^{238}U is shown in fig. I-25b. Since 100% of the flux is emitted from the surface (the sample has zero thickness) the alpha particle energies are a discrete set of lines, most of them from a single specific decay (there are a few decays that can occur at several energies). Of course, any real materials with alpha emitting impurities will have a finite thickness and most likely some distribution of impurities through its volume. The measured energy spectrum of alpha particles emitted from the surface of a thick ^{232}Th foil and a simulated spectrum expected at the silicon substrate interface of a 50:50 mix of ^{238}U:^{232}Th impurities distributed throughout

the back-end of line (BEOL) and packaging materials of a device are shown in fig. I-26. The lack of signal below 2 MeV is the cut-off energy in the discriminator of the detector (used to limit counts due to other radiations) – a smooth function down to 0 MeV is expected. These broadened energy spectra are characteristic of the alpha particle spectra in packaged ICs as the narrow discrete emission energy "lines" are "smeared-out" since alpha emitters are generally uniformly distributed thought the layers. Alpha particles emitted close to the silicon substrate interface will lose only a small energy due to traversing a small amount of material while those further away will lose more energy, so the same emission energy is shifted to lower energies with the magnitude of the shift determined by the physical distance travelled, the materials travelled through, and the initial energy of the alpha particle. From sources such as package mold compound or underfill which are essentially "thick" sources, a broadened alpha spectrum is expected. A notable exception of this broadening occurs if the alpha source is confined to a thin layer so that all the alpha particle emission essentially occurs at or very near the surface. One example of a thin source would be the residue of alpha emitting impurities left after a wet-etch with certain batches of phosphoric acid [70, 71]. Alpha particles emitted from the thin residue layer would not lose energy since the layer was much thinner than their range and thus the energy spectrum would be a sharp set of spectral lines. Another example of surface emission from thin layers is flip-chip solder bumps. It has been reported that the primary alpha emitting impurity (^{210}Po) in standard lead-based solders segregates to the surface of solder bumps [72]. This effect would also lead to a sharp spectrum. Comprehending the shape of the energy spectrum of the alpha particles incident on a silicon device is crucial for accurately determining the SEE induced by this source. Indeed, due to its pronounced non-linear Bragg peak (prominent peak where the LET is the highest) the probability that an alpha particle causes a soft error is based largely on its energy and on its trajectory. The wrong assumption about the alpha energy spectrum can lead to significant errors in estimating the SER from accelerated experiments. Since virtually all semiconductor materials are highly purified, the alpha emitting impurities will generally NOT be in secular equilibrium (a state in which the quantity of a radioactive daughter isotopes remains constant because their production rate is equal to their decay rate –

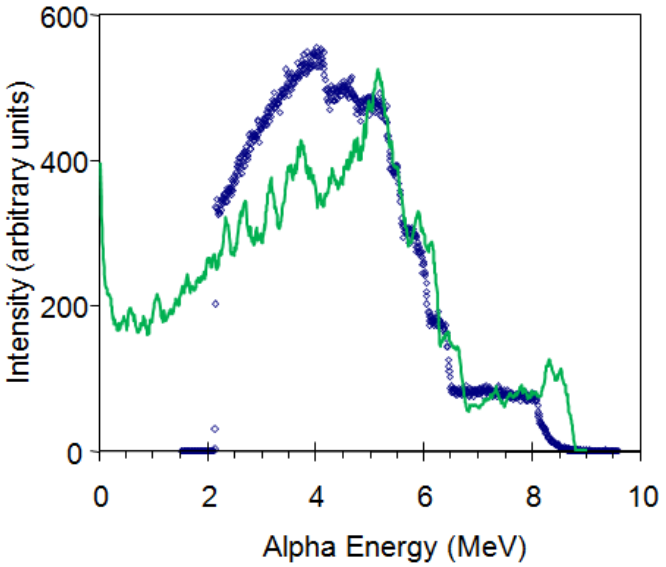


Fig. I-26 Measured alpha particle spectrum obtained from a thick foil sample of ^{232}Th (blue) and simulated spectrum (green) from ^{238}U/^{232}Th impurities in a packaged part.

this occurs when the parent isotope has a longer half-life than the daughter isotopes as with ^{238}U and ^{232}Th. Said another way, in secular equilibrium, each decay of a parent is accompanied by the decay of all its [shorter-lived] daughter products). A knowledge of the possible sources of impurities and long duration alpha counting must be used together to determine the alpha emission since the exact nature of parent/daughter distributions is seldom known. In other words, a low concentration of uranium and thorium impurities is a necessary requirement for low alpha emission but NOT sufficient. In non-equilibrium situations higher activity daughters may be present that greatly increase the alpha emission rate. This situation was highlighted during investigations into eutectic lead solders (for flip-chip bumps) in which all radioactive impurities had been eliminated except the radioactive ^{210}Pb that was chemically inseparable from the ^{206,208}Pb. Since ^{210}Pb does not emit an alpha particle when it decays, initial alpha counting measurements revealed the solder to be emitting alpha particles at extremely low levels. With the relatively short half-life of ^{210}Pb a re-growth of the alpha emitter ^{210}Po (from the decay of ^{210}Pb => ^{210}Bi => ^{210}Po) occurred and within a few months the solder alpha emission was 10x higher than initial measurements indicated. In this case impurity analysis would not have revealed any issue either since a problem can arise when the ^{210}Pb levels are far below the detection limits of any spectroscopic method. In fact a ^{210}Pb/^{208}Pb impurity ratio as low as 10^{-16} is still high enough to cause alpha emission problems! Thus for this special case even the most modern spectroscopic techniques would be hard-pressed to detect the problematic levels of ^{210}Pb (10,000 times lower than one part in a trillion). For advanced electronics most manufacturers have moved to the use of low or ultra low alpha materials. For a material be classified as ultra low alpha (ULA implies an emission at, or below, 0.002 α/hr-cm^2) the $^{238/235}$U and ^{232}Th impurity content must be below about one part per ten billion. Again, this is NOT a guarantee that the material will meet the ULA emission specification since higher activity daughters may have significant in-growth from the time the material is characterized to the time it is actually used in the final product, but these levels are required to ensure that the material has the potential of reaching ULA standards. To ensure that alpha

Table I-2. Measured alpha particle emission (α/hr-cm^2) from various packaging materials. Emission from pilot and full-flow wafers is ~10x lower that ULA packaging materials.

Sample	Type	Counts/hr-cm^2		
		CDL95	MDA95	Mean
Vendor A AgSnCu solder	FC Solder	0.00048	0.00098	0.01002
Vendor A AgSnCu solder	FC Solder	0.00049	0.00100	0.00660
Vendor B solder substrates	FC Solder	0.00048	0.00099	0.00276
Vendor C substrate solder	FC Solder	0.00048	0.00099	0.00048
Vendor C substrate solder	FC Solder	0.00049	0.00099	0.00668
Vendor B AgSn solder	FC Solder	0.00056	0.00115	0.00527
Vendor C lead free solder	FC Solder	0.00047	0.00096	0.00145
Vendor C High Lead solder	FC Solder	0.00048	0.00097	0.00237
Vendor D PbSn solder	FC Solder	0.00044	0.00090	0.01337
Vendor D LAL solder	FC Solder	0.00051	0.00105	0.00293
Vendor D PbSn solder	FC Solder	0.00047	0.00096	0.01712
DBUMP PbSn solder	FC Solder	0.00049	0.00101	0.00162
Standard solder A	FC Solder	0.00068	0.00139	6.26368
Vendor A underfill	FC Underfill	0.00015	0.00031	0.00069
Vendor B1 Underfill	FC Underfill	0.00015	0.00030	0.00122
Vendor B2 (silica) underfill	FC Underfill	0.00047	0.00096	0.00188
Vendor B3 (alumina) underfill	FC Underfill	0.00047	0.00097	0.00106
Vendor C Underfill	FC Underfill	0.00047	0.00096	0.00003
Vendor C2 Underfill	FC Underfill	0.00046	0.00093	0.00421
Vendor A Mold compound	WB Mold	0.00047	0.00095	0.00184
Vendor B1 Mold compound	WB Mold	0.00065	0.00132	0.02053
Vendor B2 Mold compound	WB Mold	0.00065	0.00132	0.02504

emission rate is low enough direct alpha counting techniques must be employed. In lead-based solders this is especially true, where chemical separation will leave known radioactive daughter products, samples should be measured several times over several months to ensure that there is no significant in-growth of alpha-emitting daughter products that would increase the material's alpha particle emission. One of the challenges of advanced technologies is verifying that all materials meet or exceed the ULA specification. Table I-2 shows the emissions from advanced semiconductor packaging materials. CDL95 is the critical detection limit at 95% confidence and represents the "background" of the system. An average (alpha signal) emission lower than CDL95 it is indistinguishable from the system background. The MDA95 is the Minimum Detectable Activity at 95% confidence. If the measured alpha flux (average) is equal to the MDA then one is 95% confident that the sample is emitting an alpha flux of this magnitude and that you have not measured an erroneously high background (related "limit of detection" approach is formalized in the JEDEC JESD-221 standard [73]). Today many vendors provide materials with alpha emission far below 0.002 α/hr-cm^2 so that the neutron SER becomes the dominant component. In flip-chip technologies with thick BEOL and under-bump metal studs (we use 15μm of copper), so many of the alpha particles will be absorbed prior to reaching the active layers that the fraction of failures induced by alpha particles will sometimes be negligible [74]. At this point, further alpha particle emission reduction becomes prohibitively expensive while providing diminishing returns since the SER is dominated by cosmic background radiation.

We have considered where alpha particles come from, their spectra in electronic devices, and the typical levels at which they are found. While alpha particles are directly ionizing, so that no nuclear reaction cross-section is involved, their LET is significantly lower than that of high-energy neutron reaction products. The curve shown in fig. I-27 denoted by solid diamonds in the chart shows the alpha particle dQ/dx (from the LET and converted to units of fC/μm) in silicon as a function of alpha particle energy. By integrating this from the energy at which the particle enters device silicon to the point at which it is stopped or leaves the sensitive volume (this includes the depletion zone and funnel and an additional term for diffusion collection) one can calculate how much charge is deposited by the event (this would

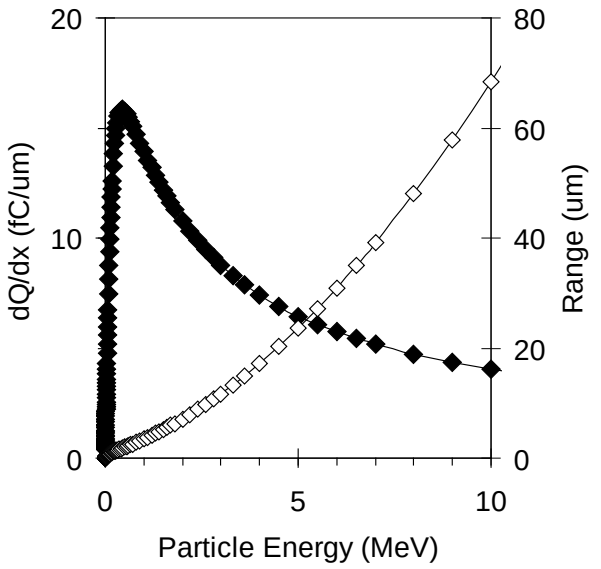


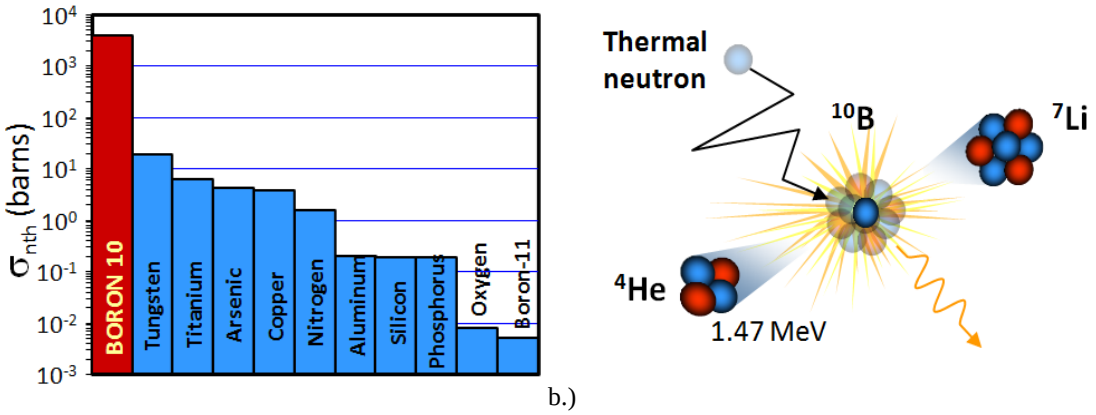
Fig. I-27 LET (solid diamonds) in fC/μm and range (open diamonds) of an alpha particle in silicon as a function of its energy. Generated by the software program SRIM [14].

be the Q_{coll} discussed in a previous discussion). Note the strongly non-linear aspect of the LET curve (the peak in the LET curve is often referred to as the “Bragg peak”) - clearly a dominant portion of the charge is generated near the Bragg peak when the alpha particle has almost been stopped in the material. This characteristic has implications for shielding – in this case, some shielding may be worse than no shielding since if the Bragg peak is deep within the substrate and some shielding is applied to the device which brings this Bragg peak closer to the active junctions, the SER may actually increase. When an alpha particle travels through a material, it loses its kinetic energy predominantly through interactions with the bound electrons of that material and thus leaves a trail of ionization in its wake. The higher the energy of the alpha particle, the farther it travels before being "stopped" by the material. The distance required to stop an alpha particle (its range shown as open diamonds) is both a function of its energy and the properties of the material (primarily the material’s density) in which it is traveling. In silicon the range for a 10 MeV alpha particle is ~ 70 μm. Thus alpha particles from outside the packaged device are clearly not a concern – only alpha particles emitted by the device fabrication and packaging materials need be considered in terms of their ability to induce SEEs.

c. Low Energy Neutrons and Boron

The third significant source of ionizing particles in electronic devices is the secondary radiation induced from the interaction of low-energy "thermal" cosmic-ray neutrons and a common isotope of boron, ^{10}B. As we have discussed in a previous section, higher energy neutrons are themselves the reaction products of even higher energy particles with nuclei in the upper atmosphere. A significant fraction of these neutrons eventually suffer enough reactions as they penetrate, deeper and deeper into the atmosphere, that their energies are reduced to thermal energies (E_k ~ 0.025eV and a mean velocity of v = 2.2 km/s). In this low energy range they "bounce around" suffering elastic collisions with gas nuclei until they are either absorbed by one of the nuclei in a thermal neutron capture reaction or they decay (neutrons bound in stable nuclei are stable but unbound neutrons are unstable, undergoing beta particle decay with a mean lifetime of approximately 15 minutes). The average sea-level flux of terrestrial thermal neutrons at N.Y.C ranges from 2 - 8 /cm^2-hr and averages approximately 4 /cm^2-hr [75].

As with the higher energy neutron flux, this flux scales with geomagnetic Lat/Long and with altitude in a manner similar to the flux of incident energetic cosmic rays. Boron is composed of two isotopes, ^{11}B (80.1% abundance) and ^{10}B (19.9% abundance). At 3838 barns (1 barn = 10^{-24} cm^2 per nucleus), the thermal neutron capture cross-section of ^{10}B is extremely high in comparison to most other isotopes present in silicon semiconductor technologies – The probability that ^{10}B will capture a thermal neutron is three to seven orders-of-magnitude larger than other materials! (fig. I-28a). To make matters worse, unlike most isotopes which simply emit gamma photons after absorbing a neutron, the ^{10}B nucleus becomes unstable and breaks apart, emitting two highly ionizing particles; an alpha particle and a lithium recoil (a fast gamma photon is soon emitted from the ^{7}Li but due to its low cross-section the gamma photon is unlikely to produce a SEE) as shown in the cartoon of fig. I-28b. In this nuclear reaction, often



a.) b.)

Fig. I-28 a.) Thermal neutron cross-section of various nuclei found in semiconductors; b.) unlike most other materials that simply emit gamma photons after neutron capture, ^{10}B disintegrates into highly ionizing secondaries.

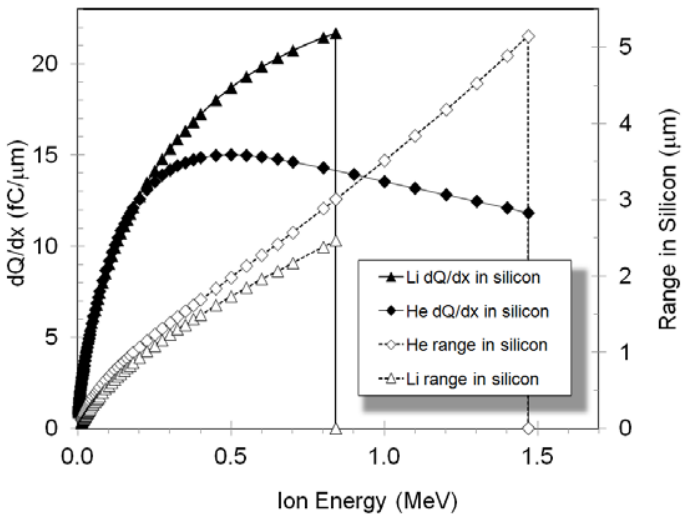


Fig. I-29 The dQ/dx and range of lithium recoil and alpha particle (emitted from $^{10}B(n,\alpha)^{7}Li$ reaction) as a function of particle energy in silicon. The cut-off for secondary is at each respective emission energy. Generated by the software program SRIM [14].

written as $^{10}B(n,\alpha)^{7}Li$, the alpha particle and the lithium recoil are emitted in opposite directions (conserving momentum), thus in most cases one ion will be aimed towards sensitive devices while the other is aimed away for the sensitive areas. The lithium nucleus is emitted with a kinetic energy of 0.84 MeV 94% of the time and 1.014 MeV 6% of the time. The alpha particle is emitted with a kinetic energy of 1.47 MeV. The lithium recoil has a peak charge generation (derived from the LET) of 22 fC/μm while that of the alpha particle is 15 fC/μm as shown in fig. I-29. Both the alpha particle and the lithium recoil are capable of inducing SEEs in electronic devices particularly advanced technologies utilizing lower operating voltages and hence lower Q_{crit}. The range of the alpha particle is about 5 μm and that of the lithium recoil in silicon is ~ 2 μm and calculations have shown that in most cases if emitted beyond ~ 3 μm from the sensitive junction area they have insufficient energy to induce soft errors (this of course depends on Q_{crit}). Thus only ^{10}B in close proximity to the silicon substrate is a reliability hazard.

Boron is used extensively as a p-type dopant and implant species in silicon and is also used in the formation of borophosphosilicate glass (BPSG) dielectric layers. Assuming maximum doping and implant levels encountered in standard silicon processes and typical boron doping of BPSG layers (5-8% boron), the ^{10}B concentration in diffusions and implants (which are predominantly ^{11}B) is thousands of times lower than that of the BPSG layer. For conventional BPSG-based semiconductor processes, BPSG layers are the dominant source of boron reactions and in most cases will be the primary cause of soft errors [76]-[79]. The SEU induced by the activation of ^{10}B in BPSG can be mitigated in several ways as illustrated in fig. I-30. The first and most direct is simply to eliminate BPSG from the process flow – this was done in response to the realization that the overall chip reliability could be improved by 4 – 8x by substituting the BPSG with PSG (phosphosilicate glass without boron doping). The advent of chemical mechanical polishing (CMP) removed the low-temperature reflow advantage of boron doped glasses, since CMP could do the planarization at even lower temperatures. Due to the limited range of the alpha and lithium recoil emitted during the $^{10}B(n,a)^{7}Li$ reaction, typically only the first few levels of BPSG need be replaced with a dielectric free of ^{10}B. In cases where the unique reflow and gettering properties of boron are still

needed (e.g. processes where CMP is not part of the flow) the regular BPSG process can be replaced by processes that use isotopically-enriched ^{11}BPSG without changing the physical or chemical properties of the film and without the requirement for new equipment or processing steps [80]. If the process cannot be changed at all, using a package with low-energy neutron shielding is a straight forward approach (mold compound with BPSG filler instead of normal silica can be used or some of the other high thermal neutron-capture cross-section materials such as cadmium [81]).

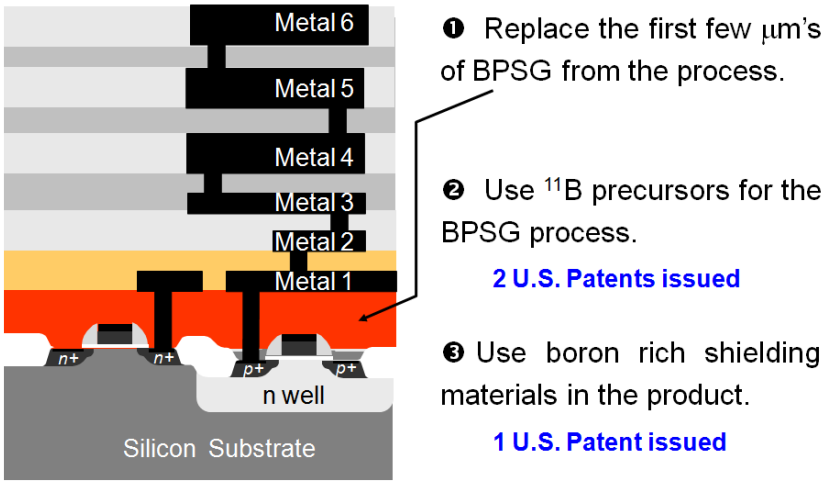


Fig. I-30 Since the alpha particle and lithium recoil in particular have a limited range, the source of ^{10}B need only be kept a few micrometers away from the active devices.

To mitigate the dominant SER threat posed by low-energy neutrons and ^{10}B, or because CMP replaced the need of using boron doping to assist planarization, BPSG has been removed from most advanced technologies. During process development any process steps that might introduce ^{10}B should be checked (implantation is usually not an issue except in cases where compounds of boron are implanted and the masses of the compound with the two boron isotopes are very similar). Recently a foundry process for tungsten plug formation based on a diborane-based carrier gas introduced very high levels of ^{10}B [82]. We investigated two different foundries and found no evidence of ^{10}B incorporation – however the lesson here is that all new processes need to be scrutinized to ensure that high levels of ^{10}B are not incorporated into chip layers. Ultimately, while large reductions in failure rates are possible, either by removing the sources of, or shielding the ^{10}B reaction products and alpha particles, some fraction of the terrestrial high-energy cosmic neutron flux will reach the devices and cause soft errors.

d. Cosmic Ray Muons?

Recently the direct ionization of silicon by protons has been demonstrated to cause upsets in advanced SRAM technologies [83, 84] and can be a dominant effect in space environments with high proton fluxes at lower energies (the cross-section due to proton direct ionization is several orders of magnitude over the saturation proton cross-section). In the terrestrial environment protons are not expected to be a significant problem as the proton flux is much lower at sea-level than that of neutrons. In the terrestrial environment muons are the most prodigious species, with a flux of ~ 60 /cm^2-hr. Muons are similar to electrons (leptons), with a negative electric charge of −1 and a spin of 1/2. The muon has a corresponding antiparticle of opposite charge (+1) but equal mass and spin called the anti-muon or positive muon. Both leptons interact by the weak nuclear force and are very penetrating particles. They are directly ionizing (like very energetic electrons) so each muon creates a charge disturbance when it traverses silicon. As compared with secondary products from neutron reactions, the amount of charge generated per event is much smaller. However since each muon creates charge (while just a small fraction of the neutrons actually react within range of the device silicon) and since the muon flux is 4x larger than the neutron flux, in certain situations, muon soft errors can become a dominant reliability concern. Recent simulation work indicates that the muon-induced error rates for 65, 45, and 40 nm SRAM nodes were negligible [85]. The same group simulated 16, 22, and 32nm SRAMs showed that as the Q_{crit} falls below 0.2 fC the soft error rate increases precipitously. As shown in fig. I-31 showing the soft error as a function of charge generated by the muon event, below 0.2 fC the soft error rate from muons grows to the same order as that generated by neutron events (green shaded region) and then quickly exceeds it. It should be noted that the work was based on bulk silicon planar designs – 3D FinFET type structures with smaller sensitive volumes will likely be less sensitive to muon events assuming the critical charge is similar to that of planar devices.

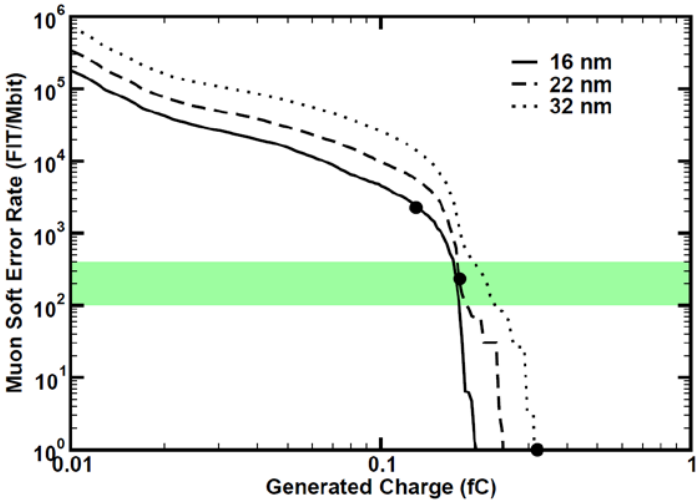


Fig. I-31 Simulation results of soft error rate produced by sea-level muons as a function of the charge they can generate in sensitive volumes of three advanced SRAM technologies. The shaded region represents the expected range of neutron-induced SER for the same devices. Adapted from [85].

The package-level radiation in the terrestrial environment is the sum of three (four if you count muons) mechanisms; terrestrial cosmic radiation in the form of penetrating high-energy neutrons that

cause nuclear reactions in the device materials producing high-ionizing secondary products, directly-ionizing alpha particles emitted from the radioactive impurities in the device and packaging materials, and the nuclear reaction between terrestrial thermal neutrons and ^{10}B in devices producing ionizing alpha particles and lithium recoils. In advanced devices where boron-doped glass is not used and/or device layers are not made with processes that leave high concentrations of ^{10}B, AND all chip materials are below the ultra-low alpha (<< 0.002 a/cm^2-hr), the high-energy cosmic ray neutrons tend to be the dominant cause of SEE. To accurately determine the SER of any product, the sensitivity for each of the three components must be accounted for by modeling or, preferably, direct characterization. Also, as indicated in the previous paragraph, as scaling continues and devices begin to operate with signals composed of fewer than 1500 electrons (0.2 fC) cosmic muon effects may become a dominant concern.

II-1. Thomas Kuhn: A "New" Theory of Scientific Revolutions

We will begin our retrospective look at SEE with a consideration of how science progresses as new discoveries are made and new theories are posited. Fifty-one years ago, Thomas Samuel Kuhn, a young physicist-turned-philosopher (it does not usually work the other way around ☺) and full professor at U.C. Berkeley, turned the philosophy world on its head by proposing a drastically different view of how advancement in science progresses. "*His 1962 book, The Structure of Scientific Revolutions, is one of the most cited academic books of all time. Kuhn's contribution to the philosophy of science marked not only a break with several key positivist doctrines, but also inaugurated a new style of philosophy of science that brought it closer to the history of science. His account of the development of science held that science enjoys periods of stable growth punctuated by revisionary revolutions* [87]." Kuhn's idea challenged the long-accepted view that science develops by a continual process of cumulative growth, with each new discovery adding to an established body of "truths", incrementally improving predictions and results while correcting minor errors. In sharp contrast to this traditional model, Kuhn showed

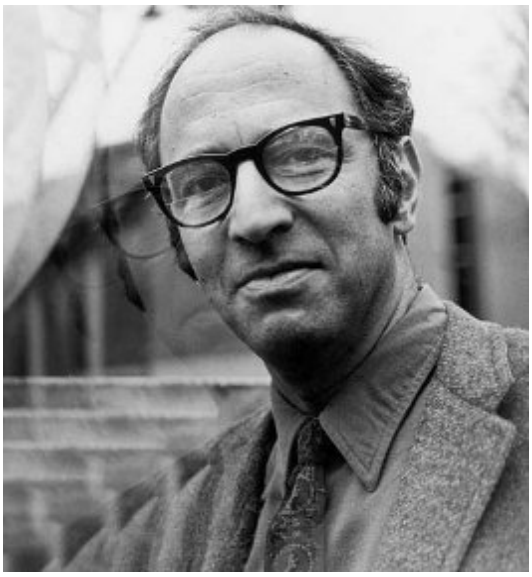
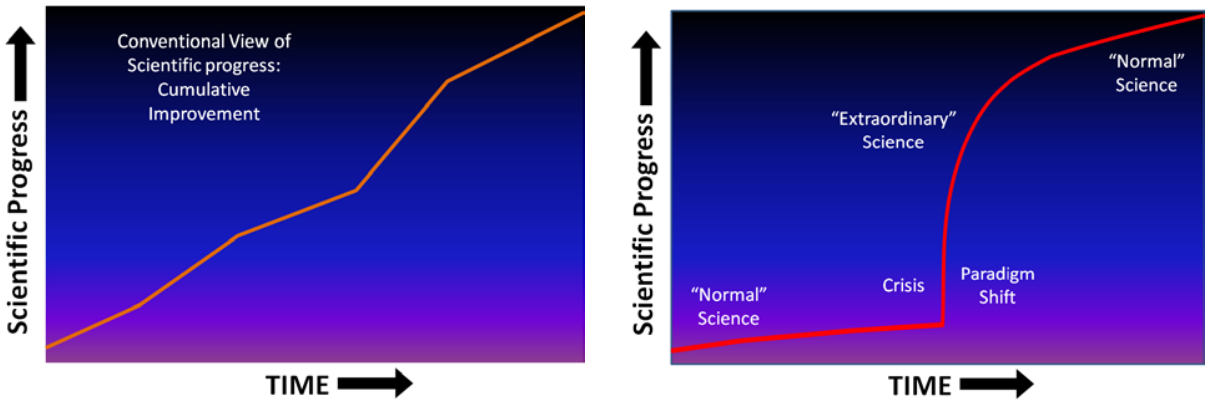


Fig. II-1 Thomas Kuhn (July 18, 1922 – June 17, 1996) American historian and philosopher of science. Image from [86].



a.) b.)

Fig. II-2 A graphical comparison of scientific progress in the traditional view (a.) where progress is cumulative and breakthroughs add to the existing paradigm. The model proposed by Kuhn (b), where breakthroughs or "extraordinary" science readjusts the existing paradigm to allow much more rapid progress. In general since the breakthroughs are incommensurate with prevailing views (the old ways are relinquished as being incorrect), the acceptance of a new paradigm is slow and receives a lot of resistance from the scientific "establishment".

scientific progress to be a process with alternating periods of “normal” science followed by periods of great upheaval in the status quo that he called “extraordinary” science [88]. Kuhn describes these periods of “normal” science as resembling the conventional models of scientific progress with “puzzle-solving” as the main activity [89]. Scientists working during a period of “normal” science expect to have a reasonable chance of solving the current problem being focused on, and that the methods of its solution will have a high degree of familiarity, being an extension of existing theory and techniques. Kuhn makes the point that “normal” science can succeed in making progress only if there is a strong commitment by the scientific community to their shared beliefs, values, techniques, and world-view – what Kuhn called their scientific “paradigm”. Thus, despite a strong desire for innovation and breakthroughs, scientists need to be conservative with respect to the generally-accepted “establishment” paradigm. Kuhn also describes highly disruptive periods of what he called “extraordinary” science, involving major revisions to an existing scientific paradigm. In many cases, when the new paradigm is eventually accepted as the “correct view”, the achievements of the preceding period will often be ejected as they are “incommensurate” with the new paradigm. The new paradigm cannot be understood through the conceptual framework/terminology of the established paradigm - like trying to explain quantum mechanics with classical notions. Periods of “extraordinary” science typically occur when there is a realization that the current paradigm is unable to solve fundamental problems or when conclusive new experimental data exists that refutes it. While anomalies are, at first, ignored or excused away, eventually the number and significance of anomalies create an environment that is ripe for a development of a new paradigm. Of course before it can be accepted as the “new truth”, the number and significance of the puzzles and anomalies solved by the new paradigm must exceed those of the old “establishment” paradigm. Unlike the conventional view that all breakthroughs are welcomed and add to existing scientific paradigm, Kuhn tells us that scientific revolutions are NOT sought out except under extreme circumstances - “crises” where the established paradigm is failing miserably. Scientific revolutions occur only when such crises force a new look at the established paradigm, highlighting and accepting its limitations and failures. This acceptance that something new is needed creates a zeitgeist encouraging the development of new or significantly revised paradigms that can solve the crisis that stymied the established paradigm. The decision to opt for a new paradigm is not a dictated by planned, logical, or a rational choice, but by a phase opening the normally conservative scientific community to consider competing and differing ideas.

In the Kuhn-sian view, which is now generally viewed as being closer to the “real” history of science than the previous cumulative progress model, science advances by “allowing” its paradigms to evolve, if somewhat fitfully, in response to the crises it encounters. With this view in mind, let us consider the dramatic development of silicon technology and the often fitful progress made in understanding the SEEs that evolved with its evolution.

II-2. From X-rays to Natural Radioactivity and Cosmic Rays

Our story starts, not too arbitrarily, just prior to the commencement of the 20th century, with the discovery of x-rays from cathode ray tubes. It was this discovery of the mysterious and invisible x-ray radiation that captured the imaginations of physicists all over Europe and catalyzed a decade of "extraordinary" research (in a literal as well as Kuhn-sian sense) that ultimately led to the understanding of the properties of radioactivity, the discovery of new radioactive elements, the radioactive decay law, the identification of protons, neutrons, and a modern concept of the nuclear structure of matter that ultimately laid the foundation for the development of nuclear, quantum, and solid state physics. A few years after accepting the Chair of Physics in the University of Würzburg, Wilhelm Röntgen became fascinated with the new generation of vacuum tubes and the emission phenomena associated with such tubes when operated at high voltages. Previous work by several others in this field had already been carried out and elucidated the properties of the cathode rays. Cathode rays are streams of electrons observed in vacuum tubes equipped with two electrodes such that when a high enough voltage is applied across the electrodes, the glass near the negatively-charged electrode (the electron source) is observed to glow as energetic electrons are stopped in the glass. They were first named, *kathodenstrahlen*, or cathode rays in 1876 by Eugen Goldstein [90]. In any case, Röntgen's real interest lay in characterizing the external effects from the various types of vacuum tube equipment when energized to high voltages. Late in 1895 Röntgen found that even with the discharge of the vacuum tube shielded to exclude any light from the cathode rays (fig. II-4), barium platinocyanide crystals placed several meters away would still fluoresce; thus, invisible rays from the discharge tube were causing the fluorescence in the crystals. In subsequent experiments Röntgen observed

Fig. II-3 Wilhelm Röntgen. Courtesy of Wikipedia commons.

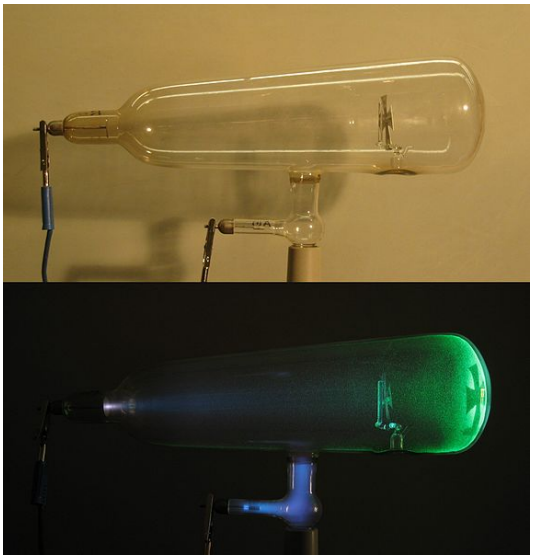
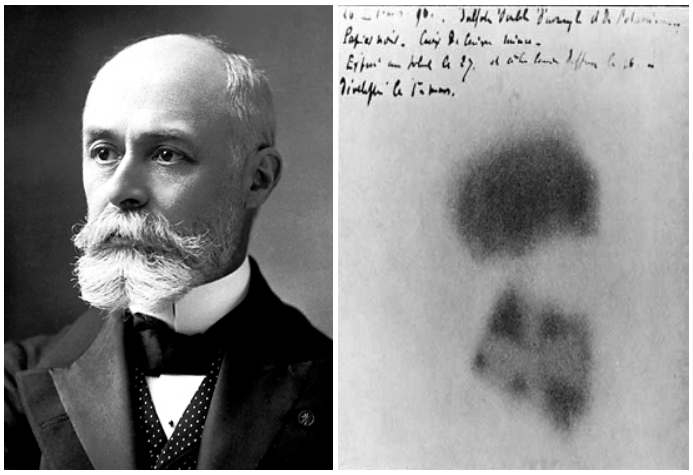


Fig. II-4 A Crookes tube (like one Röntgen was using) shown powered-down (top) and powered-on (bottom) in a darkened room. The blue glow is from the ionization of the rarified gas in the tube and the green glow is the fluorescence caused by energized electrons hitting the glass wall. Courtesy D-Kuru/Wikimedia.

that shields of different thicknesses and materials placed between the discharge tube and the fluorescing crystals changed the amount of fluorescence (brightness) – paper and thin metal foils and thick wood left the intensity of the fluorescence virtually unchanged while thick metal plates significantly reduced the intensity. In a slew of non-stop experiments (by now Röntgen was eating and sleeping in his lab), he replaced the barium platinocyanide crystals with photographic plates, repeating the shielding tests and finding that different levels of transparency (exposure) were left on the plates dependent upon the type and thickness of shielding material used. He held the hand of his wife motionless in the path of the mysteriously penetrating rays over a photographic plate (one hopes with her consent). After developing the photographic plate Röntgen observed a blurred image of his wife's hand with the flesh of her fingers painted as ephemeral shadows since the rays could penetrate easily and reach the film while the denser bones of her hand and the much denser metal ring blocked more of the rays leaving more substantial shadows. A picture of this first x-ray photograph is shown in fig. II-5. With further study Röntgen was able to show that the new rays were produced by the impact of cathode rays on the metal electrode. Because their nature was then not well established, he gave them the name "x-rays". Max von Laue and his students later demonstrated that x-rays share the same electromagnetic nature as light. For the discovery of x-rays, Röntgen was awarded the first Nobel Prize in Physics in 1901 [91].

Fig. II-5 "Hand with Ring" the famous x-ray photo of the hand of Röntgen's wife. Courtesy Wikipedia.



a.) b.)

Fig. II-6 Portrait of Henri Becquerel (a) and photo of one of Becquerel's photographic plates fogged by radiation from a uranium salt (b). The shadow of a metal Maltese Cross medal placed between the plate and the uranium salt is clearly visible. Images courtesy Wikipedia commons.

Henri Becquerel's early work was concerned with the polarization of light, the phenomenon of phosphorescence, and absorption of light by crystals. Following a discussion with Henri Poincaré on Wilhelm Röntgen's discovery a few months earlier of an invisible radiation that

accompanied a type of phosphorescence (the glow from cathode rays mentioned earlier) in vacuum tubes, Becquerel (pictured in fig. II-6a) decided to investigate whether or not there was any connection between these invisible rays and naturally occurring phosphorescence. Conveniently, he had inherited a supply of uranium salts (potassium uranyl sulfate); which phosphoresce on exposure to light. Becquerel had long been interested in phosphorescence, the emission of light of one color following a body's exposure to light of another color. In early 1896 for his first experiments Becquerel wrapped a photographic plate in two sheets of very thick light-tight black paper (so that the film could not be exposed by light). On the outside he placed a phosphorescent substance (the uranium salts from his father) and then left the whole package exposed to the sun for several hours. Upon developing the photographic plate, he could see the exposure caused by the phosphorescent substance. Becquerel repeated the experiment with metal coins and war medals between the plate and the uranium salts and saw the images of these objects appear on the plate (as shown in fig. II-6b). He correctly concluded that the phosphorescent substance in question emits rays which pass through the opaque, light-proof paper and expose the plates and that this radiation is absorbed by the metal objects and, thus, cannot expose the region beneath the objects. At that time he was convinced that uranium salts phosphoresced due to invisible rays (like Röntgen's newly-discovered x-rays) from the sun. Further experiments led him to doubt and eventually to abandon his first hypothesis. Several months later, Becquerel reported that the same uranium salts, arranged in the same way with respect to the photographic plates, but kept in the darkness (never exposed to the sun), still produced the photographic images on the plates. A month later after more experiments with non-phosphorescent uranium salts, he arrived at the correct conclusion that the penetrating radiation came from the uranium, itself, without any need of excitation by the sun's rays [92]. Ultimately, the phenomenon was found to be common to all the uranium salts and was correctly concluded to be a property of uranium atoms, themselves, and not some chemical reaction.

Becquerel made several more important contributions to the nascent field of radioactivity: He measured the deflection of beta particles (uranium emits both alpha and beta particles due to the decay of different daughter radioisotopes) in electric and magnetic fields and, from a determination of the charge-to-mass ratio, demonstrated that the beta particle was the same particle as the recently-identified electron. He also discovered that the presumed active substance in uranium, then called "uranium X", lost its activity over time while the uranium considered to be inactive when fresh, gained radioactivity over time. These observations laid the ground work for Ernest Rutherford and Frederick Soddy and led them to the transformation theory of radioactivity, which explained the phenomenon as a subatomic process allowing one element to spontaneously transmute into another. Becquerel's last major contribution was the report of the burn he received from carrying a sample of the radium in his pocket – a discovery that convinced the medical community that radiation could be effective in the treatment of cancers. For his discovery of

spontaneous radioactivity Becquerel was awarded the Nobel Prize for Physics in 1903, with Pierre and Marie Curie for their contributions to the study of Becquerel radiation.

Despite Becquerel's intriguing findings, the scientific community was largely myopically focused on Röntgen's x-rays, neglecting the far less penetrating Becquerel rays. The fact that the rays from uranium were not a popular research topic, was viewed as a plus by Marie Curie (a Polish immigrant to France who had married Pierre Curie and was studying under Becquerel) because it meant that she would not have a lot of papers to read and could begin her experimental work without delay. Utilizing the Curie brothers' new electrometer, which was far more sensitive than conventional designs of the time, she focused on making accurate qualitative measurements of the intensity of the Becquerel rays. Marie Curie confirmed Becquerel's observations that the minerals with a higher proportion of uranium emitted a higher intensity of radiation. Marie Curie first discovered that thorium gives off the same rays as uranium, and her systematic and laborious studies of various radioactive chemical compounds showed that the strength of the radiation did not depend on the chemical or physical state of the compound but only on the amount of uranium or thorium that it contained. To describe the ionizing effect of the rays emitted from the uranium and thorium, she invented the word "radioactivity". Marie Curie made the decisive conclusion that the ability to radiate did not depend on the arrangement of the atoms in a molecule, but that it must be related to the interior of the atom, itself, thus suggesting a subatomic structure. Her hypothesis was absolutely revolutionary, completely shattering the established (although weakening) paradigm that atoms were indivisible and the smallest structure of matter. Conceptually it is her most important contribution to physics. Pierre Curie gave up his research into crystals and joined Marie in her project (photo of the two in their lab is shown in fig. II-7). Her recent research had revealed that two uranium ores, pitchblende and chalcolite, were much more radioactive than pure uranium itself. She concluded that these ores might contain one or more additional, new radioactive elements. Working as a team, Pierre and Marie each took specific scientific tasks. The Curies employed various and demanding

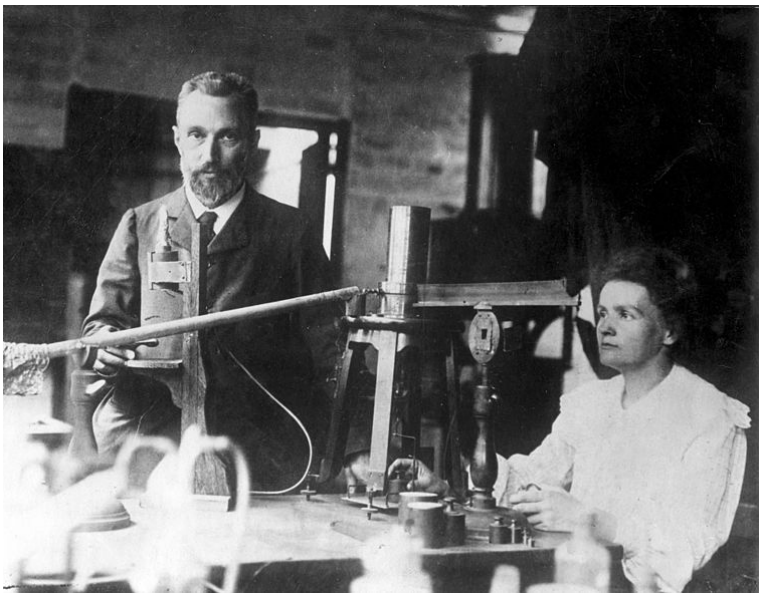
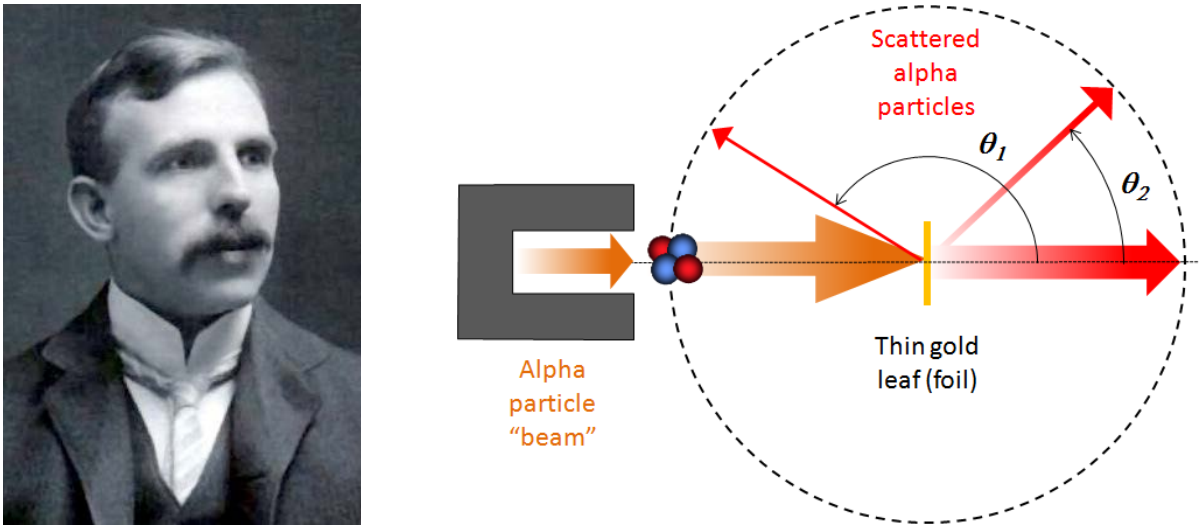


Fig. II-7 Photo of Pierre and Marie Curie in their laboratory. Image courtesy Wikipedia commons.

chemical procedures to isolate the different unknown substances present in tiny amounts in the pitchblende. After the materials were separated into different types of compounds, Marie and Pierre used the Curie electrometer to identify the most radioactive fractions. They thus discovered that two fractions, one containing mostly bismuth and the other containing mostly barium, were strongly radioactive. The bismuth fraction acted like bismuth, but since it was radioactive, it had to be something new. They named it "polonium" in honor of Poland (the country of Marie's birth). In the barium fraction they discovered another new element, which they named "radium". In 1911 Marie Curie was awarded a second Nobel Prize (the first person to be awarded twice). It was the Nobel Prize in Chemistry for her advancement of chemistry by the discovery of the two new radioactive elements, radium and polonium [93, 94].



a.) b.)

Fig. II-8 a.) Portrait of Ernest Rutherford (courtesy Wikipedia Commons), and b.) geometry of Rutherford "gold foil experiment" (actually done by Geiger and Mardsen) showing lead shield allowing only a narrow "beam" of alpha particles to escape. The beam was focused on a thin gold foil in a vacuum and the scattering of alpha particles as a function of detector angle was tabulated – the fact that back scattering (θ_1) was observed proved that the establishment paradigm of a diffuse atom was incorrect.

Ernest Rutherford, a New Zealand-born British chemist and physicist, started to explore radioactivity after hearing about Henri Becquerel's experiments with emissions from uranium salts and their effects on light-sensitive films. In 1899 he discovered two unique emissions, calling them "alpha" and "beta" rays. Rutherford characterized these emissions by studying their ability to penetrate foils of varying thicknesses and observing their behavior in magnetic fields. He also named a third, more penetrating emission discovered by Paul Villard, the "gamma" ray. Later Rutherford studied Becquerel's work on the decay in radioactivity over time and formulated the law of radioactive decay, demonstrating that the time for half the sample to decay was a constant for a given material. While Marie Curie was the first to claim that radioactivity was based on activity at the atomic level, Rutherford and Fred Soddy

“fleshed out” the concepts of “Atomic Disintegration", showing that the radioactivity from uranium and radium salts was actually produced by the spontaneous disintegration of atoms into other types of atoms. Studying the effects of alpha particle bombardment on thin gold foils (thin gold leaf, only a few micrometers thick so that the alpha particles could penetrate), Hans Geiger and Ernest Marsden, then students under Rutherford’s direction, observed that occasionally alpha particles we scattered back at high angles (back towards the point of emission). Rutherford claimed that the back-scattering of alpha particles was "*almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.*" This finding produced a “Kuhn-sian” crisis in that it could not be explained by prevailing theories of matter – the "plum pudding model" of atoms maintained that the electrons in an atom were randomly moving in a dispersed sea of positive charge (electrons like plums in a bowl of plum pudding). The significantly heavier alpha particles were expected to traverse the gold film with minimal deflection since the positive charge was assumed to be diffuse and incapable of inducing large trajectory changes (fig. II-8b). This discovery led Rutherford to propose a new model of the atom with a very small, centralized and charged nuclear mass surrounded by orbiting electrons of much lower mass. He also measured the mass-to-charge ratio of the alpha particle, leading him to hypothesize that alpha particles were doubly-charged helium ions (later shown to be bare helium nuclei). In 1907, Rutherford and Thomas Royds finally proved that alpha particles were, indeed, helium ions by allowing alpha particles to penetrate the very thin glass wall of an evacuated tube capturing a large number of the ions inside the tube. They neutralized the ions with an electric spark inside the tube and discovered that the spectrum was that of helium gas. In 1920, Rutherford also proposed, that hydrogen nuclei were a basic constituent of all matter and named them “protons” and proposed the concept of neutrons to explain the anomalous extra mass of nuclei, a theory that was verified in 1932 by James Chadwick, who discovered neutrons when bombarding beryllium films with alpha particles. For his investigations into the disintegration of the elements and the chemistry of radioactive substances, Rutherford was awarded the Nobel Prize in Chemistry in 1908 [95, 96].

In 1785 the French physicist Charles Augustin de Coulomb published three reports of electricity and magnetism where he introduced the inverse square law (previously observed by others). He used a torsion balance to study the repulsion and attraction forces of charged particles and some experiments demonstrating that charge placed on isolated bodies spontaneously discharged much more quickly than leakage currents could explain – an unknown source of charge was compensating the charged object. Many scientists pondered the meaning of this phenomenon, and in 1879 William Crookes made a break-through when he observed that the rate of discharge was decreased when the air pressure was reduced - the source of this mysterious charge was, therefore, in the air. As mentioned previously, in the late 1890s, spontaneous radioactivity had been discovered, and Marie Curie and others had observed that

electroscopes promptly lost stored charge if exposed to radioactive material. Charles Wilson, Julius Elster, and Hans Geitel, with higher sensitivity electroscope designs and thick metal shielding, independently concluded that ionizing agents were coming from outside the electroscope and that some of this radioactivity was highly penetrating [97]. The simplest hypothesis was that its origin was related to the natural decay of radioactive materials in the Earth's crust. A terrestrial origin for the radiation was a commonplace assumption. In 1901 Wilson was the first to suggest that the ionization could be caused by extremely penetrating extra-terrestrial radiation. He performed a number of experiments in deep tunnels, but because his instruments showed no reduction in ionization rate, the hypothesis of an extra-terrestrial source of radiation was abandoned.

In the early 1900s, a German physicist and Jesuit priest, Theodor Wulf, designed and built an improved electrometer. Replacing the two metal leaves with a metal-coated glass wire and adding a glass spring to maintain tension on the wire, the "Wulf Electrometer", as it came to be known, was far more compact than previous instruments. The metal-coated wire is suspended in an electric field created by two electrodes. As the wire is charged or discharged, it is deflected towards one of the electrodes. By measuring the magnitude and direction of deflection in the wire, the polarity and magnitude of the charge accumulation or discharge can be ascertained [98]. In 1909 Wulf used his new instrument to test the origin of the ionizing radiation that caused the spontaneous discharge. He measured the ionization rate from the top floor of the Eiffel Tower (a height of 273 m). He expected to find less ionization at the top of the tower than at ground level assuming the source of ionization was terrestrial in origin. However, the decrease in the ionization rate was too small a decrease to confirm the terrestrial hypothesis - he found that the amount of radiation was barely half of its ground value. If the ionization rate had been solely due to radiation originating from Earth's surface, as was the general assumption, Wulf's electrometer should have shown a halving of radiation levels every eighty meters. He should have seen a sixteen-fold drop in ionizing radiation if its sole source was terrestrial. He therefore concluded that excess energy was coming from radiation outside the Earth's atmosphere and published a paper summarizing the results of his observation on the Eiffel Tower. As one of the first experimenters to detect excess and, as of yet, unexplained atmospheric radiation,

Fig. II-9 Pacini making an electrometer measurement at sea-level. Image courtesy Wikipedia commons.

Wulf's observations were largely dismissed due to the prevailing paradigm that maintained that the source of all terrestrial radiation was emitted from radioactive impurities in the Earth's crust. Domenico Pacini (fig. II-9), an assistant meteorologist in Rome at the Italian Agency of Meteorology and Geodynamics, was another person to question the established paradigm that the ionization in air came from the radioactivity of the Earth's crust. Pacini made numerous electrometer measurements on mountain-tops, coastal areas, and out at sea. He later placed his electrometer in a water-tight copper enclosure and compared the rate of ionization at sea level and at 3 meters underwater. He found that the ionization underwater was 20% lower than at the surface, consistent with the absorption by water of the radiation coming from outside. Pacini concluded: "*a sizable cause of ionization exists in the atmosphere, originating from penetrating radiation, independent of the direct action of radioactive substances in the crust* [99]." Pacini published his results in 1912, but similar to Wulf's earlier work, this paper was largely ignored as it was eclipsed by Hess's work published in the same year.

Reading about these earlier experiments, Victor Hess, an Austrian physicist, became obsessed with the idea that the source of ionization could be located in the sky rather than the ground as Wulf's

a.)  b.) 

c.)

Mittlere Höhe über dem Erdboden m	Beobachtete Strahlung in Ionen pro ccm und sec.			
	Apparat 1	Apparat 2	Apparat 3	
	Q_1	Q_2	Q_3 (reduziert)	Q_3 (nicht reduziert)
0	16,3 (18)	11,8 (20)	19,6 (9)	19,7 (9)
bis 200	15,4 (13)	11,1 (12)	19,1 (8)	18,5 (8)
200—500	15,5 (6)	10,4 (6)	18,8 (5)	17,7 (5)
500—1000	15,6 (3)	10,3 (4)	20,8 (2)	18,5 (2)
1000—2000	15,9 (7)	12,1 (8)	22,2 (4)	18,7 (4)
2000—3000	17,3 (1)	13,3 (1)	31,2 (1)	22,5 (1)
3000—4000	19,8 (1)	16,5 (1)	35,2 (1)	21,8 (1)
4000—5200	34,4 (2)	27,2 (2)	—	—

Fig. II-10 a.) Photographs of Victor Hess (image from Wikipedia commons) and of b.) Hess (in the center of the gondola) and a colleague in the balloon prior to one of their flights (image from VF Hess Society, Echophysics, Schloss Pöllau/Austria) and c.) original table from Hess's 1912 published findings clearly showing the more than 200% increase in ionization at 5km [100, 101].

experiment had suggested. Before making any ascents, he determined the height above which ground radiation should taper off (about 500 meters) and ruggedized several Wulf electrometers so that they could operate reliably over the temperature and pressure changes encountered in balloon flights. Hess immediately put his new equipment to good use following Wulf's lead by taking measurements of radiation levels in the atmosphere but over a much larger range of altitudes. Between 1911 and 1913, Hess made ten daring ascents in an un-tethered hydrogen balloon (remember the Hindenburg) using three of his modified electrometers to make careful measurements of the ionizing radiation as a function of altitude up to a maximum altitude of 5.3 kilometers (with no bottled oxygen, he risked hypoxia). A portrait of Hess and a photo showing him prior to one of his balloon experiments are shown in II-10a and 10b, respectively. He repeatedly measured levels of ionizing radiation that were approximately 200% higher at the maximum altitude he attained (see table from his original paper of 1912 in fig. II-10c). To determine if the sun was responsible for the increase in radiation at higher altitude, Hess made five of ascents at night (one during a solar eclipse). With the Earth or, in one case, the moon, blocking most of the sun's radiation, Hess still measured similar levels of radiation at higher altitudes, thus, eliminating the sun as a possible source of the increased radiation levels. Hess published his work in the Proceedings of the Viennese Academy of Sciences concluding that there was a source of radiation penetrating the atmosphere from outer space. Robert Millikan, who confirmed Hess's results in 1925, coined the term "cosmic rays" for this radiation from above. Hess's discovery opened the door to many new discoveries in nuclear physics; a discovery for which Hess was awarded the 1936 Nobel Prize in Physics.

With the many fundamental paradigm-busting breakthroughs heralding the advent of the 20th century leading to the discovery and greater understanding of natural radioactivity, atoms, and, finally, the radiation from outer space, the stage was set for future work that was to elucidate how these mechanisms affect solid state electronic devices; a technology born of the new atomic physics gleaned during this era.

II-3. A Brief History of Electronics – from Vacuum to Solid State

The cathode ray tube (the same device that kicked off the discoveries of the x-ray and, ultimately, radioactivity and cosmic rays) were an extension of Thomas Edison's original idea of using an evacuated glass tube with two terminals to create an electric light source with a tungsten filament heated to glowing by a voltage applied across the two terminals. While investigating the cause of blackening in incandescent lamps in 1880, Edison discovered that current flow could occur inside an evacuated light bulb. At this point in time, it was not known that electrons could flow in a vacuum between two electrodes. In fact, electrons weren't identified until 1897, when Joseph Thomson developed a vacuum tube by adding a third terminal to Edison's design with the aim of better understanding the nature of cathode rays. He showed that the cathode rays were actually made up of particles, or "corpuscles", as Thomson called them, that were present in all matter. He had discovered the electron for which he received the 1906 Nobel Prize in Physics. In 1904 Sir John Fleming developed a vacuum diode that was designed into radio receivers of that day. This was the beginning of the vacuum tube dynasty.

Fig. II-11 Photo of an early triode vacuum tube, named an "audion" by its inventor Lee DeForest. This was the first vacuum device capable of amplification. Image from Wikipedia commons.

In 1908 Lee DeForest interposed a third electrode called a "grid" between the cathode and anode of the Fleming "rectifier" making it a triode device. The addition of this electrode enabled one to control the large flow of electrons from the cathode to the anode with a tiny input into the grid – in essence creating a device with gain that could be used to build electronic amplifiers. DeForest called this vacuum tube an "audion." A version of this early vacuum tube device is shown in fig. II-11. Unfortunately, the first audions were limited in dynamic range, grossly nonlinear, and had erratic performance such that even DeForest himself stated that he could not predict whether the anode current would go up or go down when a signal was fed into one of his audions. In 1912, H.D. Arnold and Irving Langmuir independently found that the DeForest audion could be made stable using a very high vacuum to form the tube [102]. From that point on vacuum tubes quickly became the dominant active component for low-power electronic circuits and played an important role in the emergence of military, industrial, and home electronics applications until the early 1960s, when they were rapidly displaced by solid-state devices [103]. It is somewhat ironic that in the 1920s, the revolutionary improvements and growing applications of the vacuum tube electronics technology (at this point the established paradigm for electronics was the

vacuum tube) totally eclipsed the first invention of the transistor [104]. As early as 1925 Julius Lillienfeld, an Austro-Hungarian physicist, applied for and received the first patent for what would now be considered a junction field-effect transistor (JFET) fabricated with polycrystalline copper sulphide as a semiconducting substrate and a gate junction formed with a thin sheet of aluminum. A 1927 patent for a bipolar transistor of semiconductor-metal-semiconductor construction clearly describes the importance of a very thin base, and a 1928 patent application by Lillienfeld describes the first MOSFET using aluminum as the gate electrode, aluminum oxide as the insulator (formed by anodization), and semiconducting copper sulphide as the substrate deposited last [105]. A combination of the Great Depression and a rapidly growing and maturing vacuum tube industry conspired against Lillienfeld and several other inventors in the U.S. and Europe that patented various transistor ideas. Ultimately, none of these inventions were developed further, and the transistor concept was buried in barren ground – electronics manufacturers were not interested in investing in a risky and unproven technology, particularly when vacuum tube business was burgeoning - so these early transistors were ignored and never developed. The transistor would have to wait until 1947 until its reinvention and introduction into electronic circuits enabled it to dethrone the vacuum tube.

William Shockley, John Bardeen, and Walter Brattain, all part of Bell Labs' newly-formed Solid State Physics Group, were tasked with understanding the nature of the electrons and conduction at the interface between metal and a semiconductor with the goal of devising useful devices. As an outshoot of this work, in December 1947, Bardeen and Brattain created the first transistor by making two point contacts very close to one another - the "point contact" transistor. An example of one of the first such transistors is shown in fig. II-12. A small positive current applied to one of the two contacts controlled the current flowing between the other contact and the base of the germanium substrate. Since a small change in the first contact current could induced a much greater change in the second contact current, this device provided amplification. A few of the point-contact transistors devices were made and assembled to make an audio amplifier that was demonstrated to Bell chief executives, who were very impressed that it didn't need time to "warm up" like vacuum tube circuits. The reinvention of the solid-state transistor ushered in a huge research effort [106]. Shortly after the invention of the point-contact transistor, Shockley invented an

Fig. II-12 Photograph of the first point-contact transistor developed at Bell Labs in 1947. Adapted from an image from http://www.porticus.org/bell/images/.

entirely new, considerably more robust, type of transistor with a layered or 'sandwich' structure. This structure went on to be used for the vast majority of all transistors into the 1960s and evolved into the bipolar junction transistor. The junction transistor was easier to understand theoretically, and could be manufactured more reliably. For their discovery of the transistor, Shockley, Bardeen and Brattain received the 1956 Nobel Prize in Physics. Germanium had large leakage currents (due to its small band gap), formed a poor quality oxide, and could not be operated at high temperatures; rendering silicon a good potential candidate for future developments - however, silicon had a unique set of challenges and many scientists predicted it would be years before a production-worthy silicon transistor process could be developed, and others theorized it could never be built in quantity. By the summer of 1953, however, a small team of engineers at Texas Instruments was working feverishly on producing silicon crystals with stable electrical junctions and developing reliable and repeatable methods of fabricating transistors from this new semiconductor. In April 1954 Gordon Teal, then the chief of Texas Instruments semiconductor R&D, recounts that prior to his talk at the Institute of Radio Engineers (IRE) National Conference on Airborne Electronics, in Dayton, Ohio, all the speakers preceding his were bearish on silicon technology development, "*One after the other (the speakers)…remarked about how hopeless it was to expect the development of a silicon transistor in less than several years. They advised the industry to be satisfied with germanium transistors for the present*." Teal then gave his presentation demonstrating that Texas Instruments had made big strides in developing reliable silicon transistors and ended with a statement that galvanized the audience - "*Contrary to what my colleagues have told you…we have three types of silicon transistors in production.*" To demonstrate that the new silicon transistors were much more stable as compared with the germanium transistors of the day, he played a record on a turntable with the germanium transistors in the amplifier section submerged in a beaker of hot oil, and the sound died away rapidly as the devices failed from the exposure to the high temperature. The demonstration was then repeated with an amplifier built with the new silicon transistors, but this time the music played on [107]. Texas Instruments was the first company to offer

Fig. II-13 When TI announced commercial availability of the silicon transistor in 1954, this photograph of a silicon transistor superimposed on a three-cent U.S. postage stamp was used to show its relative size.

silicon transistors commercially, introducing the 900 – 905 series of grown junction transistors in mid-1954. A 900 series is shown in fig. II-13. This was several years before other companies were able to produce commercial silicon units, allowing Texas Instruments to gain a large share of the highly profitable military market – silicon units performed reliably at much higher temperatures than germanium units and were in high demand by the U.S. military. The 900 series were used to create the first commercial transistor radios. The first units sold for over $100 each, and the cost remained high for many years [108]. The era of the discrete transistor had arrived, and vacuum tube applications were quickly replaced by equivalent transistor circuits that were lighter, smaller, cheaper, operated at lower power, and were more reliable. By 1961, just 14 years after the invention of the transistor, the semiconductor industry surpassed $l billion in revenue.

Between 1958 and 1959, Jack Kilby at Texas Instruments and Robert Noyce at Fairchild, independently came up with a solution to the problem of further miniaturization of discrete electronic components. Early in 1958, Kilby joined Texas Instruments in Dallas, Texas tasked with the miniaturization of discrete components. As a new employee, he had no vacation like the rest of the staff, and working alone in the lab in the summer of 1958, he made the first integrated circuit device with all the components on the same chunk (monolith) of semiconductor material. In July he wrote in his lab notebook, "*Extreme miniaturization of many electric circuits could be achieved by making resistors, capacitors, transistors and diodes on a single slice of silicon.*" Then he proceeded to show in the next pages of his notebook, how it should be done. By the end of August, a simplified version of his circuit had been built using wire bonds for interconnection and the world's first integrated circuit worked (one of the original Kilby integrated circuits is shown in fig. II-14). In January of 1959, another version of this very same idea occurred to Noyce. Fairchild was already developing a planar transistor process, and using this process Noyce focused on the matter of creating interconnections between the integrated components. While the Kilby circuit required bond wires to make circuit connections, Noyce's process used thin metal films that were photo-defined and etched in the batch processes of the semiconductor device

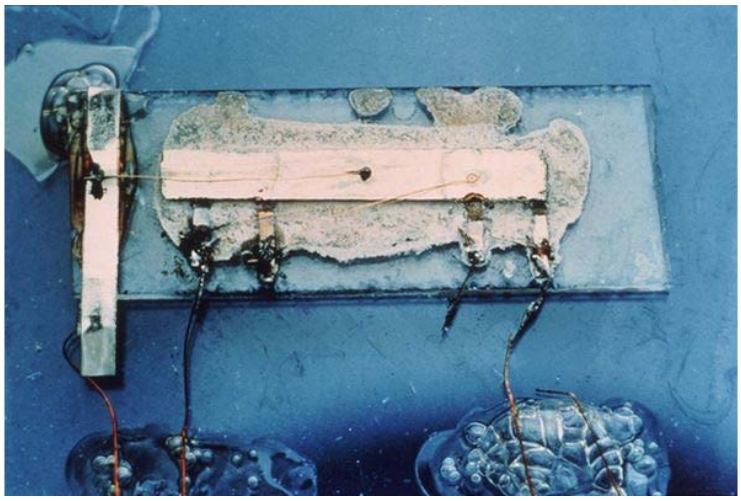


Fig. II-14 The first working integrated circuit was created by Jack Kilby in July of 1958 and contained a single transistor and supporting components on a thin slice of germanium and was 1.6 x 11.1 mm. Image courtesy Texas Instruments.

[109]. Ultimately, through the combined innovation of Kilby and Noyce and the work of many researchers to tackle the yield and reliability issues that allowed the development of reliable semiconductor manufacturing processes, the discrete transistor had evolved into the integrated circuit, ushering a new generation of electronic devices that would largely replace discrete transistors, as the production of discrete transistors had replaced vacuum tubes a decade earlier.

General Microelectronics introduced the first commercial integrated circuit in 1964. The chip was a 20-bit shift register formed from 120 PMOS transistors for use in an electronic calculator [110]. Nearly 50 years later, silicon technology scaling has enabled Intel's top-of-the-line Itanium *Poulson* eight-core CPU formed with over 3 billion transistors on a slice of silicon ~ 23mm on a side [111]. In the next section we will see two differing views from the 1960s on the potential for technology scaling.

II-4. Two Early Views on Scaling: The Pessimist vs. Optimist

At about the same time that NSREC was getting started (in 1962 the Nuclear Radiation Effects Program was part of the AIEE Summer General Meeting [112] so was a precursor to the IEEE NSREC program that started in 1964), engineers at RCA, J. T. Wallmark and S. M. Marcus, extended a simple model of packing density in integrated circuits proposed by J. J. Suran [113] in an attempt to predict the limits on technology scaling. Considering factors such as heat dissipation, fluctuation in doping density and element size, and the effects of cosmic rays, they predicted that "*for non-redundant semiconductor devices a limit on packing density, and therefore also on minimum device size, exists, set by fundamental laws of nature*". They posited that the fundamental limit to circuit scaling would be devices with dimensions in the 10 μm regime [114] as illustrated in fig. II-15. The authors found that "*the temperature limitation sets a limit to how highly the material in the active region of semiconductor devices with high packing density may be doped. For material doped below this critical density the limitations caused by cosmic rays and natural radioactivity are more severe; for material doped above this critical density, temperature is the limiting factor*". At that time state-of-the-art device dimensions were in the 50-20 μm range, so this estimate did not bode well for the future development of the semiconductor business. Luckily for us, Wallmark and Marcus's estimations were excessively conservative as devices in today's most advanced technologies measure only a few nanometers on a side – four orders of magnitude smaller than the "fundamental limit" predicted by the authors in 1962. Even if their prognostications were largely incorrect, the authors were prescient on several key aspects of SEEs and their impact to future electronic systems. They were the first to predict terrestrial cosmic rays limiting the reliability of electronics (nearly three decades before neutron-induced SEU were shown to be a dominant failure mode for SRAMs and DRAMs [116, 117]). They also provide one of the first published descriptions of a soft error, "*the ionization of the cosmic ray will create a temporary excess of hole-electron pairs which will decay with the normal lifetime of minority carriers in the device. If this lifetime is of the order of, or larger than, the minimum operating time of the device, a false signal will result.*"

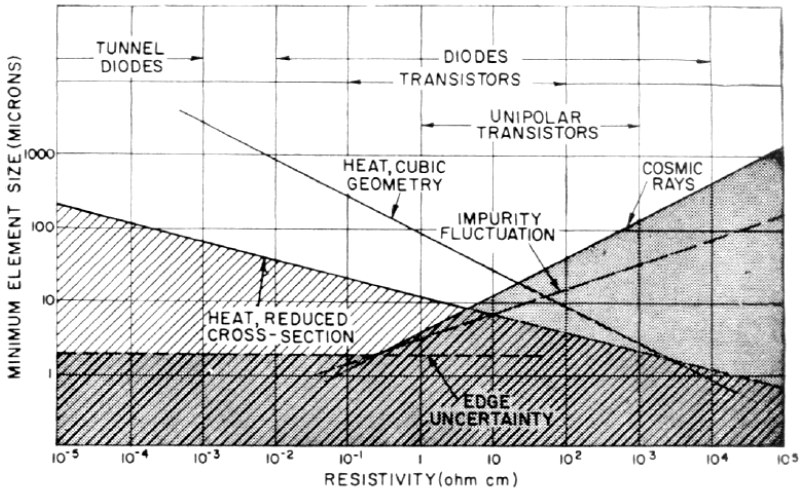


Fig. II-15 Plot of minimum element size vs. the resistivity showing the limits to scaling imposed by heat, impurity fluctuation, and cosmic rays. The limit occurs an element size of 10 μm and ρ ~ 5 Ω-cm.[115]

Just three years later, in 1965, Gordon Moore, then the director of R&D at Fairchild Semiconductor, extended the logical consequence of integrating an increasing number of components through miniaturization (technology scaling) and proclaimed that "*the advantages of integration will bring about a proliferation of electronics…will lead to such wonders as home computers…automatic controls for automobiles, and personal portable communications equipment…integrated circuits will also switch telephone circuits and perform data processing…In addition, the improved reliability made possible by integrated circuits will allow the construction of larger processing units*" [118]. The 1965 article includes a prophetic cartoon showing computers being sold in department stores like other commodities. Contradicting the thermal limits predicted by Wallmark and Marcus, Moore pointed out that as integration levels increased with continued technology scaling, the node capacitance drops so that specific circuits can operate at a lower power per area. His main focus was the decrease in cost afforded by integrating more devices, "*reduced cost is one of the big attractions of integrated electronics, and the cost advantage continues to increase as the technology evolves toward the production of larger and larger circuit functions on a single semiconductor substrate.*" Moore looked at the process from the economic point of view, with cost reduction driving the miniaturization and yield (defect density) issues limiting the degree of scaling that could be achieved. "*For simple circuits, the cost per component is nearly inversely proportional to the number of components…but as components are added, decreased yields more than compensate for the increased complexity, tending to raise the cost per component. Thus there is a minimum cost at any given time in the evolution of the technology*." At the time the paper was written, the balance between scaling and yield resulted in monolithic devices with approximately 50 components. Moore extended his scaling assumptions based on the observation that "*the complexity for minimum component costs has increased at a rate of roughly a factor of two per year. Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years.*" Based on the rate of density increase that he was predicting, Moore went on to say that he expected that "*by 1975, the number of components per integrated circuit for minimum cost will be 65,000.*" Nearly a decade later, in 1974, the 8080 (Intel) and the 6800 (Motorola) 8-bit microprocessors were introduced. Both of these products were fabricated with about 4000 transistors and a similar number of passive components. Thus Moore's prediction, while somewhat overly optimistic in terms of rate of progress, was born out. Fortunately for the semiconductor industry and electronics consumers, the prediction became a reality (and became "Moore's Law") with the component integration density doubling about every 2-3 years for more than four decades, as illustrated in fig. II-16. In closing it is interesting to note that Wallmark and Marcus assumed 3D volumes and concluded that 10^8 devices/cm^3 would be the limit. The semiconductor industry, stuck to a quasi-2D framework, managed technology

scaling far beyond the 1962 prognostications. However, the actual component count is ~ 10^9 devices/cm^3, which is within an order of magnitude of the original predictions of Wallmark and Marcus.

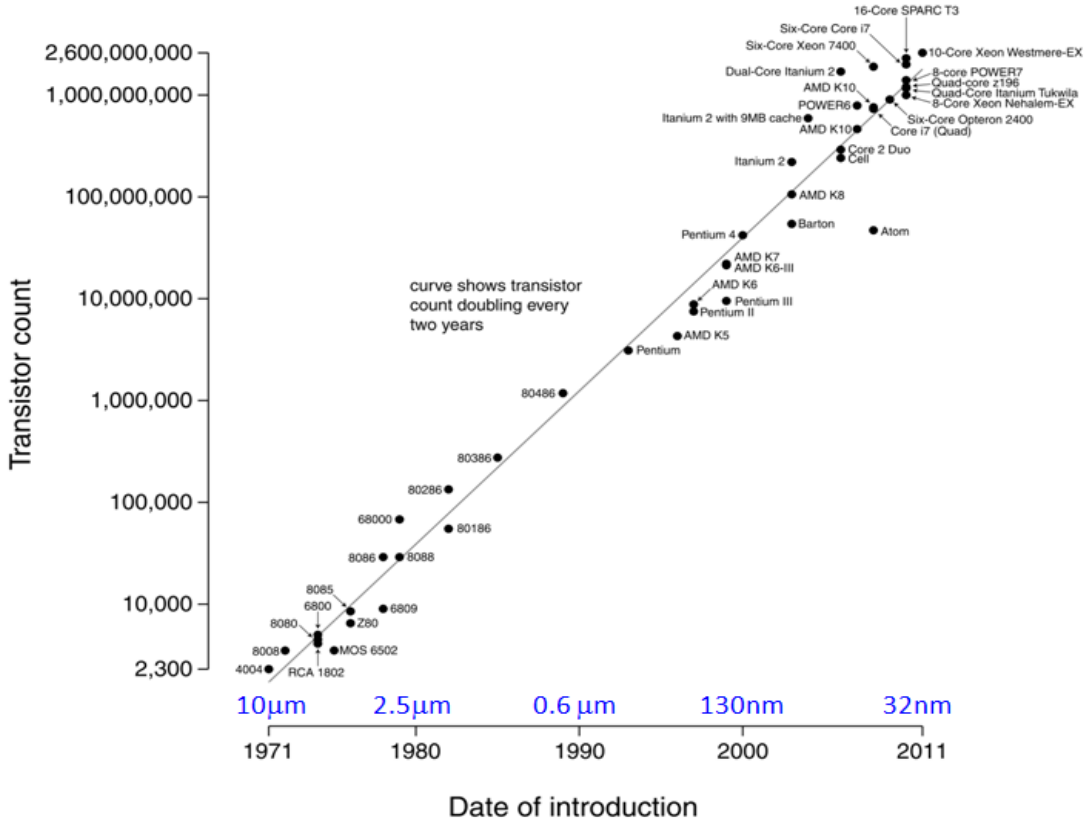
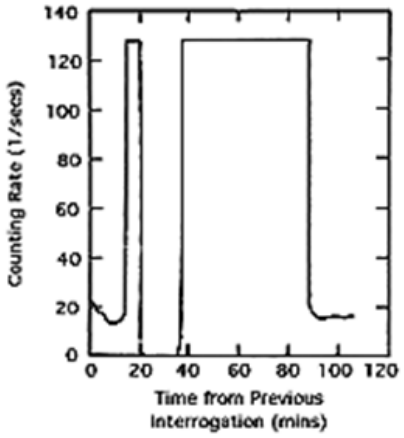


Fig. II-16 Transistor counts vs. date (and minimum feature size in blue) over the last four decades. In line with Gordon Moore's 1965 prediction of the miniaturization trend the number of transistors integrated on a single silicon chip has doubled every 2-3 years. Plot adapted from (http://en.wikipedia.org/wiki/Moore's_law).

II-5. The First Hints of Fragility - Upsets in Space

Starting in the early 1950s, James Van Allen led the exploration of the near-space environment by devising and launching a number of balloon-rocket combinations, nicknamed 'rockoons', to lift payloads of instruments into the upper atmosphere. Inexpensive military-surplus rockets were used to keep the program economical [119]. The balloons were used to lift the rockets above the densest part of the atmosphere. By avoiding the aerodynamic drag of the lower atmosphere, the rockets were able to attain much higher altitudes when fired from balloon altitudes. The most significant finding from these expeditions, launched from 1953 – 1956, was the detection of "*auroral soft radiation*" by on-board Geiger counters [120].

a.)



b.)

Fig. II-17 a.) Photo of the first American satellite, the Explorer I, launched into orbit on Jan. 31, 1958 shortly after the Russian Sputnik, and b.) the output of the Geiger tube vs. time. Note that the tube was saturated by the high number of ionizing events most of the time while the satellite traversed the Van Allen radiation belts. Both images are courtesy of NASA (http://www.jpl.nasa.gov/explorer/).

This earlier work led directly into the U.S. exploration of space. America's first satellite, the Explorer 1 (shown in fig. II-17a), was launched into orbit from Cape Canaveral on January 31, 1958 by the U.S. Army Ballistic Missile Agency. The Explorer was designed, constructed, and operated by the Jet Propulsion Laboratory. In addition to a micrometeorite detector, it carried, as its primary instrument, one of Van Allen's cosmic ray detectors. When Van Allen's team analyzed data from their instruments on Explorer I and Explorer III (launched March 26, 1958), they found surprising anomalies in the Geiger counter data (shown in fig. II-17b). The data suggested high-flux and low-flux distributions and ultimately led to the discovery of torus-shaped regions of charged particle radiation wrapped around the Earth. In addition to the Explorer satellites, the Pioneer III lunar probe, while failing in its mission to place a payload on the moon, attained an altitude of 63k miles and provided Van Allen additional data that led to discovery of a second radiation belt. By the end of 1958, Van Allen's contribution was the realization that the near-space

environment around earth contains a high level of radiation trapped by the Earth's magnetic field from several hundred miles above the surface to several thousand miles into space. The Van Allen radiation belts were named after him following the 1958 satellite missions. Thus, early in the space-age, it had become obvious that high-radiation levels were one of the challenges for space missions in the future. It would be more than a decade before the first evidence of single-event effects on electronics would be observed and attributed to cosmic ray events.

One of the first to describe radiation as a potential problem for satellite electronics [121], G. W. A. Dummer warns "*as is well-known, radiation and high energy particles can change the electrical properties of a semiconductor. Some changes are transitory and cause spurious signals, others can cause a permanent alteration to constituents of the crystal structure and therefore permanently affect the electrical properties. A solid circuit will be just as sensitive to radiations and high energy particles as a conventional transistor, and may be more so because all the "components" in a solid circuit are made of semiconductor material*." Analysis of upsets in satellite electronics being caused by cosmic ray events was first reported by D. Binder et al. in 1975 [122]. The authors reported that "*anomalies in communication satellite operation have been caused by the unexpected triggering of digital circuits*." In all, four triggering events could not be explained, and, thus, the effect of cosmic ray events as a root cause was proposed. The circuits in question were single J-K flip-flop integrated circuits built in a ~ 44 μm (the bipolar transistors are claimed to be 22 μm in radius, but it is not known if this was a minimum-sized device) bipolar technology. A focused electron beam from a scanning-electron microscope was used to identify the sensitive nodes of the flip-flops. The authors deduced that "*there were four sensitive transistors per flip-flop*." After determining the number of sensitive transistors and an estimate of the charge collection efficiency, the authors were able to calculate the type of event that could deposit enough charge into the sensitive bipolar regions to cause a soft error. They concluded that "*19 MeV must be deposited by the cosmic ray along a track length of a few microns*" and that this amount of energy could only be "*satisfied by galactic cosmic rays in the iron group*." Thus, of all the cosmic ray constituents, energetic heavy ions of iron were the only particles deemed to have sufficient LET to penetrate the shielding and cause the flip-flops to upset. The authors' characterization and analysis was combined with an early computer code for determining the event rate based on known galactic cosmic ray intensities of various particles. Their simple prediction was within 2x (higher) than the actual observed rate of four failures over 17 satellite years. This was a "landmark paper" in that it introduced the idea that upsets could, indeed, be caused by cosmic events, and it combined real satellite upset data with circuit sensitivity characterization and an event modeling to demonstrate, at least theoretically, that the observed error rate was of the same order as the number of cosmic ray events that would have been capable of producing those upsets. Three years later, J. Pickel and J. Blandford, Jr. reported a similar, but more detailed,

analysis on the effect of cosmic rays on 4kbit NMOS DRAM in the satellite environment [123]. In a simple calculation neglecting diffusion, the authors found that a cosmic ray event in the depletion region would have to deposit at least 5.6 MeV to cause a data error in the DRAM and have an LET or stopping power in excess of 4600 MeV-cm^2/gm to penetrate the aluminum shielding used in their satellite. The ions that can potentially upset the electronics are showing in the red region in fig. II-18. Since the required energy is lower than the previously mentioned example, it is not surprising that in addition to upsets by heavy ions of iron, energetic ions of aluminum and carbon would also contribute to the upset rate. Furthermore, as compared with the previously mentioned work, the upset rate is much higher since the DRAM devices have a lower critical charge and are sensitive to a larger energy range and type of particle. This paper is important in that the summary describes the key considerations in making and operating DRAM circuits so that they are more robust - "*minimum sizing on source/drain diffusion of critical transistors would improve error rate…operational parameters that reduce bit error rate include higher refresh frequency, higher voltage on the storage node, and lower temperature*." By the late 1970s, it was clear that space was a hostile environment for electronics due to the large amount of ionizing radiation trapped in the Van Allen belts. At about this same time, engineers started to notice that even electronics in the terrestrial environment were becoming sensitive to radiation.

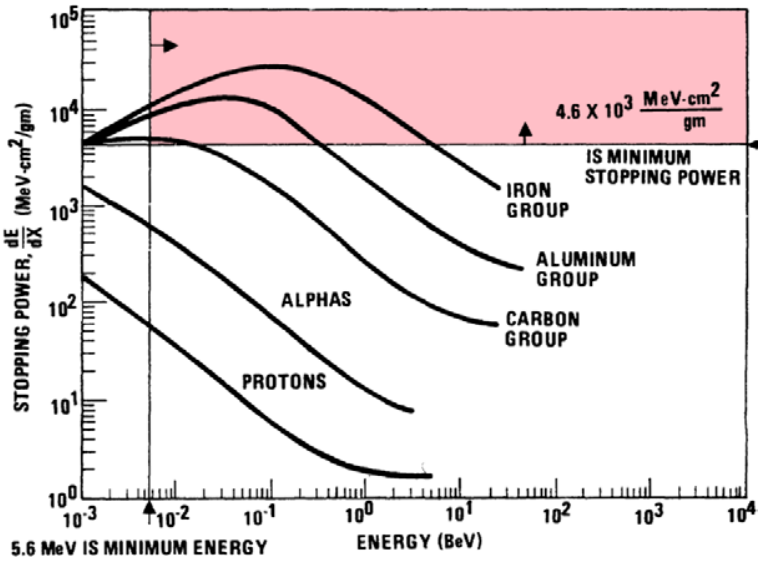


Fig. II-18 LET or Stopping power vs. particle energy. Minimum ion stopping power is defined by the aluminum shield thickness and the minimum ion energy is determined by the DRAM's critical charge and sensitive volume (adapted from [122]).

II-6. Upsets from Alpha Particles

A few months prior to Pickel and Blanford's report on the assessment of the soft error performance of 4K DRAMs in space, in April 1978, Timothy May and Murray Woods of Intel presented a paper at the Reliability Physics Symposium that rocked the commercial electronics world [124]. They demonstrated, for the first time, that failure rates in 4K and 16K DRAMs and were dominated by alpha particles produced by the natural radioactive decay of uranium, thorium, and related daughter impurities in the product packaging materials - "*a new physical mechanism for soft errors in dynamic RAMs and CCDs is the transient upset of dynamic nodes by heavily ionizing radiation from natural package materials surrounding the device*." A definitive "landmark" paper, it was the first to make the commercial semiconductor industry aware of the soft error problem in electronic devices, concluding that alpha particles were the causative agent. The authors also warned that this issue would be a persistent reliability threat and a significant challenge to continued technology scaling, "*because of the industry trend toward higher density memory devices with fewer numbers of electrons differentiating between logic states…this mechanism is expected to impose design and technology restrictions on future generations of devices, particularly at the 64K level and beyond*." The authors were also the first to introduce the concept of Q_{crit} as a metric for determining reliability performance to ionizing radiation events. "*Dynamic memories store data as the presence or absence of minority carrier charge on storage capacitors. Refresh is required to maintain the stored charge. For n-channel MOS RAM's and CCD's the charges are electrons…the number of electrons which differentiates between a '1' and '0' we define as the 'critical charge,' Q_{crit}.*" May and Woods also included a section on the packaging materials with the observation that "*reduction of alpha flux levels of the glasses, ceramics, and other materials used by the industry will be difficult because of the costs associated with purifying these materials and is unlikely, in any case, to result in the 2-3 orders of magnitude reduction in alpha flux needed for extremely dense memories of the future...inspection of alpha activity of materials will become a difficult matter in itself because of the low fluxes involved* [125]." The authors provided a plot

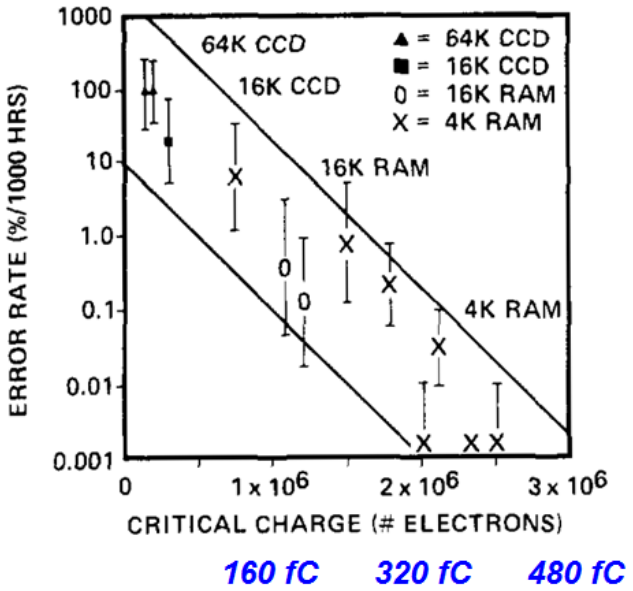


Fig. II-19 CCD and DRAM soft error rate as a function of the number of electrons and equivalent critical charge (in blue). This plot clearly shows that as the critical charge is scaled down with increasing integration density, the failure rate due to soft errors increases. Adapted from [124].

of (soft) error rate demonstrating that for a 2x reduction in Q_{crit} increases SER by more than two orders of magnitude as shown in fig. II-19.

This paper catalyzed activity in the commercial electronics world in three principle areas. First, it prompted several thorough studies of alpha particles in silicon focused on characterizing and modeling the dynamics of charge generation and collection as a function of the process, layout, and bias on silicon devices. Second, it prompted a huge effort in improving methods of alpha-particle metrology, developing manufacturing process and materials purification techniques to enable significantly lower alpha emission deemed necessary for future technology generations. In cases where materials could not be purified to hit a specified low alpha-particle emissivity, new low-alpha replacements were developed when possible, or, more commonly, methods of shielding or isolating sensitive die were employed. Third, it encouraged innovation in process design and layout specifically targeting increasing the Q_{crit} of devices, while often, correspondingly reducing their charge collection efficiency, reducing failure rates by making device technologies more robust against soft errors.

Within a year of the May and Woods report on alphas in DRAMs, there were several follow-on papers published describing experiments and simulations focusing on the properties and effects of alpha particles on silicon junctions. The first of these was a work by David Yaney, J.T. Nelson and Lowell Vanskike at Bell Labs [126], reporting that "*collection efficiency is also a function of incident particle energy and angle…most current 16K dynamic RAM'S do not lose information from the storage cells but rather are susceptible in sense amplifiers or bit lines. Because sensing is only done when the devices are active, the mode of operation in a system can affect the observed soft error rate.*" They considered sense-amplifier hits as dominant over cell hits, showing that soft errors due to cell hits were independent of duty cycle (they are sensitive virtually all of the time) while soft errors from sense amplifiers have a duty cycle dependence (since they are only sensitive during accessing). This was different than the conclusion from May and Woods that DRAM bit cells were the primary cause. They also observed the fact that near the maximum range of the alpha particle, where it is stopped in the silicon, the maximum charge density is generated (Bragg

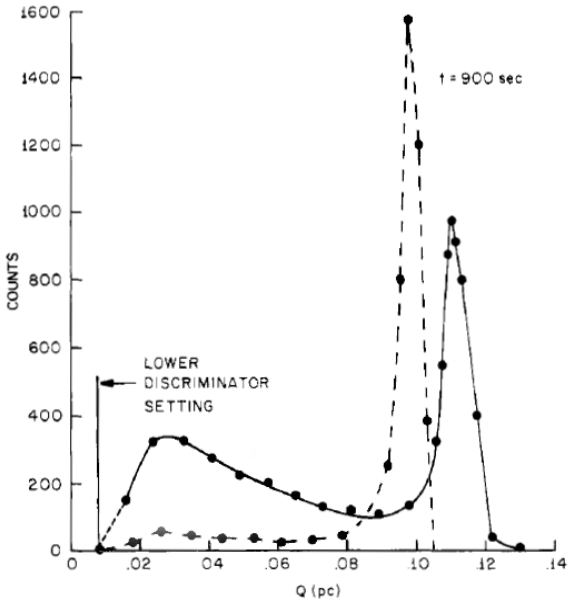
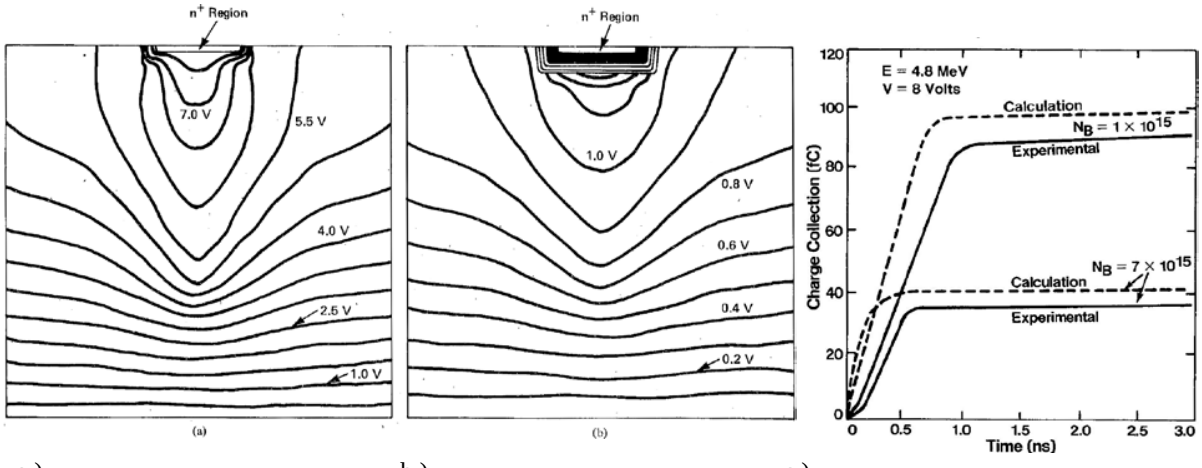


Fig. II-20 Histogram of charge collection for junction in a bulk substrate (dotted lines) and in a substrate with an epitaxial layer irradiated with 5MeV alpha particles. The well defined peaks are due to drift while the broad peak at lower collected charge is from diffusion. Adopted from [126].

peak)…"*one finds that only about 60 percent of the available charge in the track at normal incidence is actually collected…this follows from* (from the fact that) *the charge density is not uniform but is concentrated at the tail of the track…the highest charge density is, therefore, placed furthest from the collector high-field region and collected with smallest efficiency* (deep in the substrate). *The implication of this is that high-energy α's may not generate the greatest collected charge in device structures*." A final important observation in this work is that for junctions in bulk silicon, the majority of the charge collection was by the drift mechanism while for those created in an epitaxial layer, the diffusion component was much larger as illustrated in fig. II-20.

Another of the "landmark" works in this arena was a set of two papers by C. M. Hsieh, P. C. Murley and R. R. O'Brien, engineers at IBM, who considered the collection of charge by drift and diffusion from an alpha particle track in silicon junctions where they introduced the field funneling effect: "*The most important of these effects is that which we call field funneling. If the generated carriers cross a high field region such as the depletion region around a pn-junction, the presence of the carriers distorts the associated electric field and causes it to spread down along the track into previously field-free regions. The field will cause carriers in the track to be collected rapidly by drift rather than slowly by diffusion. After a few nanoseconds, the generated carrier density near the junction becomes comparable to the substrate doping, and the disturbed field relaxes back to its original position as the junction re-establishes itself. Hence the extent of this effect is strongly dependent on the substrate concentration* [127, 128]." The "funnel" shape is illustrated in fig. II-21 showing the warped equipotential lines a.) 0.1 ns and b.) 1 ns after the passage of a normally incident alpha particle. The two papers by Hsieh et al. made the case that the extension of the depletion volume by the funnel significantly increased the charge collected



a.) b.) c.)

Fig. II-21 Plot of the equipotentials [127] in an n+/p junction a.) 0.1 ns after, and b.) 1 ns after a 4.3 MeV alpha particle strike and c.) the collected charge as a function of time for n+/p junctions with two different doping levels [128]. Note that the more highly-doped sample collects much less charge than the lightly-doped sample.

by the junction. They also observed that the depth of the funnel, and hence the size of the collected charge,

was a strong function of the substrate doping. Junctions formed in more lightly-doped substrates suffered from a larger funneling effect and, hence, collected a larger amount of charge than those on highly-doped substrates (assuming the doping of the highly-doped side was the same). The effect of substrate doping on collected charge is clearly demonstrated in the plot in fig. II-21c. The junction in the more lightly-doped substrate collects more than 80fC while the junction in the more heavily-doped substrate collects less than 40 fC for the same alpha particle event. "*The funneling effect was observed in computer calculations and verified by experiment…a struck node gets most of the charge collected due to the funneling effect. The charge-sharing effect due to diffusion is small…a low resistivity substrate is superior in achieving a low SER* [128]." This work certainly served as a guide for more nearly two generations of process designers focused on developing implants and other doping modifications to reduce the collected charge from ionizing events. Although relatively short, one other paper was memorable in defining the circuit effects of alpha particles. Authored by Chenming Hu, from U.C. Berkeley, the paper describes a simple formula to calculate the length of the funnel [129]:

$$Collection\ Depth = \left(1 + \frac{\mu_n}{\mu_p}\right) W \qquad \text{(eq. II-1)}$$

The formula states that for an n+/p diode, the effective collection length (depth, if one assumes a normal hit) is the sum of the junction depletion width *W* and a term for the funneling depth, proportional to the product of the depletion width and the ratio of the electron and hole mobilities, μ_n and μ_p, respectively. Hu states that this charge collection based on this simple formula correctly predicted that the funneling depth increases as a function of $N_a^{-1/2}$ (the p-type acceptor doping level). It should be noted that for a p+/n junction, the mobility ratio in eq. II-1 is inverted (μ_p/μ_n). Hu mentions that the model predicts the effect of the angle of incidence on the current waveform and the funneling depth (collection length having a cost term) and that p+/n junctions will exhibit weaker funneling effect (because of the higher mobility of electrons). The method proposed by the paper was very effective, and I personally used it when helping design process modifications to our DRAM trench cell – many accelerated alpha-particle experiments were done that verified the accuracy of this simple formula. Hence, this was a simple but seminal work in providing researchers with a relatively simple calculation to determine the generated charge by integrating the *dQ/dx* over the calculated collection length or depth.

The May and Woods "bombshell" prompted a huge effort in improving methods of alpha particle metrology by encouraging the rapid development of manufacturing processes and materials purification techniques to enable significantly lower alpha emission deemed necessary for future technology generations. Within a year of the announcement of package alpha particles being a show-stopper, several works looking at package emission, characterization, and materials were published. Another work by the

Intel engineers E.S. Meieran, P.R. Engel, and T.C. May concluded the following "*designers will attempt to get as much reduction as possible in soft error rate out of the package fixes, to reduce their own requirements for design fixes. Hence, pressure will remain on package engineers to work with package vendors to reduce α flux levels, reductions by perhaps a factor of 5 to 10 may realistically be expected…in view of the low level of α activity required to keep soft error rates down, better radiation measurement techniques will have to be provided…very large area gas proportional flow counting, or sophisticated γ spectroscopy, will be the obvious choices* [130]." The authors summarized the techniques available at that time and their respective sensitivity levels as illustrated in table II-1. Clearly, at that time, detection of alpha emission lower than 0.01 α/hr/cm^2 was going to be a challenge. Over the next decade, most of these techniques were improved to provide 10-100x improvements in detection limits. J. A. Woolley, et al., from 3M, published a paper outlining the state-of-the-art in terms of analysis and materials and summarized those materials which would pose the biggest threat, "*only those materials opposite and adjacent to the active surface of a die need to be considered…the most critical materials are those which lie in a parallel plate relationship with the active die surface (i.e., substrates or covers)…package sealing materials are at least two orders of magnitude less critical than parallel plate sources because of their low viewing angles and smaller areas* [131]." Additionally the authors pointed out that "*formulating 99.9-percent alumina ceramic semiconductor packaging components with alpha fluxes less than 0.01 a/hr/cm^2 will require alumina raw materials with lower uranium and thorium contents than are currently commercially available*," and that in addition to materials development being required to get to lower alpha emission levels, the industry would need "*analytical techniques involving uranium and thorium separation and concentration in combination with direct alpha flux measurement methods.*" In that era, any material that was certified to have an alpha emission rate < 0.01 α/hr/cm^2 was considered "low alpha" and was sold at a premium. A 1979 paper by S. W. Levine, a chemist with Semi-Alloys, concerning the purification of package lid alloys and solder preforms is a noteworthy example of the type of activities being pursued by vendors to mitigate the alpha particle problem, "s*ince the gold plating on the lid is the dominant surface facing the semiconductor chip, it is of primary importance to reduce the thorium and uranium content of the gold as much as possible…the removal of thorium and*

Table II-1 Techniques for measurement of radioactive impurity content and alpha particle emission (circa 1979). After [130].

TECHNIQUE	SENSITIVITY	EQUIVALENT FLUX α/cm^2-hr.
Chemical:		
Emission Spectroscopy	1000 ppm	100
Mass Spectroscopy	10 ppb	0.01
Neutron Activation	10 ppb	0.01
Auger Spectroscopy	100 ppm	10-100
Electron Microprobe	0.1%	1000
Radiation:		
Alpha Scintillation	-	0.05
Gas Flow Counter	-	0.01
Gamma Spectroscopy	0.05 Pc/gm	0.001
Liquid Scintillation	-	0.01
Solid State Detectors	-	0.1
Film Track Counting	-	0.05

uranium...is achieved by the use of a chelating agent…the addition of the proprietary chelating agent to the electroplating bath again immobilizes the uranium and thorium ions and their radioactive isotopes so that they do not appear in the gold plated lid [132]." Levine implied that reductions beyond 0.01 α/hr/cm^2 were not only possible, but that "*we have received reports from several semiconductor manufacturers who have highly sophisticated measurement equipment and have measured single groups of samples taken from regular production as long as 500 hours. Their results on the gold plated lids measured indicate an activity level of between 0.01 and 0.003 count/h/cm^2*." However, characterizing materials having alpha particle emissions lower than the low alpha materials required special techniques. Levine's paper is notable in that he clearly identifies the proper statistical methods with the following introduction "*the mathematics for determining the counting time required to reach a desired level of confidence for radioactivity measurements are well known. However, using the mathematics for quality control purposes where predetermined activity level specifications at desired probability of detection levels is required is somewhat awkward in publications that have appeared in the past...since we are testing the presence of the activity of the sample exceeding a certain limit (the customer specification) it is imperative that the background alone does not exceed the threshold of an appreciable amount of time when the sample activity is not present.*" Unfortunately, the techniques outlined in Levine's work were largely ignored for the next decade by most material vendors reporting individual emission rates without any metric on the confidence or standard deviation. The problem with alpha counting is that since a signal is being subtracted from a background, if the background is too high and standard deviation/confidence intervals are not stated, it is easily possible to conclude that a material has zero emission when, in fact, it has a high alpha emission. By the late 1990s a new specification emerged based on materials purity specifications that claimed emission levels ten times lower than the low-alpha specification. Many vendors claimed to have ultra-low alpha materials, often specified to be < 0.001 α/hr/cm^2. At Texas Instruments we had a very active alpha metrology operation combining ultra sensitivity neutron activation analysis to characterize uranium and thorium impurity levels and large area alpha counting. Internally we sorted materials into three categories: standard (uncontrolled), low-alpha, and ultra-low alpha. Based on hundreds of alpha counting studies, it became obvious that if we required vendors to provide 0.001 α/hr/cm^2 materials, in general, they would not be able too. Therefore, internally at least, we designated materials that were < 0.002 α/hr/cm^2 (with a 90% confidence interval) as ultra-low alpha (ULA). We continued seeing vendors claim that their materials were < 0.001 a/hr/cm^2, but, upon obtaining and testing multiple samples from multiple lots, with very few exceptions, we found them to be > 0.001 α/hr/cm^2 (sometimes significantly so). The example shown in table II-2 exemplifies the problem. This is based on an actual alpha counting result from a material that the vendor claimed exceeded the Texas Instruments ULA specification of < 0.002 α/hr/cm^2 – they provided only a measured value (row 7). Upon further

discussions, we obtained the actual raw data of the alpha counting run. Using a 90% confidence interval, the limit of detection (row 10) given the short run time (24 hours) is actually 0.017 α/hr/cm^2, nearly 10x the ULA standard! The only thing that the vendor could actually claim for this material is that it is < 0.017 α/hr/cm^2. More recently, as part of an attempt to force the industry into adopting a standardized alpha counting technique, Texas Instruments and several other companies, developed the JEDEC JESD-221 test specification [133].

Table II-2. Results from a vendor's alpha-counting experiment that was not tested for a sufficient time. This material was claimed to have an alpha emission < 0.001 a/hr/cm^2.

1	Counter Active Area (cm2)	1000.0	cm^2
2	Sample Active Area (cm2)	100.0	cm^2
3	Background Count (B)	250	counts
4	Background Count time* (t_B)	24.0	hours
5	Gross Count (G)	252	counts
6	Gross Count time* (tG)	24.0	hours
7	**Average Alpha Emissivity**	**0.00093**	α/cm^2-hr
8	**Standard Deviation of Alpha Emissivity**	**0.01037**	α/cm^2-hr
9	**Confidence Interval**	**90.0**	%
10	**Detection Limit (at specified confidence level)**	**0.01706**	α/cm^2-hr

A decade after the May and Woods paper, their findings were so well-known that they had essentially become "gospel" among senior DRAM management and engineers – any observed soft errors had to be from alpha particles. Alpha particles were categorically the prevailing paradigm for soft errors in DRAM. I had been tasked with addressing the soft error problem for the Texas Instruments 16 Mbit DRAM program and developed a host of alpha-particle accelerated soft error rate (ASER) testing methods and flux simulation tools. Despite having improved the Q_{crit} (by specifying deeper trenches, thinner gate dielectrics) and reduced Q_{coll} (by specifying the use of higher substrate doping and an implant to reduce funneling effects at the bottom of trenches) such that ASER tests showed newer revisions to be millions-of-times less sensitive to alpha particles as compared original designs, the system SER (SSER) obtained from unaccelerated testing (requiring thousands of devices to be tested in parallel for thousands of hours), revealed no significant improvements in the soft error performance between the different versions. We had painstakingly quantified the alpha emission of all the DRAM materials with numerous neutron activation [134] and alpha-counting analyses and the sensitivity of the DRAMs to alpha particles with ASER testing so we could only conclude that either that there was a "hidden" source of alpha particles somehow being introduced during processing, or that alpha particles were categorically NOT the cause of the soft errors that we were observing (this later idea received strong resistance and ridicule from upper management since it went against the established and accepted paradigm that any observed soft errors had to be from alpha particles). I was systematically going through our process flows and had focused on the wet-etches as the most likely source of contamination. By serendipity, my supervisor at the time, Joe McPherson, had heard that phosphate rock, from which phosphoric acid is obtained, can contain high levels of alpha-emitting impurities (primarily

uranium and its radioactive daughter products). Since phosphoric acid was commonly used in semiconductor processing for the removal of silicon nitride layers, we thought we had our "smoking gun" in terms of explaining the higher-than-expected SSER. Alpha-counting and neutron activation analyses were performed to determine the impurity levels in various phosphoric acids and the extent to which these impurities were deposited on wafers during fabrication. As shown in table II-3 we found a difference of about 50x between two phosphoric acid processes. The impact of alpha-emitting impurities left by the phosphoric acid etching on production parts was ultimately determined by processing 1 megabit DRAM lots with four different phosphoric acids and measuring their actual SSER performance. While analyses revealed that, under certain conditions, wafers could become contaminated with alpha-emitting impurities from the phosphoric acid, the SSER tests performed on DRAMs processed with different phosphoric acids had the disappointing result that the concentration of alpha-emitting impurities in phosphoric acid had no measurable impact (table II-4). In light of the fact that there was nearly a 50x difference in the two acids used for processing, the lack of correlation between the acid impurity levels and the observed SSER data strongly supported that a non-alpha-related mechanism dominated the soft error rate in packaged devices. This discovery was kept as a trade secret with results published internally [135]. As with any technical discovery a similar discovery was made concurrently by other researchers [136]. Interestingly, Z. Hasnain and A. Ditali found that the phosphoric acid actually did affect their SSER rate. Ultimately our work with phosphoric acid and the other wet etches demonstrated that there were no "hidden" sources of alpha particle contaminations, and this left the fact that we were observing a relatively high soft error rate in DRAMs that was NOT dependent on the presence of alpha particles. Despite management's strong view that we had missed something or miscalculated, this "kuhnsian" crisis made the environment ripe for a new paradigm to emerge to explain the soft error rate outside the accepted view that it was alpha particles.

Table II-3. Alpha counting results of three Phosphoric acid splits used in SSER experiments.

Manufacturer	Alpha Flux (pCi/liter)
Process A	91
Process B	2

Table II-4. Results of SSER experiments on 1 megabit DRAM devices processed with various phosphoric acid splits.

Phosphoric split	# of units	device-hrs	errors	FR60 (fits)
Fresh Process A	2770	1.31×10^7	15	1276
Fresh Process B	980	4.61×10^6	6	1596
Japan Process C	596	3.17×10^6	5	1987

As a fall-back when material alpha emission cannot be reduced enough to attain specific soft error reliability goals, one approach is to deposit a material known to be low emissivity, on the die, prior to packaging, thereby blocking any alpha particles emitted by the packaging. In a paper presented at the 1981 IEEE IRPS (along-side the Hsieh paper on funneling), Malcolm L. White and three other engineers at Bell Labs reported that they did just that on a process that encapsulated SRAM devices in thick silicone

RTV rubber to eliminate alpha particles emitted from the package. "*A two mil (minimum) thick coating of silicone RTV rubber… which was determined to have an average alpha activity of .001 cm^{-2}hr^{-1}* (at least 10x lower than the best packaging materials of that era)…*over the active surface of an IC…has been used as an effective barrier to alpha particles emitted from the package materials…six hundred and eight 4K static RAMs were encapsulated with an RTV rubber coating and operated for 1655 hours…no soft errors occurred.* [137]." Based on their sensitivity measurements and the package alpha emission, the authors would have seen approximately 20 errors had a similar test been done with SRAMs with no RTV. The thick, RTV die-coating process, while effective, was never widely adopted by the industry.

The use of polyimides (a type polymer that is spin-coated and then heated to form a durable thin film on a chip prior to packaging it) eventually ended up in the production flows of many manufacturers as a mechanical stress relief layer and as an alpha particle shield, as described by Daisuke Makino, an engineer at Hitachi Chemical, "*semiconductor-grade polyimide resins are put on the market by several companies…although their structures are different, they show similar characteristics in terms of thermal, mechanical, and electrical properties…viscosity and solids content are controlled to get 2 to 4 μm-thick film by spin coating…(package) filler-induced stress is relaxed by employing a polyimide buffer coat on the chip…by forming a thin alpha-ray absorbing film on a surface of the chip…soft error can be prevented* [138]." The author continues with examples of 40-50 μm thick films of polyimide totally eliminating soft errors (fig. II-22a), and claims that on some devices, 10 μm is all that is needed to significantly reduce soft errors from alpha particles. Unfortunately, these thicker films are not feasible to manufacture in production since multiple polyimide depositions and cure cycles are required, so cycle-time becomes too long, and reliably creating bond pad openings becomes extremely difficult. Texas

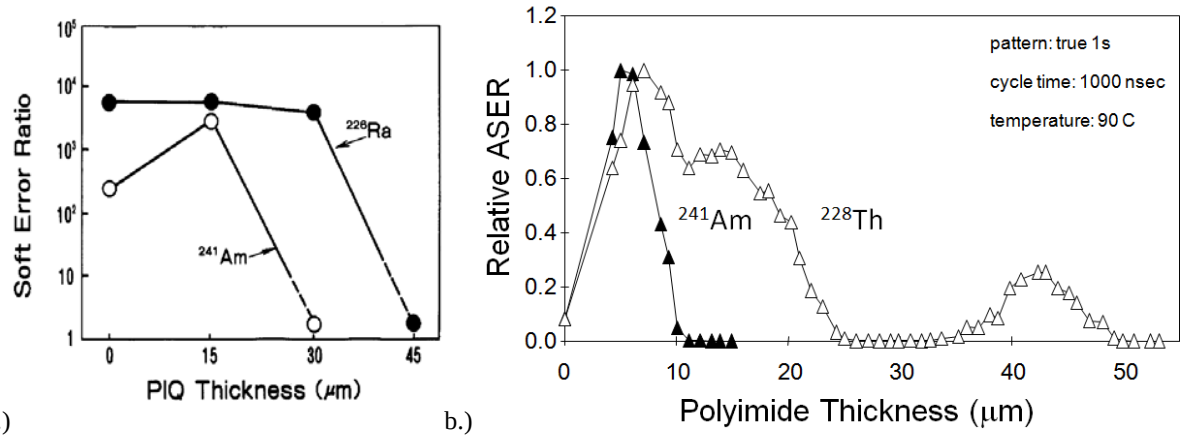


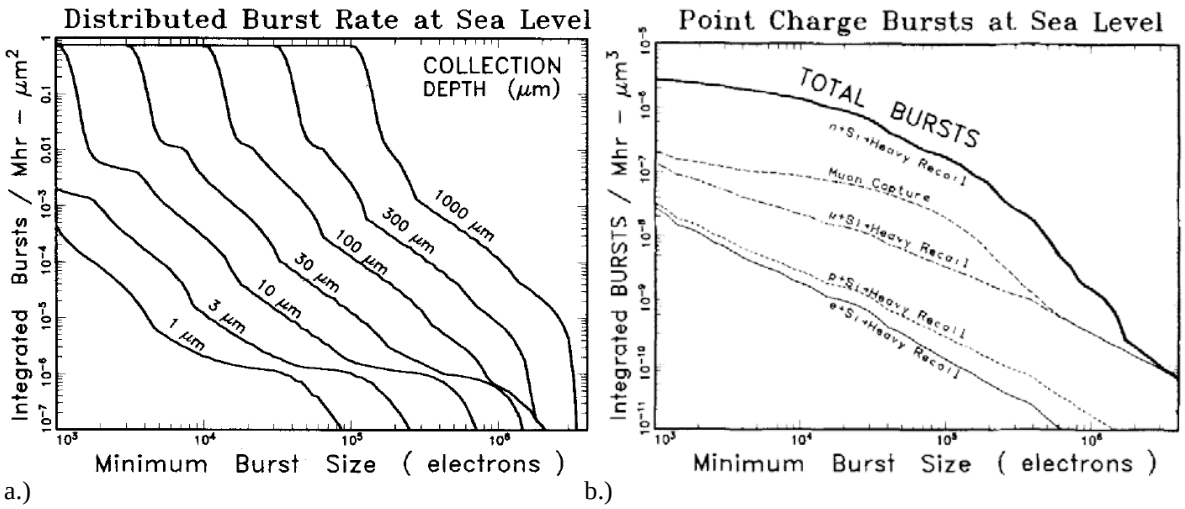
Fig. II-22 Plot of soft error vs. polyimide thickness a.) by Hitachi [138] and by b.) Texas Instruments [139]. Both investigated the same polyimide material. Note that the first study misses the peak due to higher alpha particle energies. Both plots show significant alpha particle reduction with polyimide layer of 45μm or more.

Instruments was using a production thickness of 4 μm but most of the wafers that I was first

characterizing for ASER were received without this coating – experiments revealed that the production thickness of polyimide actually made the SER worse than units with no shielding. Additional thickness of polyimide will have two effects – it will shield a larger number of high angle-of-incidence events, but at the same time, it will bring the Bragg peak of lower angle (closer to normal incidence) events closer to the active device layers. Hence the 4 μm of polyimide shielding ended stopping a greater amount of the alpha particles closer to the sensitive volume while the no-shield case allowed most of the alpha-generated charge to be harmlessly collected deeper in the substrate. We did an experiment exposing DRAM wafers with varying thicknesses of polyimide both ^{241}Am and ^{228}Th foil alpha particle sources while measuring their soft error rates [139]. As illustrated in fig. II-22b, the Thorium source covers a larger range of emission energies and, thus, is more representative of the energy range found in the actual packaging environment. Both sources reveal that having no polyimide shielding is far better than the baseline 4 μm of "shielding". The thorium source data suggests that to effectively reduce the SER by one order-of-magnitude, the polyimide would have to be ~ 40 μm thick, which, as stated earlier, could not be achieved with a viable manufacturing process. The only time shielding is an effective solution is if the barrier can block out a significant fraction of the alpha particles emitted from the packaging materials (it should be noted that comprehensive studies of chip materials compared to packaging materials has shown that the emission from chip materials is about 10x lower than that of most packaging materials).

II-7. Upsets from High Energy Neutrons

A year after the May and Woods paper on demonstrating alpha particles from the natural radioactive decay of uranium and thorium impurities as the dominant cause of soft failures, a "landmark paper" regarding the creation of soft errors by terrestrial cosmic rays at sea-level was published in the 1979 November issue of Science by James Ziegler and William Lanford [140]. As mentioned before, the first paper to suggest that cosmic rays would cause reliability problems in electronics was that of Wallmark and Marcus in 1962. However, that early claim was based on qualitative explanations and incorrectly predicted that cosmic-ray muons and gamma rays (not alpha particles) from the radioactive decay of impurities (in materials OUTSIDE of the packaged devices) would dominate the terrestrial upsets [141]. Ziegler and Lanford were the first to properly identify neutron reactions as the dominant source of ionization from terrestrial cosmic rays and to provide a quantitative assessment of its ability to cause soft errors as a function of the minimum burst size that could cause an upset in a device (its Q_{crit}). They were the first authors to demonstrate that cosmic ray effects were no longer limited to the space environment but could, indeed, be problematic for commercial electronics in the terrestrial environment. They summarized, "*we evaluated the magnitude of cosmic ray induced noise bursts in silicon and estimated the probability of memory cells being erroneously flipped to their opposite state (which is now widely called an electronic memory soft-error rate SER). We showed that because of cosmic rays most of the proposed memory circuit designs would not meet the traditional SER standard of 1/Mhr* [142]." The authors calculate the rate of charge generated, the "burst generation rate", induced by the various sea-level cosmic ray events in silicon. They divide the plethora of different events into two general types of mechanisms: direct ionization along charged particle wakes, and localized ionization from the nuclear reaction products (heavy recoils) assumed to be confined to small volumes, "*electrons, muons, and protons…leave behind weakly ionized wakes of charge…further, two possible nuclear reactions produce alpha particles…n+Si=>α and p+Si=>α which leave denser ionization wakes…alternatively, the particle may directly interact with a single silicon nucleus and cause a wide variety of heavy recoiling charged particles which…create a dense burst of electron-hole pairs*." By using published sea-level fluxes for the different cosmic ray particles and the various reaction cross-sections of said particles with silicon, Ziegler and Lanford created detailed plots of burst generation rate for both distributed ionization and point events as a function of device dimension and stored charge or Q_{crit}. These two plots are shown in fig. II-23a and b, respectively. This method, especially well-described and refined in their later work in the Journal of Applied Physics, was tremendously useful for ascertaining the failure rate of a specific device type. It should be noted that final calculated failure rate is a very sensitive function of the assumed "minimum burst size" to cause an upset (Q_{crit}) and to specific assumptions about the actual cell size and charge



a.) b.)

Fig. II-23 Integrated burst rate for a.) distributed events created by the passage of particles or nuclear reactions producing alpha particles, and b.) point charge bursts, localized disturbances caused by short range recoils as a function of minimum burst size. Minimum burst size corresponds to the minimum charge required to upset the device or Q_{crit}. Both after [142].

collection volume. In this paper the authors use a diffusion collection model first suggested by Kirkpatrick [143], which over-estimated the sensitive volume. This is important since the charge generated by alpha particle production (from the p + Si and n + Si reactions) dominates the failure rate at $Q_{crit} > 10^5$ electrons where 64kb DRAM operated. The generated charge also falls off very rapidly as the collection volume is decreased. An example of a failure rate calculation using the burst generation rate technique for one thousand 64kbit DRAM devices (8 Mbytes of memory) was provided with the result shown in fig. II-24. The authors concluded that neutron events would cause a failure rate of about one failure per day – this exceeds an average soft failure rate of 1000 FIT/device (the "standard" soft failure rate specification for DRAMs of that day) by a significant margin – Indeed Ziegler and Lanford's estimates predict an average soft failure rate of ~ 42,000 FIT/device! However, the author's assumptions about cell size and Q_{crit} are somewhat off. Based on literature of the time, commercially available 64k DRAMs had smaller capacitor areas, more shallow junctions than assumed, and Q_{crit} approaching 10^6 electrons [144 - 149]. Considering the larger Q_{crit} value only (ignoring the cell area and depth over-estimation), we get a very conservative

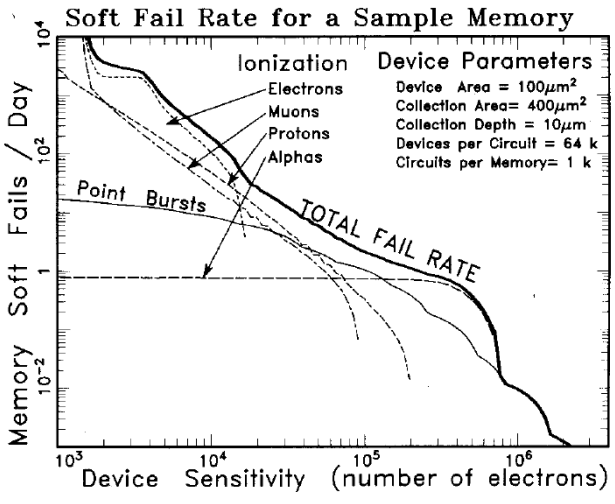


Fig. II-24 Predicted soft-error rate for one-thousand 64kb DRAM devices at sea-level, with each device having an area of about 100 μm^2. The authors estimated that the observed SER for these devices would be > 20x higher than typical SER of 1000 Fit/device. After [142].

estimate of ~ 400 FIT/device, assuming that the Q_{crit} is 10^6 electrons. Regardless of the assumption, Ziegler and Lanford's papers were revolutionary as they made clear that cosmic rays, specifically neutrons, would cause dominant reliability problems in electronics. Their predictions came true when, more than a decade later, various DRAM and SRAM manufacturers reported sea-level neutron effects limiting reliability at 16Mbit DRAM and 4Mbit SRAM nodes. It is intriguing that the neutron effects were not discussed as being a dominant failure rate mechanism for devices of that era (especially with the estimation of 42kFIT per part). In the end, while some were aware of the potential for cosmic rays to upset devices, as of 1981, the alpha particle was still considered the primary cause of soft failures in memory devices.

The first paper after Ziegler and Lanford to look specifically at sea-level terrestrial cosmic rays in commercial devices was a work by George Sai-Halasz at IBM published in 1983 [150]. Using simulations of a DRAM with minimum feature size of 1 μm (given this specification one can assume a 4Mbit device) the author came to the following conclusion, "*the Monte Carlo methods originally developed for obtaining α-particle-induced soft error rates (SER's) have been extended to calculate soft errors caused by cosmic rays. Specifically, the effect of the sea-level cosmic ray spectrum is calculated for a representative VLSI DRAM. The results show that the effects of sea level cosmic rays are minor when compared to the soft error rate caused by even moderate amounts of local radioactive contamination. At the same time, the precise cosmic soft error rate given here sets a limit on the possible reduction of the overall soft error rates by all hardware solution, including material purification.*" Sai-Halasz provides a good plot, shown in fig. II-25, showing this very effect where at high Q_{crit} the cosmic point bursts dominate, and in that region, the sensitivity to further Q_{crit} improvement is shallow. The work was in no way a paradigm-shattering statement against alpha particles. In fact, if anything, it implied that terrestrial cosmic ray events are only a "minor" effect, but the interesting statement is that if device parameters are tuned such that alpha particle effects are mitigated, then cosmic point bursts will limit any further improvements since the reduction in sensitivity to these events is nearly flat,

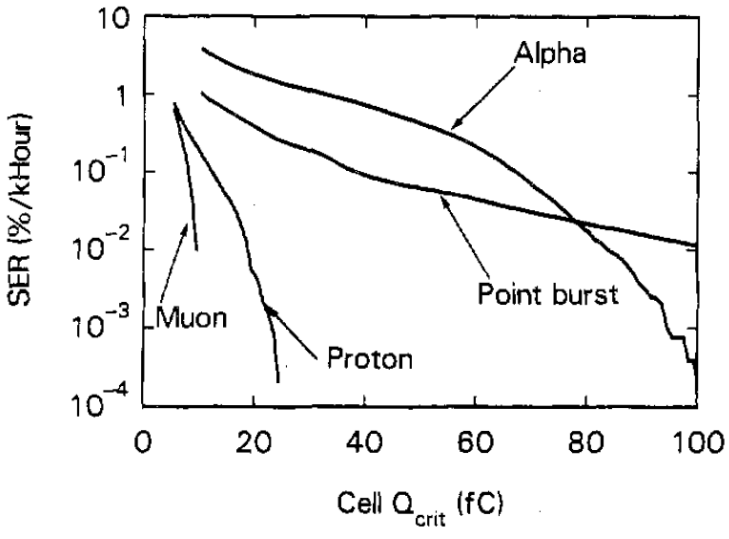


Fig. II-25 Plot of soft error vs. critical charge induced by alpha particles and various terrestrial cosmic ray components derived from Monte Carlo simulation. After [150].

so big improvements in Q_{crit} lead to only small improvements in SER. A decade later, empirical evidence would show that neutrons did, indeed, limit the reliability of devices whose sensitivity to alpha particles had been reduced or mitigated.

Several researchers from various companies were already focused on cosmic ray neutrons, and from 1993 to 1996 things came to a head – the days of alpha particles being considered the dominant cause of soft errors were over. The first landmark paper of this era was presented at the 1993 IEDM by Craig Lage et al. at Motorola, who reported "*a quantitative model which attributes most soft errors in dense SRAMs not to alpha particles as is commonly accepted, but to cosmic ray events. This work also elucidates for the first time the stored charge required in SRAM cells to achieve acceptable soft error rates* [151]." The work was impressive as the results were largely based on empirical tests from a large number (many millions of device-hours) of SRAM SSER experiments. The authors used several SRAMs designs and operated them at several different voltages to obtain a wide range of different Q_{crit} values. The assumption was that if the Q_{crit} were increased, the device sensitivity to alpha particles should drop off. They show that, indeed, this occurs, but only to an extent because the observed SSER saturates at a higher value than the alpha particle trend can explain. C. Lage and his colleagues attributed this anomalous failure rate to point-bursts (first described by Ziegler and Lanford) from the interaction of cosmic ray neutrons with silicon. Referring to the plot shown in fig. II-26, the authors explain that "*It is generally assumed that alpha particles…are the cause of the SSER events. If this were so, a linear correlation between these two types of measurements would be expected…these data represent more than 10 million device-hours of SSER testing…the assumption that alpha particles are the predominant source of the system soft errors provides a poor fit to the data. This discrepancy can be quantitatively explained by including in the SSER events the additional contribution of cosmic ray events.*" In other words, IF the observed SSER failures were due only to alpha particle events, then the plot of SSER as a function of alpha particle accelerated testing (ASER) should be linear – clearly this is not the case, and as Q_{crit} is increased (such that alpha particle events no longer dominate the failure rate), the data does not fit the linear expectation. Indeed, instead of approaching a zero failure rate as alpha sensitivity is reduced by many orders-of-

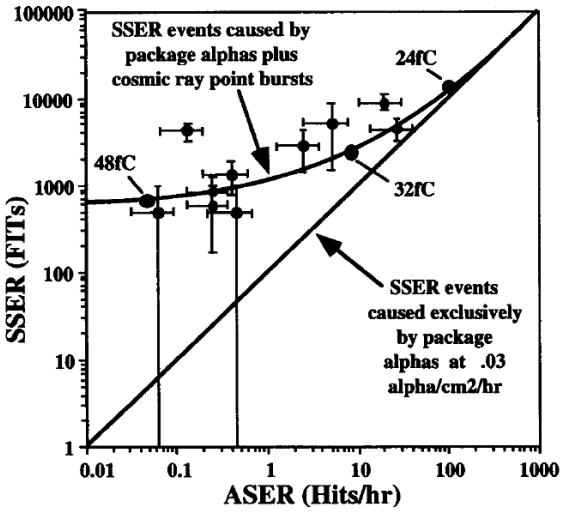
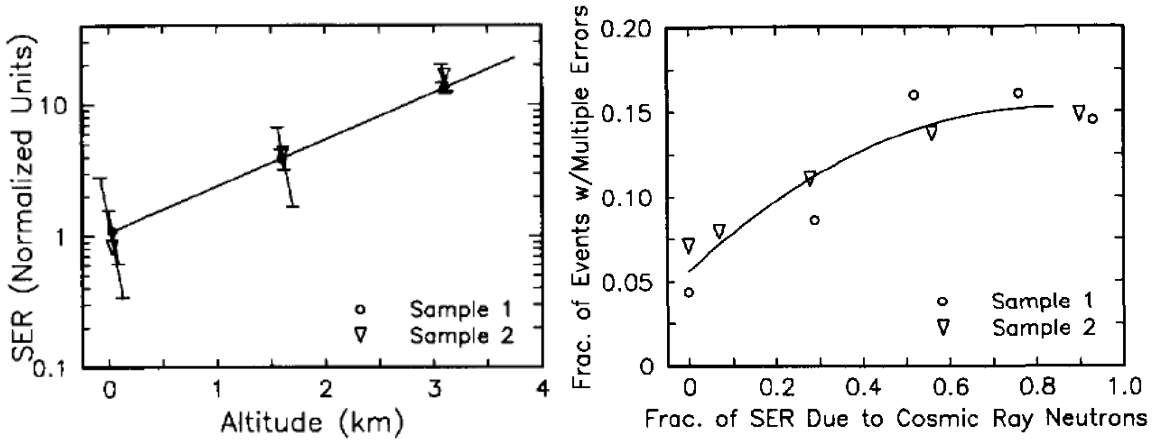


Fig. II-26 Plot of unaccelerated system soft error rate (SSER) data observed from long time testing vs. alpha-particle accelerated SER (ASER) test results. The fact that the SSER and ASER do not correlate at higher Q_{crit} implies another mechanism is dominating the SSER in this region. After [150].

magnitude, the left-hand side of the curve saturates to a non-zero failure rate, indicating that observed SSER fails in this regime are caused by another mechanism. As predicted by the simulation of G. Sai-Halasz, the cosmic ray neutron events are limiting the reliability of the SRAMs, NOT alpha particle events. The Lage paper is a very clear and empirical proof that alpha particles do not dominate the SSER of devices with sufficient Q_{crit}, and provides a credible argument that the observed failure rate is from cosmic ray neutrons.

The second blow to the alpha particle paradigm came just four months after the work of Lage et al. Tim O'Gorman, an engineer at IBM, makes a similarly definitive statement that "*conclusive evidence that cosmic rays cause soft errors in commercial dynamic RAM (DRAM) chips at ground level. Cosmic-ray-induced soft errors in electronic components have long been a problem for the designers of satellites and spacecraft, but they have not generally been considered to be an important influence on memory chip soft error rate (SER) in terrestrial environments…even at sea level there is a significant component of the SER that can be attributed to the effects of cosmic rays, and that the magnitude of the effects increases dramatically at higher altitudes* [152]." The data was persuasive as the tests were all SSER (unaccelerated) tests with thousands of units for thousands of hours. By testing at different locations with very different altitudes and neutron fluxes, O'Gorman clearly demonstrated that once above sea-level the



a.) b.)

Fig. II-27 Plot of a.) SSER for DRAMs tested at various altitudes and plot of b.) the soft fails that were fraction multi bit as a function of the fraction of the upset due to neutrons (altitude). After [152].

observed failures were largely due to neutron events as illustrated in fig. II-27a. Note that the uncertainty (denoted by the error bars) decreased at higher altitudes because the number of observed errors increased at the higher elevations. He also made the observation that the fraction of events that were multi-bit upsets (MBUs) increased as the fraction of them that were from neutrons increased (fig. II-27b). Since most error correction schemes employ single–error correction, the presence of multiple-bit errors defines the lower limit of the failure rate. O'Gorman showed that the MBU fraction increased nearly 3x from ~ 5% at

sea-level to 15% at 3 km where almost all the events were due to neutrons. The Lage and O'Gorman papers were both landmark papers that were directly responsible for changing the paradigm that the SER observed in commercial electronics in the terrestrial environment was solely an alpha-particle problem. A demonstration that many major companies were studying neutron effects, Texas Instruments came to the same conclusions with very similar experiments done at various altitudes, and Fujitsu used nuclear and device simulations to ascertain that neutrons dominate the observed SER. These results followed on the heels of Lage and O'Gorman and were reported at the 1996 IEEE IRPS [153] and VLSI symposium [154], respectively. By now the alpha-particle "spell" had been broken, and the establishment paradigm over-turned. Alpha particles were no longer considered the only cause of soft errors.

As is usual with any "revolutionary" or "extraordinary" science, it takes some time for the news to get around. As late as 2000, a major manufacturer was still circulating DRAM data sheets claiming outrageously low SSER for its DRAM. The claim was erroneous on two points: first, the unaccelerated field data was not run long enough – indeed, given the short reported run time and number of units, DRAMs with a mean SSER greater than 5000 FIT would have exhibited the reported zero failures the majority of the time (the target soft fail specification at that time was an SSER of ~1000 FIT for DRAMs, so virtually all commercially available DRAM would have shown zero soft errors under such a inadequate test). Second, and more glaringly, despite the wealth of published literature available for a half-decade demonstrating that DRAM SSER was currently dominated by cosmic ray neutrons, this manufacturer showed extrapolated failure rates from accelerated alpha SER results, again meaningless since DRAM of that era were basically nearly impervious to alpha particles. This data sheet, and a few other examples like it, were one of the driving forces behind making sure that the JEDEC JESD89/89A specification clearly stated that one could NOT report SER results based on a single mechanism and claim low failure rates. The following text is included in similar form in each chapter on accelerated testing for each of the three unique soft error mechanism to ensure that users understand that all mechanisms must be accounted for when reporting a product's failure rate in the terrestrial environment - "*High-energy neutron SER data cannot be used to predict alpha particle or low-energy neutron (i.e., thermal neutron) cosmic-ray induced failure rates. Conversely, neither can alpha particle nor thermal neutron SER data be used to predict high-energy neutron-induced failure rates. An overall assessment of a device's soft error sensitivity is complete ONLY when the alpha, high-energy neutron AND thermal neutron induced failure rates have been accounted for*."[155]

II-8. Upsets from ^{10}B and Thermal Neutrons

The ^{10}B(n, α)^{7}Li reaction was well known in nuclear circles, but somehow its impact was either over-estimated or ignored completely during the first few decades of semiconductor development. High concentrations of ^{10}B were incorporated into semiconductor products as borosilicate dielectric layers, until this practice led to the spectacularly public failure of a new series of enterprise server manufactured by SUN Microsystems [156]. One of the first mention of this reaction causing problems in semiconductor devices was a paper published by I. Arimura and C. Rosenberg of Boeing [157], demonstrating thermal neutron dose-related bipolar gain degradation in PNP bipolar devices that they attributed to the ^{10}B(n, α)^{7}Li reaction. As illustrated in fig. II-28, they found that "*high energy alpha recoil then penetrates into the base producing damage in that region and, consequently, degrades the gain of the device. The lithium recoil, if generated close to the base-emitter boundary, is also capable of producing considerable damage...the gain degradation is approximately linear with neutron fluence similar to that observed for fast neutron irradiations indicating that bulk damage effects are probably the cause of the observed degradations.*" Although not an issue at the relatively low terrestrial neutron fluxes, this paper already points to the concern with the presence of an unstable (to neutrons) isotope of boron of the primary p-type dopant in silicon. Interestingly, in one of his numerous follow up papers [158], May discusses the ^{10}B(n, α)^{7}Li reaction as a potential soft error concern only to dismiss it as negligible because of two errors in his cross-section calculation.

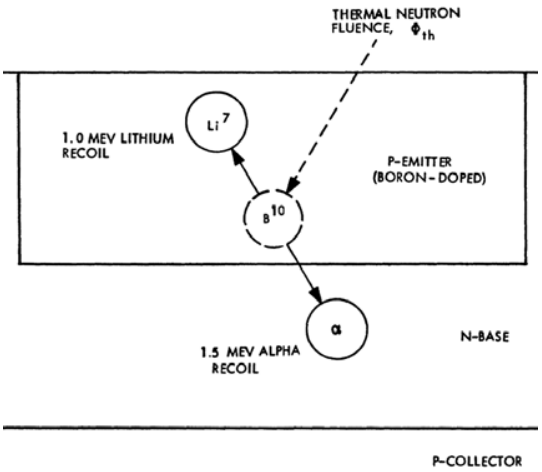


Fig. II-28 Diagram in the first work demonstrating the effect of thermal neutrons on semiconductor devices. The authors reported bipolar gain degradation induced by defects from the recoils of the ^{10}B(n, α)^{7}Li reaction as a function of neutron dose. Adapted from [157].

A decade after the Arimura and Rosenberg paper, a good theoretical treatment was provided by Robert Fleischer [159] of General Electric, concluding that the ^{10}B(n, α)^{7}Li reaction would, indeed, be a problem for silicon devices – stating that the "*most sensitive location was in the p-type regions close to junctions in silicon semiconductor devices…Packaging of devices, particularly for space applications, is often in ceramic packages which are hermetically sealed with borosilicate glass that contains 7% to 12% boron by weight…in some cases are not shielded from direct view of the circuit elements and hence could inject ^{4}He or ^{7}Li…*" and that these "*heavily ionizing particles…can contribute to soft errors in integrated circuits. The effect from boron exceeds that from cosmic rays incident on silicon nuclei if the boron*

concentration exceeds 10^{20}/cc, which is often present in p-type Si and which is exceeded in common packaging sealing glasses." This paper was a well-thought out diatribe on the ^{10}B issue, warning of the effects of the ^{10}B(n, α)^{7}Li reaction. The conclusion regarding glass frit used in hermetic packages, while not incorrect, turns out not to be a major issue for commercial electronics since they are packaged predominantly with plastic encapsulants. Even in ceramic packages, integrated circuits soon started being fabricated with more and more metallization and dielectric layers making it much less likely that the lithium and helium recoils would have enough energy to reach the active devices. In most technologies, particularly MOS technologies, substrate doping was reduced several orders-of-magnitude below 10^{20} cm^{-3}, thus minimizing this source of ^{10}B as a significant contributor to soft errors. Fleischer clearly demonstrated that a substrate doping density of 10^{20} cm^{-3} created an SEU rate of the same order as that from high-energy neutrons. However, as illustrated in fig. II-29, the range of typical substrate doping for advanced CMOS technologies, shown in green, is substantially below this – implying that reactions from the substrate are bound to contribute to a few percent at most. So, while Fleischer's work is unquestionably a landmark work in terms of providing a clear and very early warning to manufacturers about the risk associated with the ^{10}B(n, α)^{7}Li reaction and the use of boron compounds in semiconductor processes, the two sources that he was concerned about were, in general, negligible contributors to product failure rates.

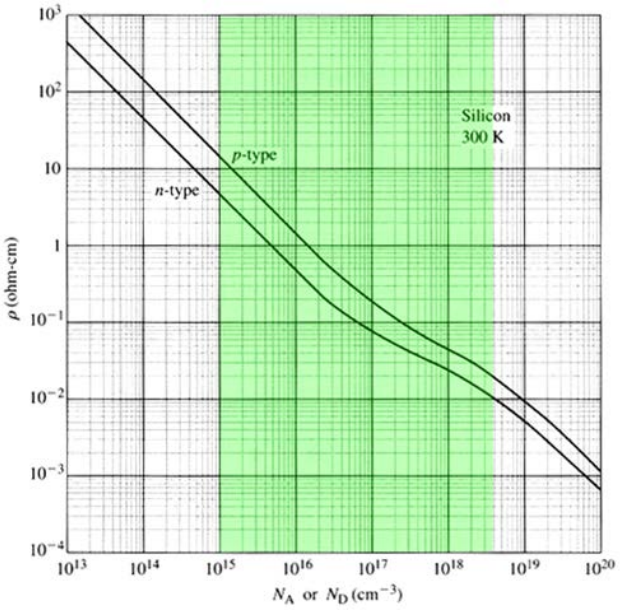


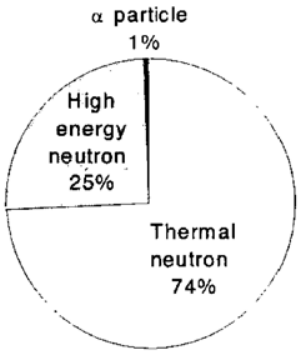
Fig. II-29 Silicon resistivity as a function of dopant density. Note the green region delineates the range of substrate doping commonly used in commercial MOS devices – SEU from thermal neutrons and ^{10}B substrate concentrations will be at least an order-of-magnitude below that generated by high-energy neutrons.

From 1990 to 1992 Tim Hossain and I, were working together on high-sensitivity neutron activation analysis to determine levels of uranium and thorium impurities in an ongoing effort to find the elusive "hidden" source of alpha particles responsible for DRAM SSER (blindly assumed to be the cause because the established paradigm was that alpha particles were the sole cause of SSER in DRAMs as mentioned earlier). We soon realized that ^{10}B and thermal neutrons might be the source of the DRAM soft errors. The three factors in favor of this theory were as follows: highly-doped (5-8%) boro-phosphosilicate glass (BPSG) was ubiquitous (due to its gettering and low-temperature planarization properties) as a thick dielectric layer close to the silicon interface in commercial semiconductor products. ^{10}B had an extremely

high neutron capture cross-section as compared to other semiconductor materials and represented nearly 20% of the BPSG doping. Finally, and most importantly, unlike most neutron capture reactions where excited nuclei release gamma photons, ^{10}B produced two highly ionizing recoil products, each capable of producing a soft error [160, 161]. Ironically, the effect was not dominant in DRAM devices that spurred the discovery since most had been optimized to have bit cell critical charges far in excess of the charge generated by either of the recoil products (in addition, our initial estimation of the effect was over-estimated because the thermal neutron flux reference we had used was about 10x higher than the actual flux, later reported by Dirk et al. [162]). After five years of experiments with SSER and a change in focus from DRAMs to SRAMs, I did additional work with E.B. Smith, demonstrating that this effect was a concern for SRAMs and, in fact, THE dominant soft error mechanism in devices with BPSG. We concluded that SER from thermal neutrons and reactions in the BPSG was about 4x the other mechanisms depending on the voltage of operation and the amount of boron doping in the BPSG [163]. A table included in the paper comparing the two CMOS technologies with and without BPSG is shown in fig. II-30a. As a result of this work, Texas Instruments moved away from the use of BPSG in its advanced technologies. In later (unpublished) work on more advanced technologies operating at lower voltages, the effect was seen to be as much as 8x higher than the other SER mechanisms. The concept of using a ^{11}B enriched source material to eliminate the ^{10}B soft error problem was patented along with a shielding solution using high-thermal neutron cross-section material in the chip encapsulant (conformal) or at the board level [165,166]. Our result were reproduced in similar technologies by H. Kobayashi et al. at Sony [164]. These authors confirmed that the $^{10}B(n, \alpha)^{7}Li$ reaction in the BPSG was the dominant source of soft errors in submicron SRAM technologies. In their assessment the SER from BPSG and thermal neutrons was 3x higher than that from high-energy neutron reaction and alpha particles as shown in fig. II-30b. The result of all these studies was to ultimately convince manufacturers to remove BPSG from semiconductor processes that were sensitive to soft errors. The net result of this discovery was a direct 3x to 8x reduction in product failure rates – one of the biggest reliability improvements afforded by a single process change.

Component	0.25μm SRAM	0.18μm SRAM
Alpha particles	4%	18%
High-E Neutrons	15%	82%
^{10}B Fissions	81%	0%
Total SER (a.u.)	7.5	1.0

a.)



b.)

Fig. II-30 The first report of the effect of BPSG on SRAM a.) was demonstrated to be the dominant cause of soft errors and about 4x larger than the other mechanisms at the 250 μm node [163]. This result was later confirmed [164] with SSER measurements and reported b.) to be about 3x larger than the other mechanisms.

As with the phosphoric acid process that has been shown to potentially introduce alpha-emitting impurities into semiconductor products, the use of diborane source gas (or any other boron containing agent) has been scrutinized prior to its inclusion in production flows. When Texas Instruments was developing its tungsten contact metallization flow in the late 90s for ensuring adhesion of thin TiW sticking layers for W plugs, we did extensive alpha counting and materials analysis to ensure that the process would not be plagued with high concentrations of ^{10}B left behind by the process. The primary concern is that diborane or other boron-based gas processes includes a natural abundance of the isotope ^{10}B (unless it has been intentionally removed). ^{10}B diffused into the chip dielectric and/or metal layers or incorporated as borides might be in high enough concentration to pose a soft error risk. In fact this very problem is claimed to have been observed by S. Wen et al. at Cisco Systems [167] that did thermal neutron assessments and SIMS in various SRAM technologies as shown in fig. II-31. The authors claim that in SRAM samples on certain technologies "*the thermal neutron SER rate is 100-300 FIT/Mbit…comparable to the high-energy neutron failure rate. As technology scales, the impact of B^{10} to SER rate may increase. Plus, if the B^{10} source is from a fab process material that is not B^{10} controlled, the amount of B^{10} contamination may vary lot to lot. Based on B^{10} location from SIMS result, we looked at the processing steps…to identify the possible B^{10} source. We found that indeed in advanced technologies, many fabs are using B_2H_6 based material for W plug liner layer…If this B_2H_6 based material is not B^{11} purified, then it will introduce B^{10} during this liner process step.*" The lesson here is that while the whole industry, with few exceptions (some companies avoided BPSG SER by serendipity, having replaced BPSG with PSG due to adopting CMP), made a conscious decision to remove BPSG from their processes, without vigilance, during the development of new process technologies, it is always possible that another high-density source of boron may be introduced, so new process flows should be scrutinized – this is particularly true of companies working with foundry processes.

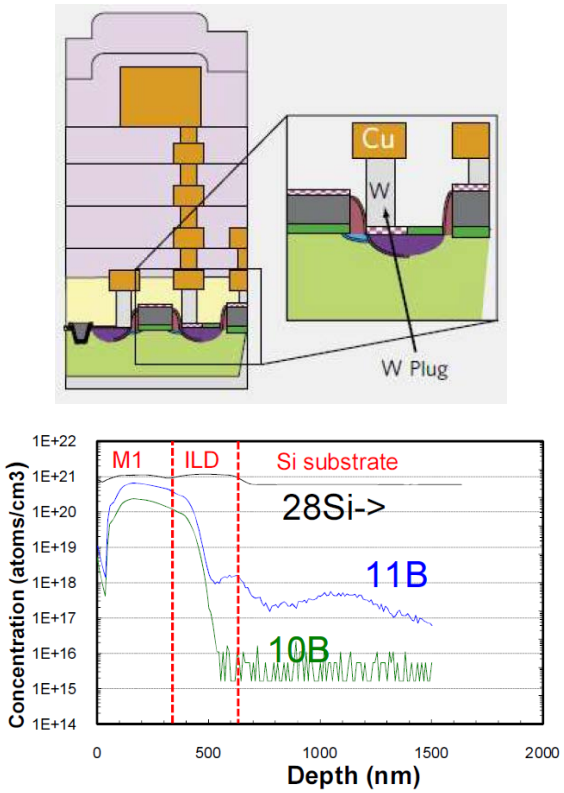


Fig. II-31 a.) Diagram showing the location of the tungsten plugs. The process for forming a sticking layer uses diborane gas and apparently leaves a high concentration of ^{10}B, as illustrated by the b.) SIMS data. The 10B concentration was apparently high enough to cause an SEU rate of hundreds of FIT/Mbit. Both images from [167].

II-9. Summary

In closing, I hope this historical retrospective has been an illustrative example of the way in which science progresses as first proposed by Thomas Kuhn more than 50 years ago. He claims that science does NOT progress as a cumulative accumulation of facts and theories as has been traditionally posited. Rather, Kuhn paints a picture of a conservative scientific establishment resistive to new ideas, if not openly hostile, until enough data has been accumulated proving that the new paradigms are so much better than the established ones that they must ultimately replace them. Thus, in discontinuous fits and starts does science progress, with tumultuous periods of "extraordinary science" (where old ideas are actually jettisoned to make room for better new paradigms) followed by periods of advanced based on "normal science". I have shown several examples of parallel work being done by researchers unaware of each other's progress, often leading to reinvention (the transistor for example). In any scientific endeavor, there will be early (pre-paradigm shift) discoveries that are often ignored or sidelined until an understanding of their true nature has had time to percolate through the "system" (Like Wulf and Pacini's earlier experiments and conclusions hinting about the existence of cosmic rays, only later "accepted" as truth after the balloon experiments of Victor Hess).

I have tried to relate how the study of SEEs led to the acceptance of new paradigms for circuit reliability and how these paradigms shifted with discontinuous bursts of activity. With the development of soft errors and SEEs, upsets in space hinted at the sensitivity of electronics but it was not until the May and Woods paper of 1978 that people fully appreciated that terrestrial radiation could dominate semiconductor reliability. Once entrenched however, the concept that something other than alpha particles could dominate SER was resisted. It was not until 15 years later, with an over-abundance of proof from multiple researchers, that cosmic neutron events were accepted as a dominating SER mechanism – despite the Zeigler paper clearly showing neutrons to pose a threat to electronics. In that same era the problem with BPSG and cosmic neutrons was also discovered, but, like the Zeigler paper, the finding was ignored by most of the industry until thermal neutron effect had a stunning impact on the SRAM-caches in high-end servers. Recent work studying the effects of muon SER imply that this may become a dominant reliability concern for future technologies as critical charge continues to drop (interestingly predicted as a dominant mechanism limiting technology scaling in 1962).

In closing I would like to remind the "youngsters" in the audience (those having reached their young adult-hood in the age of the internet) that most of the science and engineering discussed in this short course was done "B.C." - Before Connectivity. In B.C. times, where printed references had to be painstakingly looked up in volumes of bound physical journals, the scope of the search was limited by a researcher's time and the quality and breadth of the library in which the researcher was doing his/her research – in short, it was easy to miss crucial papers that would have speeded up progress or redirected

research in more fruitful direction. I can personally attest to having missed some pretty fundamental work that would have shaved years off my research had I found such critical references at the time. In the current, A.C. (After Connectivity) age of instant gratification, huge numbers of references can be searched by keyword across multiple journals, the majority being instantly accessible, enabling current and future researchers an unheard of advantage to be better-acquainted with past work and contemporary developments while not wasting time on larks (papers that sound really cool from the abstract or title but simple don't live up to expectations). So my advice to those wanting to make the biggest breakthroughs, either pick an area that is less popular (like Marie Curie is said to have done) so that there are fewer reference works to slog through, or do a thorough literature search, now made vastly simpler by the huge number of works online that can be accessed by simple keyword searches.

Given the limited time to cover this material, I have chosen to focus on the aspects of SEE with which I was directly involved in, the discovery and elucidation of new mechanisms that defined the soft error rate and the characterization and processing technology tweaks done to mitigate the problems in commercial CMOS technologies. I have also limited the focus of this discussion to digital commercial technology despite the wealth of material and effects in bipolar technologies. However since digital CMOS technology has been a key driver for a large fraction of the landmark innovations in electronics functionality I do not think this to a major flaw. My biggest regret though, and a much larger omission in my opinion, is that fact that I had to ignore the huge body of excellent work on simulation of SEE that has evolved from simple charge collection estimations of the 60s and 70s to full-on event simulation where specific nuclear reaction products injected into dynamic and self-consistent 3D models of a silicon device embedded in a circuit response simulation. SEEs are so complex that simulation is often the only tool capable of revealing essential details useful for hardening a design or simply gaining insight on the aspects that matter.

References

[1] G. E. Moore, “Cramming More Components onto Integrated Circuits”, *Electronics*, 38(8), pp. 4-7, Apr. 19, 1965.

[2] Daniel Lyons, “Sun Screen”, *Forbes.com*, Nov. 13, 2000, (http://www.forbes.com/forbes/2000/1113/6613068a.html)

[3] SUN senior management, Private communication, May 2001.

[4] “In-flight upset, 154 km west of Learmonth, WA, 7 October 2008, VH-QPA Airbus A330-303,” ATSB Transp. Safety Report - Aviation Occurrence Investig., AO-2008-070, pp. 1 – 313, Dec. 2011.

[5] W. Cui, M. Li, X. Luo, J.L. Drewniak, T.H. Hubing, T.P. VanDoren, and R.E. Du Broff, “Anticipating EMI from coupling between high-speed digital and I/O lines”, *IEEE International Symp. On Electromagnetic Compatibility*, Vol. 1, pp.189 – 194, Aug. 1999.

[6] B.R. Farris and J.P. Rohrbaugh, “Recommended EMI test procedures for MMICs”, *IEEE Internat. Symp. On Electromagnetic Compatibility*, pp. 118 – 123, Aug. 1990.

[7] A. Vittal and M. Marek-Sadowska, “Crosstalk reduction for VLSI”, *IEEE Trans. on Computer-Aided Design of ICs and Systems*, Vol. 16(3), pp. 290 – 298, Mar. 1997.

[8] K. Aingaran et al., “Coupling noise analysis for VLSI and ULSI circuits”, *IEEE Symp. On Quality Electronic Design*, pp. 485 – 489, Mar. 2000.

[9] S.H. Russ and C.O. Alford, “Coupled noise and its effects on modern packaging technology”, *IEEE Trans. on Components , Packaging, and Manufacturing Technology*, Vol. 13, No. 4, pp. 1074 – 1082, Dec. 1990.

[10] J.W. McPherson, “Reliability Trends with Advanced CMOS Scaling and the Implications for Design,” *IEEE Custom Integrated Circuits Conf.*, pp. 405 – 412, 2007.

[11] M. White, “Scaled CMOS reliability and considerations for spacecraft systems: Bottom-up and top-down perspectives,” in *Proc. IEEE Internat. Rel. Physics Symp*., pp. 4B.4.1 - 4B.4.5, 2012.

[12] E. L. Peterson, “Nuclear Reactions in Semiconductors”, *IEEE Trans. on Nucl. Sci.*, 27(6), pp.1494-1499, Dec. 1980.

[13] E. S. Meieran, P. R. Engel, T. C. May, “Measurement of Alpha Particle Radioactivity in IC Device Packages”, *in Proc. IEEE Int. Reliability Phys. Symp.*, pp. 13 – 22, 1979.

[14] J. F. Ziegler and J. P. Biersack, *SRIM - 2008 software*, downloaded from (http://www.srim.org/).

[15] J. F. Ziegler, J. P. Biersack, and U. Littmark, “The Stopping and Range of Ions in Matter”, Vol. 1, N.Y., 1985. Pergamon.

[16] M. P. King, et al., “Radial characteristics of heavy-ion track structure and implications of delta-ray events for microelectronics”, *Appl. Physics Letters*, 101(5), pp. 53509-53509-3, 2012

[17] M. T. Robinson, Mark; I. M. Torrens, "Computer simulation of atomic-displacement cascades in solids in the binary-collision approximation", *Phys. Rev. B*, 9(12), pp. 5008-5024, Jun. 15, 1974.

[18] J. R. Key, T. A. Rabson, “Average energy expended per ionized electron-hole pair for α particles in Si(Li)”, *IEEE Trans. on Nuclear Sci.*, 18(1) , Part: I, pp. 184 – 186, 1971.

[19] 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*, EDL-2, pp. 103-105, Apr. 1981.

[20] C. Hu, “Alpha-Particle-Induced Field and Enhanced Collection of Carriers”, *IEEE Electron Device Letters*, EDL-3, No. 2, pp. 31-34, Feb. 1982.

[21] C. M. Hsieh, P. C. Murley, and R. R. O’Brien, “Dynamics of charge collection from alpha-particle tracks in integrated circuits,” *Proc. IEEE Int. Reliability Phys. Symp*., pp. 38–42, 1981.

[22] D. S. Yaney, J. T. Nelson and L. L. Vanskike, *IEEE Trans. Electron Devices*, ED-26, pp. 10-16, Jan. 1979.

[23] S. Kirkpatrick, “Modeling diffusion and collection of charge from ionizing radiation in silicon devices,” *IEEE Trans. Electron Devices*, ED-26, p. 1742-1753, Nov. 1979.

[24] P. E. Dodd and L. W. Massengill, "Basic mechanisms and modeling of single-event upset in digital microelectronics," *IEEE Trans. Nucl. Sci.*, 50(3), pp. 583-602, Jun. 2003.
[25] F. W. Sexton, "Destructive single-event effects in semiconductor devices and ICs," *IEEE Trans. Nucl. Sci*., 50(3), pp. 603–621, Jun. 2003.
[26] S. Satoh, Y. Tosaka, and S. A. Wender, "Geometric effect of multiple-bit soft errors induced by cosmic ray neutrons on DRAM's," *IEEE Electron Device Letters*, Vol. 21, pp. 310 - 312, Jun. 2000.
[27] J. Maiz, S. Hareland, K. Zhang, and P. Armstrong, "Characterization of multi-bit soft error events in advanced SRAMs," *IEDM Tech. Dig*., pp. 21.4.1 - 21.4.4, Dec. 2003.
[28] O. Musseau, et al., "Analysis of multiple bit upsets (MBU) in a CMOS SRAM," IEEE Trans. Nucl. Sci., vol. 43, pp. 2879–2888, Dec. 1996.
[29] X. Zhu, et al., "A Quantitative Assessment of Charge Collection Efficiency of N+ and P+Diffusion Areas in Terrestrial Neutron Environment", *IEEE Trans. Nucl. Sci.*, Vol. 54, pp. 2156 – 2161. Dec. 2007.
[30] R. Koga, et al., "Single Event Functional Interrupt (SEFI) sensitivity in microcircuits," *Proc. 4th RADECS*, pp. 311-318, Sept. 1997.
[31] S. M. Guertin, et. al., "Dynamic SDRAM SEFI Detection and Recovery Test Results", *IEEE Rad. Effects Data Workshop*, pp. 62-67, 2004
[32] J. Benedetto, et al., "Heavy Ion-Induced Digital Single-Event Transients in Deep Submicron Processes," *IEEE Trans. Nucl. Sci.*, vol. 51, pp. 3480-3485, Dec. 2004.
[33] G. Bruguier and J-M. Palau, "Single Particle-Induced Latchup," *IEEE Trans. Nucl. Sci*., vol. 43, pp. 522-532, April 1996.
[34] P. E. Dodd, et al., "Neutron-induced latchup in SRAMs at ground level," *Proc. IEEE Int'l Rel. Physics Symp. (IRPS)*, April 2003, pp. 51-55.
[35] M. A. Clemens, et al., "The Effects of Neutron Energy and High-Z Materials on Single Event Upsets and Multiple Cell Upsets," *IEEE Trans. Nucl. Sci.*, Vol 58, Part 1, pp. 2591 – 2598, Dec. 2011.
[36] P.E. Dodd, et al., "Production and Propagation of Single-Event Transients in High-Speed Digital Logic ICs," *IEEE Trans. Nucl. Sci*., vol. 51, No. 6, pp., Dec. 2004.
[37] B. Gill, et al., "Comparison of alpha-particle and neutron-induced combinational and sequential logic error rates at the 32nm technology node," *Proc. IEEE Int. Rel. Phys. Symp*., pp. 199 – 205, 2009.
[38] Boselli, V. Reddy, and C. Duvvury, "Latch-up in 65 nm CMOS technology: A scaling perspective," *Proc. IEEE Int. Rel. Phys. Symp.*, pp. 137–144, 2005.
[39] D. G. Mavis and D. R. Alexander, "Employing radiation hardness by design techniques with commercial integrated circuit processes," Proc. 16th AIAA/IEEE DASC, vol.1, pp. 26–30, Oct. 1997.
[40] N.A. Dodds et al., "Selection of Well Contact Densities for Latchup-Immune Minimal-Area ICs", *IEEE Trans. Nucl. Sci.,*vol. 57, Dec. 2010, pp. 3575-3581.
[41] J. Salzman, private communication, June 2012.
[42] E. Takeda, et al., "A new soft-error phenomenon in VLSIs: the alpha-particle-induced source/drain penetration (ALPEN) effect," *Proc. IEEE Int. Rel. Phys. Symp.*, pp. 109-112, 1988.
[43] N. C. Lu, "Advanced cell structures for dynamic RAMs," *IEEE Circuits and Devices Magazine*, Vol. 5(1), pp. 27 – 36, Jan. 1989.
[44] R. Edwards et al., "Technical standard for atmospheric radiation single event effects, (SEE) on avionics electronics," IEEE Rad. Effects Data Workshop, pp. 1-5, 2004.
[45] Borucki, L. et al., "Comparison of accelerated DRAM soft error rates measured at component and system level," *Proc. IEEE Int. Rel. Phys. Symp.*, pp. 482–487, 2008.
[46] S. Carcia, "Build the Micro D-Cam solid-state video camera Part 1: The IS32 Optic RAM and the micro D-Cam hardware," Carcia's Circuit Cellar, *BYTE Magazine*, pp. 19-30, Sept. 1983.
[47] P. E. Dodd and F. W. Sexton, "Critical Charge Concepts for CMOS SRAMs," *IEEE Trans. Nuc. Sci.*, Vol. 42, pp. 1764-1771, Dec. 1995.

[48] X. Zhu, et al., “Comparison of product failure rate to the component soft error rates in a multi-core digital signal processor,” *Proc. IEEE Int. Rel. Phys. Symp*, pp. 209 – 214, Apr. 2005.
[49] H. T. Nguyen et al., “Chip-level soft error estimation method,” *IEEE Trans. Dev. and Materials Reliability*, Vol. 5, pp. 365-381, Sept. 2005.
[50] S. Yamamoto, et al., “Neutron-induced soft error in logic devices using quasi-monoenergetic neutron beam” *Proc. IEEE Int. Rel. Phys. Symp*, pp. 305 – 309, Apr. 2004.
[51] J. F. Dicello, et al., “The Relative Efficiency of Soft-Error Induction in 4K Static RAMS by Muons and Pions,” *IEEE Trans. Nucl. Sci.*, Vol. 30, p. 4613-4615, Dec. 1983.
[52] J. F. Ziegler, “Terrestrial Cosmic Ray Intensities,” *IBM J. Res. Develop.*, Vol. 42(1), p. 125, Jan. 1998.
[53] Evaluated Nuclear Data File (ENDF), Database Version of Sept. 17, 2012 International Atomic Energy Agency: (http://www-nds.iaea.org/exfor/endf.htm).
[54] J. F. Ziegler and W. A. Lanford, “The effect of sea level cosmic rays on electronic devices,” *J. Appl. Phys.*, 52, pp. 4305-4311, 1981.
[55] N. Fukataa, et al., “Formation energy of vacancy in silicon determined by a new quenching method”, *Physica B: Condensed Matter*, Vols. 308–310, pp. 1125–1128, Dec. 2001.
[56] K. S. Krane, “Nuclear Reactions”, Introductory Nuclear Physics, New York: John Wiley & Sons, 1988, pp. 419.
[57] J.L. Autran, et al., “Soft-Error Rate Induced by Thermal and Low Energy Neutrons in 40 nm SRAMs,” *IEEE Trans. Nucl. Sci.*, Vol. 59, p. 2658-2665, Dec. 2012.
[58] M. S. Gordon, et al., "Measurement of the Flux and Energy Spectrum of Cosmic-Ray Induced Neutrons on the Ground," *IEEE Trans. Nucl. Sci*, vol. 51, pp. 3427-3434, Dec. 2004.
[59] E. Normand, “Single event effects in avionics,” *IEEE Trans. Nucl. Sci.*, vol. 43, p. 461-474, Apr. 1996.
[60] “Measurement and Reporting of Alpha Particle and Terrestrial Cosmic Ray-Induced Soft Errors in Semiconductor Devices”, *JEDEC Test Standard JESD-89A*, Sep. 2006.
[61] Flux calculation web site uses models generated by M.A. Shea and D.F. Smart using International Geomagnetic Reference Field for 1995. (http://www.seutest.com/cgi-bin/FluxCalculator.cgi)
[62] J. F. Ziegler, “The effect of concrete shielding on cosmic ray induced soft fails in electronic systems,” *IEEE Trans. Electron Devices*, Vol. 28, pp. 560 – 565, May 1981.
[63] M. V. Medvedev and A. L. Melott , “Do extragalactic cosmic rays induce cycles in fossil diversity?”, *The Astrophysical Journal*, Vol. 664(2), pp. 879-889, 2007.
[64] E. Stoupel et al., “Twenty years study of solar, geomagnetic, cosmic ray activity links with monthly deaths number”, (n-850304) *J. Biomedical Science and Eng*., 4, pp. 426-434, 2011.
[65] E. Stoupel et al., “Are neutrons involved in the pathogenesis of life-threatening cardiac arrhythmias?”, *J. Basic Clinical Physiology and Pharmacol*, 17(1), pp. 55-62, 2006.
[66] P. D. Bradley and E. Normand, “Single event upsets in implantable cardioverter defibrillators,” *IEEE Trans. Nucl. Sci*., vol. 45, pp. 2929–2940, Dec. 1998.
[67] T. C. May and M. H. Woods, “Alpha-particle-induced soft errors in dynamic memories,” *IEEE Trans. Electron Device*s, vol. 26, pp. 2–9, Feb. 1979.
[68] K. S. Krane, “Beta Decay”, Introductory Nuclear Physics, New York: John Wiley & Sons, 1988, pp. 273-277.
[69] K. S. Krane, “Radioactive Decay”, Introductory Nuclear Physics, New York: John Wiley & Sons, 1988, pp. 161.
[70] R. C. Baumann and J. W. McPherson, “Impact of Alpha-Particle Emitting Impurities in Phosphoric Acid on the Soft Error Rate of DRAMs.” Texas Instruments, Dallas, TX, #03-92-17, Nov. 1992.
[71] Z. Hasnain and A. Ditali, "Building-in Reliability: Soft errors – A Case Study," *IEEE Proc. IRPS Symp.*, pp. 276-280, 1992.
[72] *Low-Alpha Lead Symp.*, Lawrence Livermore National Laboratories, Feb. 1997.
[73] “Alpha Radiation Measurement in Electronic Materials”, *JEDEC Test Standard JESD-221*, May 2011.

[74] N. Seifert and M. Kirsch, “Real-Time Soft-Error Testing Results of 45-nm, High-K Metal Gate, Bulk CMOS SRAMs,” *IEEE Trans. Nucl. Sci.*, Vol. 59, pp. 2818-2823, Dec. 2012.
[75] Lt. J. D. Dirk et al., “Terrestrial Thermal Neutrons,” *IEEE Trans. Nucl. Sci*., Vol. 50, pp. 2060-2064, Dec. 2003.
[76] R.C. Baumann and E.B. Smith, “Neutron-induced 10B fission as a major source of soft errors in high density SRAMs, *Proc. IEEE Int. Reliability Physics Symp.*, p.152-157, 2000.
[77] H. Kobayashi, et al., “Soft Errors in SRAM Devices Induced by High Energy Neutrons, Thermal Neutrons and Alpha Particles,” *Proc. IEEE Int. Elec. Dev. Meeting*, pp. 337–340, 2002.
[78] E. Normand, et al., “Quantifying the Double-Sided Neutron SEU Threat, From Low Energy (Thermal) and High Energy (> 10 MeV) Neutrons,” *IEEE Trans. Nucl. Sci.*, Vol. 53, pp. 3587 – 3595, Dec. 2006.
[79] M. Olmos, et al., “Investigation of thermal neutron induced soft error rates in commercial SRAMs with 0.35 m to 90 nm technologies,” *Proc. IEEE Int. Reliability Physics Symp.*, 2006.
[80] R.C. Baumann and T.Z. Hossain, “Electronic Device And Process Achieving A Reduction In Alpha Particle Emissions For Boron-Based Compounds Essentially Free From Boron-11,” U.S. Patent 5 395 783, March 7, 1995.
[81] M. Hwang, W.R. McKee, and R.C. Baumann “Thermal neutron shielded integrated circuits,” U.S. Patent 6 239 479, May 29, 2001.
[82] Shi-Jie Wen et al., “B10 finding and correlation to thermal neutron soft error rate sensitivity for SRAMs in the sub-micron technology,” *IEEE Int. Integrated Reliability Workshop*, pp. 31-33, Oct. 2010.
[83] K. Rodbell, et al., “Low-energy proton-induced single-event-upsets in 65 nm node, silicon-on-insulator, latches and memory cells,” *IEEE Trans. Nucl. Sci.*, vol. 54, pp. 2474–2479, Dec. 2007.
[84] B. Sierawski, et al., “Impact of low-energy proton induced upsets on test methods and rate predictions,” *IEEE Trans. Nucl. Sci*., vol. 56, no. 6, pp. 3085 –3092, Dec. 2009.
[85] B. D. Sierawski, et al., “Effects of scaling on muon-induced soft errors,” in *Proc. IEEE Int. Rel. Phys. Symp.*, pp. 247–252, 2011.
[86] Public domain image of Thomas Kuhn obtained from (http://www.molwick.com/en/scientific-methods/041-scientific-methodology.html).
[87] A. Bird, "Thomas Kuhn", The Stanford Encyclopedia of Philosophy (Spring 2013 Edition), E. N. Zalta (ed.), (http://plato.stanford.edu/archives/spr2013/entries/thomas-kuhn/).
[88] T. S. Kuhn, The Structure of Scientific Revolutions, Chicago: University of Chicago Press, 1996, 3rd edition (with 1969 postscript), pp. 82-89.
[89] T. S. Kuhn, The Structure of Scientific Revolutions, Chicago: University of Chicago Press, 1996, 3rd edition (with 1969 postscript), pp. 36-39.
[90] J. Keithley, The story of electrical and magnetic measurements: from 500 B.C. to the 1940s: John Wiley and Sons, p. 205, 1999.
[91] “Nobel Lectures, Physics 1901-1921”, Elsevier Publishing Company, Amsterdam, 1967.
[92] "Henri Becquerel - Biography". Nobelprize.org. 30 Apr 2013 (http://www.nobelprize.org/nobel_prizes/physics/laureates/1903/becquerel-bio.html).
[93] “Marie Curie and the Science of Radioactivity”, American Institute of Physics, (http://www.aip.org/history/curie/resbr2.htm)
[94] "Marie and Pierre Curie and the Discovery of Polonium and Radium". Nobelprize.org. 30 Apr 2013 (http://www.nobelprize.org/nobel_prizes/physics/articles/curie/).
[95] "Ernest Rutherford - Biography". Nobelprize.org. 30 Apr 2013 (http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1908/rutherford.html)
[96] “Alpha Particles and the Atom: Rutherford at Manchester, 1907–1919”, American Institute of Physics, (http://www.aip.org/history/exhibits/rutherford/sections/alpha-particles-atom.html)
[97] Domenico Pacini and the origin of cosmic rays, CERN Courier, Jul 18, 2012. (http://cerncourier.com/cws/article/cern/50216).
[98] Paul Frame, “String Electrometers”, Oak Ridge Associated Universities,

(http://www.orau.org/ptp/collection/electrometers/stringelectrometers.htm).
[99] P. Carlson and A. De Angelis, "Domenico Pacini and the discovery of cosmic rays," Eur. Phys. J. H 309, 2011.
[100] V. F. Hess, "Über die absorption der g-strahlung in der atmosphre," Phys. Zeits. 12, 998, 1911.
[101] V. F. Hess, "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten ," Phys. Zeits. 13, 1084, 1912.
[102] D. Hesselgrave, "A Short Technical Early History of Vacuum Tubes (Electron Valves)," Dhh:\Vacuum Tubes and Transistors\Vacuum Tube History.doc, Mar. 2004, (http://www.moah.org/education/transHist1.html)
[103] I. Brodie and P. R. Schwoebel, "Vacuum microelectronic devices," IEEE Proc., Vol. 82(7), pp.1006 – 1034, July 1994.
[104] W. Gosling, "The pre-history of the transistor," Radio and Electronic Engineer, Vol. 43(1.2), pp. x, 1973.
[105] M. Riordan, "The lost history of the transistor," IEEE Spectrum, Vol. 41(5), pp. 44-49, 2004.
[106] M. Riordan, "The lost history of the transistor," IEEE Spectrum, Vol. 41(5), p. 45, 2004.
[107] W. F. Brinkman, "The transistor: 50 glorious years and where we are going," in digest of *IEEE International Solid-State Circuits Conference*, pp. 22 – 26, 1997.
[108] Jack Ward, Texas Instruments 900 Series, Transistor Museum, (http://www.transistormuseum.com)
[109] M. Riordan, "The lost history of the transistor," *IEEE Spectrum*, Vol. 41(5), p. 45, 2004.
[110] "1964 - First Commercial MOS IC Introduced", Computer History Museum, (http://www.computerhistory.org/semiconductor/timeline/1964-Commecial.html).
[111] Transistor Count: Microprocessors, Wikipedia, (http://en.wikipedia.org/wiki/Transistor_count).
[112] R. L. Pease, D. G. Platteter, K. G. Kerris and K. F. Galloway, "Fifty Years of Radiation Hardening", NSREC Fact Book, IEEE Nuclear and Plasma Sciences Society, 2013 edition, p. 5.
[113] J. J. Suran, "Circuit considerations relating to microelectronics," in *Proc. IRE*, vol. 49, pp. 420-426, Feb. 1961.
[114] J. T. Wallmark and S. M. Marcus, "Minimum Size and Maximum Packing Density of Nonredundant Semiconductor Devices," in *Proc. IRE*, pp. 286 - 298, March 1962.
[115] J. T. Wallmark and S. M. Marcus, "Minimum Size and Maximum Packing Density of Nonredundant Semiconductor Devices," in *Proc. IRE*, p. 293, March 1962.
[116] C. Lage, et al, "Soft Error Rate and Stored Charge Requirements in Advanced High-Density SRAMs," in *Digest IEEE International Electron Devices Meeting*, pp. 821-824, 1993.
[117] T.J. O'Gorman, "The effect of cosmic rays on the soft error rate of a DRAM at ground level," *IEEE Trans. Elec. Dev.*, vol. 41, pp. 553-557, April 1994.
[118] G. E. Moore, "Cramming More Components onto Integrated Circuits", *Electronics*, 38(8), pp. 4-7, Apr. 19, 1965.
[119] G. Ludwig, "James Van Allen, From High School to the Beginning of the Space Era: A Biographical Sketch," for Van Allen's 90th Birthday Celebration at the University of Iowa, Oct. 2004. (http://en.wikipedia.org/wiki/James_Van_Allen)
[120] J. A. Van Allen, "Direct Detection of Auroral Radiation with Rocket Equipment", in *Proc. National Academy of Sciences*, Vol. 43 (1), pp. 57-62. January 1957.
[121] G.W.A. Dummer, "The Reliability of Components in Satellites," *J. of Brit. IRE.*, Vol 22(6), pp. 457 – 463, Dec. 1961.
[122] D. Binder, E. C. Smith, and A. B. Holman, "Satellite anomalies from galactic cosmic rays," *IEEE Trans. Nucl. Sci.*, Vol. 22(6), pp. 2675-2680, Dec. 1975.
[123] J. C. Pickel and J. T. Blandford, Jr., "Cosmic ray induced errors in MOS memory cells," in *IEEE Trans. Nucl. Sci.*, Vol. 25(6), Dec. 1978.
[124] T. C. May, M. H. Woods, "A New Physical Mechanism for Soft Errors in Dynamic Memories" in *Proc. IEEE Reliability Physics Symposium*, pp. 33-40, April 1978.
[125] T. C. May and M. H. Woods, "Alpha-particle-induced soft errors in dynamic memories," *IEEE Trans. Electron Device*s, vol. 26, pp. 2–9, Feb. 1979.

[126] D. S. Yaney, J. T. Nelson and L. L. Vanskike, *IEEE Trans. Electron Devices*, ED-26, pp. 10-16, Jan. 1979.
[127] 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*, EDL-2, pp. 103-105, Apr. 1981.
[128] C. M. Hsieh, P. C. Murley, and R. R. O'Brien, "Dynamics of charge collection from alpha-particle tracks in integrated circuits," in *Proc. IEEE Int. Reliability Phys. Symp*., pp. 38–42, 1981.
[129] C. Hu, "Alpha-Particle-Induced Field and Enhanced Collection of Carriers", *IEEE Electron Device Letters*, EDL-3, No. 2, pp. 31-34, Feb. 1982.
[130] E. S. Meieran, P. R. Engel, T. C. May, "Measurement of Alpha Particle Radioactivtiy in IC Device Packages," *IEEE Reliability Physics Symposium*, pp. 13 – 22, 1979.
[131] J. A. Woolley, L. E. Lamar, N. H. Stradley, and D. Max Harshbarger, "Low Alpha-Particle-Emitting Ceramics: What's the Lower Limit?" *IEEE Trans. Components, Hybrids, and Manuf. Tech.*,Vol. CHMT-2(4), pp. 388-390, Dec. 1979.
[132] S. W. Levine, "Alpha Emission Measurements of Lids and Solder Preforms on Semiconductor Packages" *IEEE Trans. Components, Hybrids, and Manuf. Tech.*,Vol. CHMT-2(4), pp. 391-395, Dec. 1979.
[133] "Alpha Radiation Measurement in Electronic Materials," *JEDEC Standard JESD221*, May 2011.
[134] R. C. Baumann and T. Z. Hossian, "Determination of Uranium and Thorium Concentration in 16Mbit DRAM Metallizations by Neutron Activation Analysis." Texas Instruments, Dallas, TX, #03-91-05, Mar. 1991.
[135] R.C. Baumann and J. McPherson, "The Impact of Alpha-Particle Emitting Impurities in Phosphoric Acid on the Soft Error Rate of DRAMs." Texas Instruments, Dallas, TX, #03-92-17, Mar. 1992.
[136] Z. Hasnain and A. Ditali, "Building-in Reliability: Soft Errors - A Case Study," in *Proc. IEEE Internat. Reliab. Physics Symp.*, pp. 276-280, Apr. 1992.
[137] M. L. White, J. W. Serpiello, K. M. Striny, and W. Rosenzweig, "The Use of Silicone RTV Rubber for Alpha Particle Protection on Silicon Integrated Circuits," in *Proc. IEEE Int. Reliability Phys. Symp*., pp. 34–37, 1981.
[138] D. Makino, "Application of polyimide resin to semiconductor devices in Japan," *IEEE Electrical Insulation Mag*., Vol. 4 (2), pp. 15 – 23, 1988.
[139] R. C. Baumann, "Investigation of the Effectiveness of Polyimide Films for the Stopping of Alpha Particles in Megabit Memory Devices." Texas Instruments, Dallas, TX, #03-91-11, Mar. 1991.
[140] J. F. Ziegler and W. A. Lanford, "The Effect of Cosmic Rays on Computer Memories", *Science*, Vol. 206, No. 4420, pp. 776, 1979.
[141] J. T. Wallmark and S. M. Marcus, "Minimum Size and Maximum Packing Density of Nonredundant Semiconductor Devices," in *Proc. IRE*, p. 287, March 1962.
[142] J. F. Ziegler and W. A. Lanford, "The effect of sea level cosmic rays on electronic devices," *J. Appl. Phys.*, 52, pp. 4305-4311, 1981.
[143] S. Kirkpatrick, "Modeling Diffusion and Collection of Charge from Ionizing Radiation in Silicon Devices," *IEEE Trans. on Elect. Devs*., Vol. ED-26, No. 11, pp. 1742-1753, 1979.
[144] J. Y. Chan, J. J. Barnes, C. Y. Wang, J. M. De Blasi, M. R. Guidry, "A 100 ns 5 V Only 64K x 1 MOS Dynamic RAM," *IEEE J. Solid-State Circuits*, Vol. SC.15 (5), pp. 839-846, Oct. 1981.
[145] A. F. Tasch, Jr. , P. K. Chatterjee, H-S. Fu, and T. C. Holloway, "The Hi-C RAM Cell Concept," *IEEE Trans. Electr. Dev*., vol. ED-25, pp.33-42, Jan. 1978.
[146] N. Ieda, E. Ami, Y. Ohmori, and K. Takeya, "A 64K MOS RAM design," in *Digest IEEE Solid State Devices Conf*., pp. 17-18, Aug. 1977.
[147] T. Wada, O. Kudoh, M. Sakamoto, H. Yamada, K. Nakamura, and M. Kamoshita, "A 64K X 1 dynamic ED-MOS RAM," *IEEE J. Solid-State Circuits*, Vol. SC-13, pp. 600-606, Oct. 1978.
[148] T. Wada, M. Takada, S. Matsue, M. Kamoshida, and S. Suzuki, "A 150 ns, 150 mW, 64K dynamic MOS RAM," *IEEE J. Solid-State Circuit*s, Vol. SC-13, pp. 607-611, Oct. 1978.

[149] K. Naturi., M. Ogura, H. Iwai, K. Meaguchi, and S. Taguchi, "A 64K bit MOS dynamic random access memory," *IEEE Trans. Electron Devices*, Vol. ED-26, pp. 560-563, Apr. 1979.
[150] G. A. Sai-Halasz, "Cosmic Ray Induced Soft Error Rate in VLSI Circuits", *IEEE Electron Device Letters*, Vol. EDL-4 (6), pp. 172-174, June 1983
[151] C. Lage, et al, "Soft Error Rate and Stored Charge Requirements in Advanced High-Density SRAMs," in *Digest IEEE International Electron Devices Meeting*, pp. 821-824, 1993.
[152] T.J. O'Gorman, "The effect of cosmic rays on the soft error rate of a DRAM at ground level," *IEEE Trans. Elec. Dev.*, vol. 41, pp. 553-557, April 1994.
[153] W. R. McKee, et al, "Cosmic ray neutron induced upsets as a major contributor to the soft error rate of current and future generation DRAMs," in *Proc. IEEE International Reliability Physics Symposium*, pp. 1 – 6, 1996.
[154] Y. Tosaka et al, "Impact of cosmic ray neutron induced soft errors on advanced submicron CMOS circuits," in *Digest IEEE VLSI Technology Symposium*, pp. 148-149, 1996.
[155] "Measurement and Reporting of Alpha Particles and Terrestrial Cosmic Ray-Induced Soft Errors in Semiconductor Devices," *JEDEC Standard JESD89A*, Oct. 2006.
[156] Daniel Lyons, "Sun Screen", *Forbes.com*, Nov. 13, 2000, (http://www.forbes.com/forbes/2000/1113/6613068a.html)
[157] I. Arimura and C. Rosenberg, "Anomalous Damage Mechanism in PNP Silicon Transistors due to Thermal Neutrons," *IEEE Trans. Nucl. Sci.*, Vol. 20(6), pp. 274-279, Dec. 1973.
[158] T. C. May, "Soft Errors in VLSI: Present and Future," *IEEE Trans. Components, Hybrids, and Manufacturing Technology*, Vol. 2(4), pp. 377-387, Dec. 1979.
[159] R. L. Fleischer, "Cosmic Ray Interactions with Boron: A Possible Source of Soft Errors", *IEEE Trans. Nucl. Sci.*, Vol. 30(5), pp. 4013-4015, Oct. 1983.
[160] R. Baumann and T. Hossain, "Boron as a Dominant Source of Ionizing Radiation in Semiconductor Devices", Texas Instruments, Dallas, TX, #03-94-04, Mar. 1994.
[161] R. Baumann, T. Hossain, S. Murata, H. Kitagawa, "Boron compounds as a dominant source of alpha particles in semiconductor devices", in *Proceedings of IEEE International Reliability Physics Symposium*, 1995, pp. 297-302.
[162] Lt. J. D. Dirk et al., "Terrestrial Thermal Neutrons," *IEEE Trans. Nucl. Sci.*, Vol. 50, pp. 2060-2064, Dec. 2003.
[163] R. Baumann, E. B. Smith, "Neutron-induced boron fission as a major source of soft errors in deep submicron SRAM devices", in *Proceedings of IEEE International Reliability Physics Symposium*, 2000, pp. 152-157.
[164] H. Kobayashi et al., "Soft errors in SRAM devices induced by high energy neutrons, thermal neutrons and alpha particles", in *Proceedings of International Electron Devices Meeting*, 2002, pp. 337-340.
[165] R.C. Baumann and T.Z. Hossain, "Electronic Device And Process Achieving A Reduction In Alpha Particle Emissions For Boron-Based Compounds Essentially Free From Boron-11," U.S. Patent 5 395 783, March 7, 1995.
[166] M. Hwang, W.R. McKee, and R.C. Baumann "Thermal neutron shielded integrated circuits," U.S. Patent 6 239 479, May 29, 2001.
[167] S. J. Wen, S. Y. Pai, R. Wong, M. Romain, N. Tam, "B10 finding and correlation to thermal neutron soft error rate sensitivity for SRAMs in the sub-micron technology," *IEEE International Integrated Reliability Workshop Final Report*, pp. 31 – 33, Oct. 2010.