Scaling and Technology Issues for Soft Error Rates

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

Abstract - The effects of device technology and scaling on soft error rates are discussed, using information obtained from both the device and space communities as a guide to determine the net effect on soft errors. Recent data on upset from high-energy protons indicates that the soft-error problem in DRAMs and microprocessors is less severe for highly scaled devices, in contrast to expectations. Possible improvements in soft-error rate for future devices, manufactured with silicon-on-insulator technology, are also discussed.

I. INTRODUCTION

Soft-errors from alpha particles were first reported by May and Woods [1], and considerable effort was spent by the semiconductor device community during the ensuing years to deal with the problem of errors from alpha particles in packaging, metallization and other materials. This included modifications in device design to reduce the inherent sensitivity to extraneous charge, as well as application of topical shielding and improvements in material purity. Atmospheric neutrons can also produce soft errors. One example is shown in Figure 1, after Lage, et al.[2], which shows the increase in measured soft error rate when experiments were done on SRAMs using alpha sources with different intensities. The increased error rate is due to the presence of atmospheric neutrons that have a larger relative influence when alpha experiments are done for long time periods using low-intensity sources.

Figure 1. Increase in soft-error rate of an SRAM when low-intensity alpha particle sources are used.

The research in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA) under the NASA Electronics Parts and Packaging Program (ERC).
space community also routinely tests devices with heavy ions. For example, Figure 3 shows how the cross sections of internal registers and cache memory of a microprocessor increase when tests are done with long-range ions that have different values of LET. (For reference, the LET of a 5-MeV alpha particle is 1 MeV-cm²/mg). The heavy-ion data provides a direct indication of the specific ionization track density required for upset, and is a useful way to compare the upset sensitivity of different devices.

![Figure 3](image)

**Figure 3.** Dependence of cross section on linear energy transfer for a microprocessor. The upset threshold is slightly above the LET for a 5-MeV alpha particle.

**II. SPACE AND TERRESTRIAL ENVIRONMENTS**

**A. Comparison of Environments**

The high-energy cosmic rays that spacecraft encounter are largely shielded by the earth’s atmosphere. Consequently they do not occur in significant numbers in terrestrial applications. However, the interaction of cosmic rays in the upper atmosphere produces secondary particles, including neutrons. The neutrons have a low interaction probability with the atmosphere, and significant numbers of energetic neutrons arrive at the earth’s surface.

Although spacecraft are usually not concerned with neutrons, most space environments include high-energy protons with a wide distribution of energies. Direct ionization from protons is generally too low to be of concern in microelectronics. However, protons can produce nuclear reactions in silicon, and the secondary reaction products from reactions or nuclear collisions -- which have relatively short range -- produce ionization tracks with much higher charge densities compared to those generated along the path of the high-energy protons that initiate the reaction. However, the cross section for such indirect processes is much smaller than for direct ionization because the incoming particle has to undergo a nuclear collision in order to produce the reaction (nuclear cross sections are typically $10^{-4}$ to $10^{-5}$ lower).

Protons with energies above 50 MeV have cross sections for nuclear reactions that are nearly identical to those for high-energy neutrons. This allows test results for high-energy protons to be extended directly to neutron environments, which has been noted by workers in both the semiconductor and space environments [10,11]. The proposed standard for evaluating SER rates in terrestrial applications allows protons to be used as an alternative to tests with neutron sources [12].

Although cross sections for protons and neutrons are nearly the same at high energies, that is not the case for lower energies (<50 MeV). At low energies, protons cross sections increase because of Coulomb interactions with the lattice atoms, causing the cross section to increase with decreasing energy.

In addition, the neutron energy distribution of the uncharged neutrons in terrestrial environments extends to much lower energies than the distribution of protons within spacecraft because even moderate amounts of shielding eliminate most of the low-energy protons in space. Figure 4 compares the distribution of atmospheric neutrons (actually the energy distribution of the white neutron source at Los Alamos National Laboratory) with the typical distribution of proton energies in spacecraft.

![Figure 4](image)

**Figure 4.** Energy distributions of atmospheric neutrons on the ground and protons in an earth-orbiting spacecraft.

The point of this figure is to show that the proton spectrum in space decreases at low energies, reducing the importance (and interest) of low energy protons in causing errors in spacecraft. On the other hand, the relative number of neutrons in terrestrial environments continues to increase at low energies. Thus, the soft-error rate problem on the ground is much more affected by reductions in the threshold energy for errors from neutron-induced reactions compared to the soft-error problem in space (from protons).

**B. Recoil Energy Distribution**

A great deal of experimental work has been done to investigate the distribution of recoil energies when experiments are done with monoenergetic protons [13,14]. That work has shown that there is a continuous distribution of recoil energies that decreases up to a maximum energy. For silicon, the maximum recoil energy due to elastic collisions is 13.6% of the energy of the incident particle. This means that for 100 MeV protons (or neutrons), the...
maximum total energy of a recoil is 13.6 MeV. However, there are very few particles with maximum recoil energy. The energy distribution decreases roughly as 1/E up to the maximum recoil energy.

The range of the recoil products varies with energy. For the case where the incident particle energy is 100 MeV the more energetic recoils have ranges up to 5 µm, but most of the particles have ranges below 2 µm.

For the real environments in terrestrial or space environments the net distribution of recoils is a superposition of the distribution of energetic primary particles along with the energy distribution of the recoil products. Most of the recoils will have relatively short ranges, but “short” is a relative term. For older technologies, one could make the assumption that the entire recoil energy was absorbed within the sensitive volume of a sensitive device because device dimensions were considerably larger than the range of most recoils. That is no longer true for highly scaled devices, which affects the collected charge.

C. Effective Volume for Recoil Products

Many of the recoil products have ranges that are considerably longer than the average charge collection depth. Thus, particles that undergo collisions outside the charge collection region may have sufficient energy to reach the sensitive region and deposit part of the energy in that region. One way to deal with this is to define a generation rate for recoils (referred to as the burst generation rate) and then assume an extended region outside the charge collection region for the total sensitive volume [15-18].

This approach has been reasonably effective for older generation devices, but needs to be modified to handle modern devices with smaller sensitive volumes [ref]. It is also important to realize that the charge collection region is not really constant, but changes dimensions depending on the total deposited energy and location because the depletion region collapses if the energy is sufficiently high.

Charge collection volumes for DRAMs are quite complex because diffused charge in the substrate, well beyond the depletion region, can be collected by a reverse-biased junction. Figure 5 shows the results of calculations of the charge collected by particles that produce charge tracks well away from the depletion region [19] (similar to the situation in the bulk substrate of a DRAM). These calculations were done with the PISCES device analysis program. Note that nearly all of the deposited charge is collected, but the time over which the charge collection occurs extends to relatively long time intervals. For DRAMs this is an important effect because the charge will be effective in fully or partially discharging the storage capacitor. Although the same charge collection process will take place in SRAMs or logic devices, they are generally only sensitive to charge collected in short time periods, unlike DRAMs. Thus, charge collection depends not only on substrate properties, but also on the way in which the circuit responds to longer duration charge pulses.

Figure 5. Charge collection from short range recoils at various distances from the depletion region of a p-n junction.

III. DEVICE SCALING TRENDS

A. Scaling for Microprocessors and Logic

Device scaling is an extremely complex topic that involves many assumptions about technology evolution. The earliest studies were done by Dennard, et al. at IBM [20]. More recent studies by Davari, et al. [21] and Hu [22] have discussed scaling trends for two basic technology directions: high performance, where power dissipation is a secondary concern; and low power, where power and performance are both considered. Table 1, after Davari, presents the results based on the state-of-the-art in 1995. Those predictions have been reasonably accurate in predicting future technology trends. For radiation susceptibility, the key parameters are gate oxide thickness, relative speed and density, and power supply voltage which affect charge collection and critical charge.

Table 1. Scaling Predictions for High-Performance and Low-Power Logic Circuits (after Davari, et al. [21])

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply voltage (V)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High performance</td>
<td>5</td>
<td>5.03</td>
<td>3.2/2.5</td>
<td>2.5/1.5</td>
<td>1.5/1.2</td>
<td>1.0/</td>
</tr>
<tr>
<td>Low power</td>
<td>-</td>
<td>3.3/2.5</td>
<td>2.5/1.5</td>
<td>1.5/1.2</td>
<td>1.0/</td>
<td>1.0/</td>
</tr>
<tr>
<td>Lithog. resolution (µm)</td>
<td>1.25</td>
<td>0.8</td>
<td>0.5</td>
<td>0.35</td>
<td>0.25</td>
<td>0.18</td>
</tr>
<tr>
<td>Channel length (µm)</td>
<td>0.9</td>
<td>0.60/0.45</td>
<td>0.35/0.25</td>
<td>0.20/0.15</td>
<td>0.1/</td>
<td>0.07</td>
</tr>
<tr>
<td>Gate oxide thickness (nm)</td>
<td>23</td>
<td>15/12</td>
<td>9/7</td>
<td>6/5</td>
<td>3.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Relative density</td>
<td>1.0</td>
<td>2.5</td>
<td>6.3</td>
<td>12.8</td>
<td>25</td>
<td>48</td>
</tr>
<tr>
<td>Relative speed</td>
<td>1.0</td>
<td>1.4/2.0</td>
<td>2.7/3.4</td>
<td>4.2/5.1</td>
<td>7.2/</td>
<td>9.6/</td>
</tr>
<tr>
<td>High performance</td>
<td>-</td>
<td>1.0/1.6</td>
<td>2.0/2.4</td>
<td>3.2/3.5</td>
<td>4.5/</td>
<td>7.2/</td>
</tr>
<tr>
<td>Low power</td>
<td>-</td>
<td>1.0/1.6</td>
<td>2.0/2.4</td>
<td>3.2/3.5</td>
<td>4.5/</td>
<td>7.2/</td>
</tr>
</tbody>
</table>

B. Scaling for Memories

Memory technologies involve quite different scaling assumptions. There are two reasons for this: first, memory technologies require very low leakage and hence must impose a much higher ratio between “on” and “off” transistors within the memory; and second, memories are generally sold at very low prices compared to microprocessors and other high-performance devices, resulting in much more conservative approaches for device design.
There have been many changes in memory technology. For DRAMs, storage capacitor technology is one factor that has continually evolved. Although it is not possible to cover all aspects of DRAM scaling in this paper, Figure 6, after Itoh, et al. [23] shows how cell area and storage capacitor size have changed over several DRAM generations. Device dimensions have scaled inversely with memory size, as expected from general scaling trends. On the other hand, storage capacitance has decreased far more slowly. One reason for this is the need to keep the total stored charge above the charge generated by alpha particles, as well as other noise sources. The charge generated by a 5-MeV alpha particle in one micrometer of path length is indicated on this figure for comparison. This trend in storage capacitance affects soft-error rates for proton or neutron reactions because it establishes a “floor” for charge generation.

![Figure 6. Cell area and stored charge for DRAMs over several generations (after Itoh, et al. [23]).](image)

**C. Scaling Implications for SER Sensitivity**

**Charge Collection Depth**

The depth over which charge from a long-range ion is collected generally decreases with scaling. This is due to several factors, including the decrease in active layer thickness and increase in channel doping, which reduces the depletion width as well as charge funneling.

Substrate properties also affect charge collection depth, particularly for epitaxial CMOS. Present state-of-the-art CMOS devices use epitaxial layers that are approximately 2 µm thick, effectively limiting charge collection to that region. That is not true for bulk substrates (which are typically used for memories), and thus the charge collection depth for memories can be up to an order of magnitude higher. As discussed later, circuits designed with SOI technology will further reduce charge collection depth, potentially reducing SER sensitivity by up to an order of magnitude.

**Switching Amplitude**

Power supply voltage (and total switching amplitude) directly affect critical charge and upset sensitivity. For logic circuits and SRAMs, the decrease in switching voltage reduces the voltage swing, directly affecting critical charge. That factor, along with the decrease in device dimensions and increase in switching speed, will compete with the decrease in charge collection depth.

For DRAMs the effect of internal voltage is less apparent because of the effect of charge leakage, which reduces the internal cell voltage between successive refresh cycles. DRAM architecture is also a factor. Most DRAMs use boosted word lines, and the internal noise margin is affected by the ratio of the bit line capacitance and the cell storage capacitance. Changes in DRAM architecture have resulted in block-oriented designs that have allowed the capacitance ratio to increase for scaled devices [24]. Architecture and device design play a large role in determining DRAM sensitivity, and make it difficult to establish general scaling trends.

**Critical Charge**

The effects of device scaling on soft-error rate depend on several competing factors. The critical charge required to upset a memory element (or active transistor) is expected to decrease with scaling, but this is partially offset by the decrease in charge collection depth as well as by device architecture. Although earlier studies predicted a steady decrease in critical charge with scaling [25], the complex interaction of device design with other scaling issues was not taken into account. Even though critical charge is expected to decrease with scaling, the net effect on threshold conditions is less obvious, and may actually improve with scaling as discussed in the next section. Designs that take upset from alpha particles into account will reduce the critical charge compared to earlier projections.

**Cross Sectional Area**

Finally, even if the first two factors cancel the decrease in device area that results from scaling will lower the total cross section per bit. Thus, the net upset rate per bit is expected to decrease with scaling even if the minimum ionization track charge density is unchanged.

Other factors must also be considered. For example, devices with complex internal architectures (e.g., special internal test modes in DRAMs) may allow new types of errors to be introduced that are difficult to identify and may interfere with standard ways to detect and correct errors. The total functional capability of VLSI devices continues to increase with scaling, and thus the probability of an error within a chip may increase substantially even if the error rate per bit decreases. This has certainly been an issue in the space environment.

**IV. TEST RESULTS FROM THE SPACE COMMUNITY**

**A. Microprocessors**

Extensive work has been done on microprocessor testing during the last 15 years. That work has shown that the dominant effect of heavy ions and protons on microprocessors is upset in internal registers. Upset in random logic has not been a significant factor up to now, although this may change as microprocessors continue to evolve.
One result which was certainly unexpected is that the threshold LET required to upset registers in NMOS and CMOS microprocessors has essentially not changed during the last fifteen years [26]. As shown in Table 2, the feature size used in manufacturing these devices has changed by more than one order of magnitude during that time period. The last entry in this table is for a very high speed modern microprocessor, the Power PC750. This may be influenced by overall concerns about alpha particle sensitivity by microprocessor manufacturers.

Table 2. Threshold LET for Microprocessors

<table>
<thead>
<tr>
<th>Device</th>
<th>Manuf.</th>
<th>Year</th>
<th>Feature Size (approx.)</th>
<th>Threshold LET (MeV-cm²/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z-80</td>
<td>Zilog</td>
<td>1986</td>
<td>3 µm</td>
<td>1.5 - 2.5</td>
</tr>
<tr>
<td>8086</td>
<td>Intel</td>
<td>1986</td>
<td>1.5 µm</td>
<td>1.5 - 2.5</td>
</tr>
<tr>
<td>80386</td>
<td>Intel</td>
<td>1991</td>
<td>0.8 µm</td>
<td>2 - 3</td>
</tr>
<tr>
<td>68020</td>
<td>Mot.</td>
<td>1992</td>
<td>0.8 µm</td>
<td>1.5 - 2.5</td>
</tr>
<tr>
<td>LS64811</td>
<td>LSI</td>
<td>1993</td>
<td>1.2 µm</td>
<td>2 - 2.5</td>
</tr>
<tr>
<td>90C601</td>
<td>MHS</td>
<td>1993</td>
<td>1.2 µm</td>
<td>2 - 2.5</td>
</tr>
<tr>
<td>80386</td>
<td>Intel</td>
<td>1996</td>
<td>0.6 µm</td>
<td>2 - 3</td>
</tr>
<tr>
<td>PC603e</td>
<td>Mot.</td>
<td>1997</td>
<td>0.4 µm</td>
<td>1.7 - 3</td>
</tr>
<tr>
<td>Pentium</td>
<td>Intel</td>
<td>1997</td>
<td>0.35 µm</td>
<td>2 - 3</td>
</tr>
<tr>
<td>Power PC750</td>
<td>Mot.</td>
<td>2000</td>
<td>0.25 µm</td>
<td>2 - 2.5</td>
</tr>
</tbody>
</table>

It is also useful to examine trends in proton upset for microprocessors. Figure 7 shows the dependence of proton upset cross section on proton energy for register errors, normalized per bit, for two relatively advanced processors. The threshold energy is essentially the same, but the cross section of the newer Power PC750 processor is more than an order of magnitude lower than that of the older PC603e. Both processors use thin epitaxial layers over highly doped substrates. The PC603e was designed with a feature size of 0.35 µm, while the feature size of the Power PC750 was 0.25 µm.

The fact that the threshold energy is approximately the same is consistent with the results in Table 2 showing nearly the same threshold LET for the two processors. However, there is an important consideration for proton recoils beyond that of threshold LET for long-range particles. As devices become smaller, the range of the recoil atoms from indirect processes increases relative to device feature size. This allows reaction products with lower energy to upset the device, potentially increasing the upset rate in the terrestrial neutron environment.

These results are indicative of mainstream microprocessors that are fabricated on epitaxial substrates. The newer Power PC750 operates at very high clock frequencies (700 MHz). Test results on both types of processors indicate that their responses to both protons and heavy ions are dominated by errors in registers and the cache memory. It is possible that faster devices may be susceptible to transient errors in logic or other regions of the processor, creating a scenario where the upset rate may increase with further scaling. That topic is currently under investigation at JPL using next-generation microprocessors as a vehicle for study. Once that threshold is reached, it may become quite difficult to use devices in space. The error rate in the terrestrial environment will also increase, but will be of less importance because the particle flux is so much lower for the terrestrial environments.

B. DRAMs

Trends in DRAM responses are also of great interest. However, two points must be kept in mind: first, DRAM scaling involves very different assumptions about device parameters; and second, DRAMs are almost always fabricated on bulk substrates and are sensitive to charge collection over far longer distances. The latter factor is highly significant when comparing upset from heavy ions (with long range) and the short-range recoil products from proton or neutron reactions.

Figure 8 shows how the sensitivity of proton upset rates have changed as DRAMs have evolved. The upset rate has steadily declined, on average over several DRAM generations. This figure does not consider any of the details involving the structure of individual devices, and it is apparent that there are large differences in the upset rate of devices from different manufacturers. This is probably related to the design of the wells; most DRAMs use a triple-well structure in order to reduce leakage current in the memory array.
Heavy-ion results have shown that the threshold LET for DRAMs has remained between 1 and 2 MeV-cm²/mg for an extended period. In this case, the threshold is almost certainly related to concerns about alpha-particle sensitivity, which involves relatively long-range particles with LET values that are slightly below 1 MeV-cm²/mg. Although earlier generation devices often used internal die coatings to reduce sensitivity to upset from those sources, newer generation DRAMs have taken alpha particle sensitivity into account as part of the design process.

V. PREDICTIONS FOR MORE ADVANCED TECHNOLOGIES

A. High-Speed/Low-Voltage Designs

As discussed earlier, one of the major concerns about high-speed devices is sensitivity to upset from internal transients in logic circuits. Clocked logic circuitry generally reduces sensitivity to such transients because the timing window is limited to periods when the clock is undergoing a logic transition. However, as devices are pushed to higher and higher speeds the relative size of the sensitive timing window increases. Reduced switching voltages increase overall sensitivity to this type of error, particularly when power dissipation is reduced.

Recent simulations and models by Shin, et al. have shown how charge collected from alpha particles in very small devices is affected by junction area and switching voltage [27]. Figure 9 shows the results of those calculations, which are done for a triple-well structure in an advanced DRAM. The collected charge does not change nearly as rapidly with junction area as one would expect. This is mainly due to the collapse of internal electric fields during the first 200-500 ps after the ion strike. Dodd has investigated charge collection for basic p-n structures with various conditions, including epitaxial substrates [28]. However, more work needs to be done on charge collection to investigate these dependencies for the more compact structures that occur in circuits. This requires three-dimensional simulations which are time consuming and difficult to interpret because of the many variables involved.

B. Silicon-on-Insulator Technology

Silicon-on-insulator technology has been studied for many years. For the first time mainstream high-performance devices are being produced [29], although their availability is extremely limited at the present time. No radiation test results are available at this time, but is expected in the near future.

Results for SOI processors from much older technologies indicated that the threshold LET was approximately five time higher for equivalent SOI structures [30]. This agrees closely with elementary calculations using the charge collection depth in combination of the assumption of a somewhat lower critical charge. If the newer processors behave similarly, the net effect will be to reduce upset rates in either space or terrestrial environments by about one order of magnitude. Although this is a significant decrease, it does not eliminate concern about upset rates in either the terrestrial or space environments.

Another important effect is bipolar action due to the short channel length [31], which remains an important problem for partially depleted devices that are currently used for SOI.
devices [32]. The bipolar parasitic transistor between source and drain can be turned on by heavy ions or recoil products from protons or neutrons, providing far more current than that due to the ionization. Recent work on SOI structures has investigated bipolar action triggered by terrestrial neutrons [33], relying on modified contacts to reduce vulnerability.

Commercial manufacturers are continuing to evaluate the response of SOI devices to neutrons. Figure 10 shows recent results from a group at Fujitsu Semiconductor [34]. These measurements were done on large-area SOI structures which allow considerably more total charge to be collected compared to the compact, small area regions in an actual SOI transistor. The work shows how the distribution of charges collected in experiments at the Los Alamos white neutron source depends on SOI film thickness.

![Figure 10. Charge collection in large-area SOI test structures at the Los Alamos neutron source for various film thicknesses.](image)

VI. CONCLUSIONS

This paper has discussed various factors that affect soft error rates in advanced devices using a slightly different approach that takes advantage of the body of test data and analysis in the space community on the radiation response of memories and microprocessors. Such results are largely empirical because the space community lacks the thorough understanding of internal device design that is present in the semiconductor device community. However, the data taken in space environments provides a direct comparison of different device types and manufacturers, and is readily available in the literature.

Results with high-energy protons are generally applicable to the atmospheric neutron environment, provided the threshold energy remains above approximately 30 MeV. Recent data for high-performance processors shows that the threshold energy is still above that range, and that errors in current-technology processors are still dominated by registers and cache memories.

Data for high-density memories show that the overall cross section for upsets from protons has steadily decreased as DRAMs have evolved. This trend is consistent with the elementary concept of a near constant threshold for sensitivity to long-range ions, along with the reduced area of advanced devices. However, charge collection in these structures is a complex process, and more modeling studies need to be done in order to improve the level of understanding of charge collection in structures that closely resemble real semiconductor devices.