4.2. Si CMOS Platforms: Radiation

Lloyd W. Massengill, Vanderbilt University

4.2.1 Introduction [1-7]

In extreme environments, as expected in earth orbit, exploratory space, or even in the specialized environments of nuclear reactors or nuclear weapons, microelectronic circuitry must endure a host of radiation hazards, including exposure to electromagnetic radiation in the form of ionizing gamma- and X-rays, as well exposure to highly energetic particles from a variety of sources: cosmic, solar, and terrestrial. In terms of the latter, even microelectronics operating in the not-so-extreme environment of a home or office\(^1\) are exposed to irradiation from natural sources.

Silicon is the most ubiquitous platform for microelectronic circuits, weathering decades of challenges from potential successors: III-V and II-VI semiconductors, ternary and quaternary compound systems, and now elemental carbon-based materials. All of these non-silicon systems have sterling attributes in various areas and many hold prominence in niche markets (many of which are described in other chapters of this book). However, silicon, and more precisely silicon CMOS, has enabled – and continues to enable – Moore’s Law and the information processing revolution enjoyed by our society.

Previous chapters have discussed the mechanisms associated with the long-term exposure of microelectronic devices to a wash of ionizing radiation – called total ionizing dose or TID. In

\(^1\) Extreme is, of course, a contextual term, and here unfortunately excludes synonyms of exciting, thrilling, risky, and/or dangerous, which we know describes most office environments.
CMOS, TID leads to parametric shifts in transistor transconductance, threshold voltage, and subthreshold leakage; and because the exposure is distributed, the effects are global across the circuit platform, affecting every transistor (to first order) without regard to the circuit design topology. As a result, TID is often addressed via the integrated circuit fabrication process or library transistor layout (all discussed in other chapters), as opposed to solutions based on circuit topology or function\(^2\).

Silicon CMOS operating in extreme environments is particularly susceptible to highly energetic particles, both cosmic and terrestrial. Unlike total ionizing dose radiation and its accumulated degradation of device parameters due to spatially- and temporally-averaged energy deposition, a *single* ionizing particle is capable of swiftly dashing the proper operation of a modern circuit.

Other chapters have introduced the mechanisms of single ionizing particle interactions with MOS devices (single events) – the Rutherford (Coulombic) scattering and occasional nuclear spallation, the induced energy transfer to semiconducting material (energy deposition), the liberation of mobile charge carriers (charge deposition), the conduction of this charge to metallurgical junctions and/or contacts (charge collection), and the terminal response of CMOS devices to the excess charge (single event photocurrent). These physical effects will not be repeated here; nor will this chapter attempt to provide a history lesson or comprehensive review of the broad radiation effects field. This chapter will concentrate on modern CMOS technology as a design platform for circuits destined for extreme environments.

From a circuit designer’s perspective, a single event interaction is a spatially and

\(^2\) There are exceptions to this generalization, such as precision analog current sources or voltage references, but these specialized topics are beyond the scope of this chapter
temporally localized effect, and can lead to a seemingly spontaneous signal transient within a singular region of a circuit. If this transient influences a critical circuit node (or nodes), it may lead to an incorrect, unreadable, or unstable system state; corrupt the legitimate information; and cause misinterpreted data – a possibility most design engineers would prefer to avoid.

CMOS technology has an interesting history as a platform for extreme environment circuit design – it has experienced a divergent evolution regarding radiation effects. On the one hand, Moore’s Law scaling\textsuperscript{3} of CMOS has forced the advancement of low-defect material systems and the elimination of trapping layers, leading to enhanced resilience to total ionizing dose. If not outright immunity, the incremental change in device parameters by radiation (such as threshold voltage shift) are often masked by much larger parametric effects (such as statistical variability) that are prevalent and unavoidable in scaled technologies. An example of this is the radiation-induced subthreshold leakage in sub-100nm CMOS – explained in Section 4.2.2.3 of this chapter. On the other hand, Moore’s Law scaling has decreased the switching energy of digital and noise margins of analog designs, seriously reducing the resilience of the technology to transient radiation effects such as single event ionizing particles.

This chapter describes silicon CMOS as a platform for circuit design for extreme environments – in particular, radiation environments. It addresses the unique radiation attributes of modern CMOS from three perspectives: device parameters, circuit topology, and layout geometry.

\textsuperscript{3} Since integrated circuit technology scaling long ago departed from rigorous constant-field or classical voltage scaling, the term ‘Moore’s law scaling’ is used throughout this chapter as a catch-all descriptor for dimensional scaling, energy scaling, and the introduction of novel materials – all to achieve the functionality progression best described by Gordon Moore, Intel co-founder, in 1965.
4.2.2 CMOS for Radiation Environments – Device Parametrics [8-15]

The Si CMOS platform presents circuit designers with the challenge of persistent parametric shifts when exposed to the extreme environment – the accumulating effect of total ionizing dose (TID). Fig. 4.2.1 shows the effects of TID on the current-voltage response of a sub-100nm CMOS NFET. Three primary device response mechanisms that impact circuit operation are typical, as described below.

4.2.2.1 Radiation-Induced Threshold Voltage Shift

Ionizing radiation can induce trapped charge in the insulating and interfacial gate regions of MOS structures. The electric field distortion created by this immobile charge modulates the effectiveness of the applied gate bias in inducing a drain-source conducting channel at the silicon surface, resulting in a shift in the effective threshold voltage of the device. Radiation-induced threshold voltage shift has been the most extensively analyzed, modeled, measured, and manipulated CMOS parameter in the history of radiation effects study (see Chapter 2.4), and mitigation solutions to this mechanism are presented elsewhere (Chapter 7.2).

However, it is not material innovations or circuit techniques (of which there are many) that present the best solution to TID-induced MOS threshold voltage shifts in modern CMOS platforms, instead it is Moore’s law scaling. Dimensional and voltage scaling of modern CMOS (below the 180nm technology node) has reduced the volumetric dimension of gate insulators to a range that does not support charge trapping – charge escapes the insulator readily through transport mechanisms such as tunneling. Gate oxide thicknesses less than 10nm show virtually no threshold voltage shift to ionizing radiation.

Even so, the TID issue has not been eliminated from CMOS. Modern CMOS platforms
still exhibit significant vulnerabilities to total ionizing dose. Even though gate stack materials have been thinned with dimensional scaling, isolation insulators and buried oxides remain thin enough to trap charge and induce undesirable effects in highly-scaled technologies. More on this later.

4.2.2.2 Radiation-Induced Transconductance Degradation

Trapped charge at, or near, an MOS conducting channel interface can reduce the device transconductance through two ways: carrier mobility degradation in the channel or increase in surface resistivity in ancillary regions such as LDD. Modern CMOS relies on high-gain transistors to achieve the switching speeds and signal propagation required by energy scaling. Transconductance degradation in CMOS platforms can lead to dynamic failures due to (1) incomplete switching at critical nodes, (2) signal propagation delays along synchronous data paths, or (3) the inability to drive output internal or external logic fanout. Weakened-drive transistors test the operational resilience designed into the CMOS circuit topology, especially at highly-scaled technology nodes.

4.2.2.3 Radiation-Induced Leakage Current

While Moore’s law scaling has virtually eliminated the problem of TID-induced MOS threshold voltage shift described in Section 4.2.2.1, and robust circuit design can control the deleterious effects of TID-induced transconductance degradation described in Section 4.2.2.2, scaling has severely exacerbated a third TID-induced radiation effect: radiation-induced leakage current.

The thinning of gate insulators and defect reduction at critical interfacial regions
produced by dimensional scaling do not necessarily extend to other insulating regions, such as isolation layers or buried oxides. For example, ‘trench isolation’ is a common technology practice to reduce unwanted inter-device signal communication – yet can trap charge in an extreme environment. TID-induced trapped charge accumulates in these thick layers, leading to surface and buried leakage (conduction) paths both inter- and intra-device. Fig. 4.2.2 shows a representation of these conduction paths.

Modern CMOS platforms are particularly susceptible to leakage, especially low-power variants. Leakage can induce significantly reduced performance, or even functional failure, of CMOS circuits because of (1) charge accumulation on dynamic circuit nodes, (2) unexpected current compensation via an ‘off’ device driving a switching node, or (3) standby power increasing beyond the acceptance limit. TID-induced current leakage is a critical problem for modern CMOS platforms, and the severity increases with dimensional scaling.

4.2.2.4 Silicon on Insulator (SOI) and other Isolation Technologies

The discussion of this section has centered on bulk CMOS, but silicon-on-insulator (SOI) technologies also represent a mainstream variant of the silicon CMOS platform. SOI has served the military and aerospace enterprise well for several decades as a radiation-hardened technology. More recently, SOI has found widespread acceptance in the commercial community.

The inherent radiation resiliency of SOI resides in the geometrically-constrained active volumes (limited silicon regions for radiation energy deposition and ionization) and inter-device isolation (limited device-to-device communication of deposited charge) – more on these two topics in the next sections. However, this SOI isolation is achieved via insulating regions
(typically deposited SiO₂) that are often much thicker than gate insulator layers and possess a higher level of charge trap precursors than the highly-optimized gate stacks. As a result, SOI isolation regions exhibit all of the radiation responses described in Sections 4.2.2.1 – 4.2.2.3, usually to a more troubling degree than gate regions. For example, in very thin SOI devices, the buried insulating oxide (BOX), and any associated TID-induced trapped charge, may act as a secondary gate, modulating a pseudo-channel at the bottom of the MOS structure (inducing back-side leakage currents) or electrostatically coupling to the top surface channel (modulating the device threshold voltage). Modern multi-gate, vertically-stacked structures (e.g. MUGFETs, FINFETs, TRIGATEs) also exhibit TID-induced leakage. In modern low-power SOI technologies destined for extreme environments, these parasitic effects are of significant concern.
4.2.3 CMOS for Radiation Environments – Circuit Topology [16-47]

Si CMOS is acutely sensitive to transient radiation effects that generate illegitimate signals that may perturb, overwhelm, or even circumvent the signals impressed on the platform by the circuit topology – the phenomenon of single ionizing particle interactions or single events.

From a physical perspective, a single event phenomena (SEP) or just a single event (SE) refers to the interaction of a single ionizing particle with a semiconductor device. Each event is a localized interaction that does not depend on flux or total exposure – i.e., it is considered (to first order) spatially and temporally random, and uncorrelated. It has been said (though not scientifically verified) that single events perfectly obey Murphy’s Law.

From a functional perspective, the single event terminology has roots in the response of various components or circuits to the event itself – i.e., the effect upon the platform, rather than the cause.

One such platform effect is the single event upset (SEU) – the corruption of a digital bit of information within a memory circuit, such as a static random access memory (SRAM) or a dynamic random access memory (DRAM). SEUs are usually considered directly observable and measurable, vice single event faults and transients discussed in Section 4.2.3.3. One often finds the term SEU used interchangeably with single event error or soft error – the observable, measurable manifestation of an SE as an incorrect circuit operation (usually a system response involving corrupt data or incorrect state). The term “soft” implies that the error is electrically correctable, as opposed to a permanent, or “hard,” error. The single event error rate or soft error rate (SER) is the frequency of errors in a particular environment (e.g. a specified orbit, mission trajectory, or terrestrial locale). For space applications, the SER is typically quoted in
error per bit-day or errors per device-day. For terrestrial applications, the SER is typically quoted in failures in time (FITs) representing errors per billion hours of operation. Soft errors are discussed in Section 4.2.3.2.

Another platform effect is the **single event transient or SET**. DSETs (digital) and ASETs (analog) represent propagating single event pulses that may compete with legitimate signal flow in all types of analog, digital, and mixed-signal platform circuitry. This very important class of radiation effects is discussed in Section 4.2.3.3.

### 4.2.3.1 Basics of CMOS Circuit Response to Ionizing Particles

In order to understand the response of complex CMOS circuitry to single events, it is advantageous to start with the inverter, as shown in Fig. 4.2.3. The CMOS inverter represents the constitutive template of all CMOS static logic – its dynamic characteristics form the basis for discussing the response of all CMOS digital circuits to single-event-induced currents.

The bi-stable switching nature of an inverter can be characterized by the ability of the inverter to charge and discharge a capacitive load at the point when the output and input voltage are equal, known as the inverter unity transfer voltage, \( V_I \). This drive capability tracks the weaker of the pull-up or pull-down devices, defined by the worst-case transconductance \( K \). The ability of an inverter to drive a capacitive load is known as the **intrinsic time constant of the inverter**, \( \tau_i \) – a figure of merit parameter closely tied to the technology node - and is given by

\[
\tau_i = \frac{C_O}{V_{DD}^2} K
\]

where \( C_O \) is the total output capacitance and \( V_{DD} \) is the power rail voltage.

The characteristic rise and fall times of an inverter (the transition between 0.1 and 0.9 of the rail voltage) are directly related to \( \tau_i \) and the threshold voltages of the driving devices.
through a numerically-derived mapping shown in Fig. 4.2.4, where $\alpha_n = V_{TN}/V_{DD}$, $\alpha_p = |V_{TP}|/V_{DD}$, and the rise/fall time is a multiple of $\tau_i$. This relationship maps the response of CMOS circuitry to both total ionizing dose and single event transients back to the platform technology, and is important to upcoming discussions.

In the early days of single event observation, the charge collection time constant for the event ($\tau_E$ - typically on the order of picoseconds) was significantly faster than the intrinsic time constant of the digital gates of the day ($\tau_i$ - on the order of nanoseconds at that time). Thus, the single event photocurrent modulated the information node voltage on a time scale much faster than the circuitry could respond. In a sense, the voltage change from the event was “instantaneous” (relative to any circuit time constant), as was the change in capacitive charge on the affected node. Thus the details of the current waveform could be mathematically absorbed in a Dirac delta function and removed from consideration. This led to the construction of a modulus based on charge, rather than current, that has persisted to this day.

$4.2.3.2$ Integral Charge Collection and Circuit Upset – The Concept of Critical Charge

Under the conditions of $\tau_E \ll \tau_i$, the voltage perturbation induced on the output node of an inverter by a single event is given by:

$$V = \int i_{SE}(t) \frac{dt}{C_N} = \frac{Q_{collected}}{C_N}$$

(4.2.2)

where $Q_{collected}$ is the total integrated charge delivered to/from the node by the single event photocurrent and $C_N$ is the total nodal capacitance. Further, if a voltage perturbation threshold for information loss (i.e. the noise margin $V_{nm}$) can be defined, then the amount of charge necessary to induce that voltage can be calculated from Eq 4.2.2. From this analysis, the quantity of $Q_{crit}$ or $Q_c$ – the critical charge – emerges:
\[ Q_{\text{crit}} = C_N \cdot V_{nm} \]  \hspace{1cm} (4.2.3)

\( Q_{\text{crit}} \) is usually defined as the minimum amount of charge collected at a sensitive node necessary to cause a circuit upset. It is important to note that \( Q_{\text{crit}} \) is a circuit response metric, not a technology or physical quantity. A circuit possesses a certain \( Q_{\text{crit}} \) regardless of the environment – it is not a property imposed by, nor tied to, a single event.

**SEUs in DRAMs and Dynamic Logic**

The CMOS design platform with the most straightforward response to charge collection and \( Q_{\text{crit}} \) is the dynamic random-access memory (DRAM). DRAM technology refers to the broad class of information storage platforms, usually one-transistor designs, which store packets of charge to represent binary information. The key to DRAM radiation faults is that the information storage is passive (no active regeneration path), and any (no matter how small) disturbance of the stored information by a particle strike will persist until corrected by external circuitry. There is no inherent refreshing of this charge packet (e.g., charge resupply through a load device) and no active regenerative feedback as one observes in latches and SRAM cells. What is so often referred to as a ‘bit flip,’ the transition from one stable binary state to the other, is not required in DRAMs for an SEU to occur. A degradation of the stored signal to a level outside the noise margin of the supporting circuitry is sufficient to lead to erroneous interpretation and a resultant error.

The most prevalent soft error source in DRAM arrays is single-event charge collection within each binary cell, caused by a single-event strike in or near either the storage capacitor or the source of the access transistor. A resultant cell upset is usually observed as a 1→0 transition; the collected charge relaxes a stored charge state.
Upsets can also occur in DRAMs due to bit-line strikes. When the bit lines are in a floating voltage state (e.g., during a read cycle), DRAMs are sensitive to the collection of charge into diffusion regions that are electrically connected to the bit access lines. This collection could arise from any of the access-transistor drains along the bit-line length, or from a direct strike to the differential sense amplifier. The bit-line SEU mechanism is the reduction of the sensing signal due to a charge imbalance introduced on the precharged bit lines, either prior to or during the sensing operation. Because bit-line strikes are only possible during the floating precharge and sensing stages of operation, temporal characteristics of the strike in relation to the clocking signals are critical. Also, because the duty cycle of these stages to the overall cycle time increases with increasing overall clock frequency, the bit-line soft error rate is inversely proportional to DRAM cycle time. In contrast, cell upsets are independent of the DRAM cycle time. Bit-line errors also show a strong inverse correlation with the signal charge. With technology scaling, bit-line errors have become increasingly important.

DRAM platforms generally lead the Moore’s law scaling trend with the most aggressive feature size and density scaling. The soft error performance of this family of circuits is intimately tied to scaling, so DRAMs are often used as technology roadmap monitors. However, for extreme environment resiliency, this assumption is debatable. First, many different physical storage structures are used in DRAM manufacture, ranging from various stacked capacitor designs to buried trench capacitors. These physical structures have been shown to have a dramatic effect on the observed soft error rate. Second, the concomitant effects of voltage/energy scaling on the stored bit energy and the effects of dimensional scaling on the capacity of the storage capacitance both reduce $Q_{\text{crit}}$. However, dimensional scaling has also reduced the projected surface area of each bit cell, so the cross-section presented to the
environment is reduced. These competing effects on soft errors has created a situation in modern platforms where the soft error rate per bit has fallen with scaling. However, the overall error rate per chip remains relatively constant.

A close cousin to dynamic memory is dynamic logic. Like the DRAM, dynamic logic relies on passive charge storage for information representation and operation. Also, like the DRAM, this information is extremely vulnerable to single event charge collection because there exists no resupply mechanisms to restore depleted charge. The concept of $Q_{crit}$ applies similarly to dynamic logic.

SEUs in SRAMs

The upset process in SRAMs is quite different from DRAMs, due to the active feedback in the cross-coupled inverter pair that forms a typical 6-transistor SRAM memory cell. When an energetic particle strikes a sensitive location in an SRAM (typically the reverse-biased drain junction of a transistor biased in the “off” state), charge collected by the junction results in a transient current in the struck transistor. As this current flows, the restoring transistor (“on” transistor) sources current in an attempt to balance the particle-induced current. The current mismatch at the common drain node modulates the node voltage – it is this voltage transient that triggers a potential upset in the SRAM cell. Similar to an asymmetric write pulse, this voltage transient can overpower the positive feedback of the cell and induce an erroneous static state in the memory cell.

From a technology standpoint, the recovery time of an SRAM cell depends on the restoring transistor current drive, the temporal characteristics of the single event charge collection transient, and the time constant of the positive feedback loop through the opposing
inverter. This latter time constant is related to the cell write time and in its simplest form can be thought of as twice (one complete feedback loop) the minimum propagation delay time of the inverter, as discussed in Section 4.2.3.1. This delay time constant is thus a critical parameter for determining SEU sensitivity in SRAMs – the smaller the delay, the faster the feedback mechanism can respond to voltage transients, latching the perturbed state. Thus, with the scaling of $\tau_i$, the more susceptible the SRAM becomes to single events. Obviously this has implications for the sensitivity of high-speed CMOS platforms, as discussed in the next section.

4.2.3.3 Temporal Charge Collection and Circuit Response – When Critical Charge is Not Really ‘Critical’

Thanks to Moore’s law scaling, CMOS circuits are not only smaller, but faster with each technology node. The assumption that $\tau_E$ – the charge collection time constant of the single event – is much faster than $\tau_i$ – the intrinsic time constant of the CMOS gate – is dubious; as is the direct applicability of $Q_{crit}$. Single event effects in a modern CMOS platform involves the engagement of charge collection mechanisms at the semiconductor regions and electrical response of the circuitry, all within the same temporal window. In these cases, the voltage signal on a node is no longer simply controlled by $Q_{collected}$ - collected charge, but rather by the dynamics of the collected charge over time – $Q_{collected}/t$, which is, of course, current. In other words, the time profile of single event charge delivery is critical.

The manifestation of this “temporally-responsive” charge collection in CMOS platforms is a nodal response that is not fixed by the energy of the impinging particle, but is instead modulated by the dynamics of the connected circuitry. For example, in the early days of single event study, an adequate circuit representation for the charge collection process was the universal
“double-exponential current source.” In modern circuitry, we find that the temporally-responsive nodal charge collection current behaves as shown in Fig. 4.2.5. The ‘shelf’ or ‘plateau’ region of the photocurrent, which is simply an induced balance of charge collection current and resupply current, plays an important, and in some cases, dominant role in modern CMOS technologies. Long node voltage pulses, much longer than predicted by the physics of direct charge collection of Section 4.2.3.2, have been observed. Such complex temporal characteristics are critically important to the ultimate response of CMOS platform circuitry to ionizing particles – these temporal voltage transients are known as ‘single event transients’ or SETs.

The following subsections discuss these temporally-complex single event transients in both digital logic (digital single event transients or DSETs) and analog mixed-signal circuitry (analog single event transients or ASETs).

**Digital Logic and DSETs**

In CMOS synchronous logic\(^4\) circuits, the concepts of ‘upsets’ and ‘errors’ are distinct from the memory circuits of Section 4.2.3.2. A single event soft fault (SESF or SEF) is the corruption of a digital *bit* of information within a synchronous circuit, usually in a register or embedded memory element, that exists but may be latent and has not yet propagated to – or caused effect on – an observable state of the system. Many soft faults remain latent, or are masked by logical or temporal conditions, and never become physically observable; thus an SEF

\(^4\) For the remainder of this discussion, we will use the term *synchronous logic* to denote a circuit composed of combinational logic gates intermixed with storage elements (registers/latches) that operate in a clock-synchronized fashion, i.e., static, not dynamic, logic.
may never become an “error” and thus are difficult to directly measure. A single event error or soft error represents the observable, measurable manifestation of an SEF as an incorrect circuit operation (usually a system response involving corrupt data or incorrect state).

In order for a single ionizing event to affect CMOS logic as an error, four events must occur: (1) the event must generate a voltage transient that propagates out of the affected gate and through the logic signal path (the complement of this occurrence is called electrical masking) – this propagating transient signal is called a digital single event transient or DSET, (2) the transient signal must find an open propagation path through active combinational logic paths to a latch or register (the complement of this occurrence is called logical masking), (3) the transient must arrive at a latch element and meet the setup and hold characteristics of the latch (the complement of this occurrence is called temporal masking), and (4) once latched, the fault must impact the operation of the digital state machine in some way (i.e. not be overwritten by a subsequent state change or be ignored by a branch operation, called operational masking). The resiliency of modern CMOS technology to single events is intimately tied to these four operations, so each is discussed here from the perspective of the CMOS design platform.

Electrical Masking:

A detailed analysis of electrical masking is, of course, best performed by advanced EDA/CAD tools under the conditions of circuit operation and with nonlinear parasitics included. However, much can be learned about the impact of CMOS technology attributes on logic upset through a solution of the first order analytical MOS I-V equations driving a capacitive load for a few specialized single-event charge collection waveforms. A pulse that propagates is considered a DSET. On the other hand, a pulse that does not propagate is considered electrically masked, as
shown in Fig. 4.2.6.

A single-event-induced voltage transient of rail-to-rail excursion will propagate unhindered (no electrical masking) provided the following two conditions are met:

1) the slowest of the rise or fall time constant of the originating SET voltage transient is faster than the characteristic rise/fall time of the inverter, AND

2) the pulse width of the originating SET voltage pulse, measured at $V_I$, is greater than $\tau_R + \tau_F$

If we assume that the SET pulse immediately converges to the characteristic waveform of the gate (a remarkably good assumption), then the inverter pair following the strike must transition with characteristic rise/fall time constants. For one inverter pair to perform a complete transition to a stable output, the first inverter must complete a characteristic high-to-low transition ($\tau_F$) and the next inverter must complete a characteristic low-to-high transition ($\tau_R$). The output voltage is then, and only then, at full rail with no memory of the previous state. This is critical for unattenuated propagation. Pulses that do not meet these conditions are attenuated by subsequent gates, will diminish, and eventually disappear (electrical masking).

SET voltage pulses that do not initially reach the rail will be amplified, eventually reaching rail-to-rail excursions, and propagate indefinitely as characteristic pulses if more stringent requirements are met:

1) the slowest of the rise or fall time constant of the originating SET voltage transient is faster than the characteristic rise/fall time of the inverter, and

2) the pulse width of the originating SET pulse, measured at $V_I$, is greater than $\tau_i \sigma$, where

$$\tau_i \sigma = \frac{V_{DD}^2}{\beta \left| V_{SR} - V_{TN} \right|^2 (V_{DD} - V_{SR} - V_{TP})^2}$$

where $\tau_i$ is the technology intrinsic time constant of Eq 4.2.1, $\beta = K_p / K_n$, and $V_{SR}$ is the
sub-rail pulse voltage magnitude.

A third case of DSET pulse propagation is an initial SET pulse with a slower rise/fall time than the characteristic time constant of the initiating inverter. In the previous two cases the output waveform parameters $\tau_R$, $\tau_F$, and $\tau_D$ are defined completely by the drive capabilities of the inverter and the load capacitance. However, for SETs that generate a rising or falling edge slope less (slower) than the characteristic inverter slope, the inverter output time constant is modulated by the input time constant. The shape of the output waveform is nearly exponential, even given a slow ramp input. For either a slow input ramp or slow exponential, the output pulse has a delay characteristic given by:

$$
\Delta = C_1 \frac{t}{C} \quad \text{if} \quad \tau_{IN} < \tau_C
$$

(4.2.5)

and

$$
\Delta = C_2 \frac{t}{IN} + C_3 \frac{t}{IN} \quad \text{if} \quad \tau_{IN} > \tau_C
$$

(4.2.6)

where $\Delta$ is the delay time of the falling (rising) edge of the output pulse measured from the rising (falling) edge of the input pulse, $C_1, C_2, \text{ and } C_3$ are empirical constants, $\tau_C$ is the characteristic $\tau_R$ or $\tau_F$ depending on the transition, and $\tau_{IN}$ is the rising (falling) edge time constant of the input pulse. Thus, the output edge delay is dependent on the input edge time constant if the input is slower than the characteristic inverter time constant. Under certain conditions of asymmetry, this delay of the output edge relative to the input edge can lead to an output pulse that is wider (or "broader") than the input – ‘single-event pulse broadening’, a phenomenon that has been experimentally observed in modern CMOS platforms.
CMOS platform effects on electrical masking:

SETs with very narrow pulse widths (i.e., short transients) relative to the characteristic time constant of the logic tend to be filtered (attenuated) along the data path by the finite response time of subsequent logic gates, while long pulses can propagate unattenuated deep into the logic. Thus, Moore’s law scaling reduces electrical masking – as circuits become faster, a higher percentage of single-event-generated transients meet the requirements for infinite propagation. Natural electrical masking cannot be assumed for modern Si CMOS design platforms.

Logical Masking:

Logical masking is the termination of a DSET because of a blocked logic path due to the system state at a particular time. As Fig. 4.2.7 shows, the active combinational path at any point in time depends on the dynamic state of the logic as determined by the particular code vectors under execution (the present ‘state’ of the logic). If such a path exists, the DSET will propagate through the intervening gates en route to a latch. If an active path does not exist, the DSET will be masked by a blocking gate.

CMOS platform effects on logical masking:

Logical masking in modern CMOS platforms is not heavily dependent on technology scaling. Logic is logic, there is just more of it in modern CMOS. Therefore, we see more potential combinational logic paths in scaled platforms, and as such more chances for DSET propagation. But the results of logical masking are so intimately entwined with the function of the circuit, the particular dynamic operation at the time of exposure, and the stochastic nature of
the state machine, that it is difficult to draw direct parallels between logical masking trends and the technology platform.

Temporal Masking:

Temporal masking is one area where clear links to the CMOS technology platform have been observed. This form of masking can best be described using the setup and hold characteristics of an edge-triggered latch. In order for any signal to be captured by a latch, it must meet the inherent setup and hold requirements: the signal must (1) arrive at the latch input at a time sufficiently before the clock signal transition so that transitional dynamics have settled out when the clock edge appears (setup) and (2) maintain a fixed level for sufficient time after the clock edge for the latch to act upon the data through a potential change of state (hold).

For single event transients, of irregular pulse widths, and arriving at times unsynchronized to the system clock, meeting the setup and hold requirements of a latch is not at all guaranteed (or perhaps even likely) – if not, the SET will not be captured by the latch – it will be temporally masked.

A temporal measure of latch vulnerability to a DSET arrival is the ‘window of vulnerability’ or WOV – defined as the time duration in which a latch element is sensitive to an SET, as shown in Fig. 4.2.8. The related ‘sensitivity window plot’ is a representation of the dynamically-varying sensitivity of a latch to a DSET arrival as a function of time (as shown in Fig 4.2.9) when extended beyond a single latch to a complex circuit topology with time-varying windows of vulnerability.

CMOS platform effects on temporal masking:
The CMOS platform has a direct impact on temporal masking. Switching energy scaling allows increased clock frequencies with each technology node. Scaled clock frequencies implies increasing clock transitions per unit time, and since SET latching is tied to the window-of-vulnerability associated with each edge trigger, increasing edge frequency induces a similar increase in SET latching (to first order) – this intuitive relationship has led to the often quoted (and more often extrapolated) linear dependence of SET faults with technology clock frequency.

However, the relationship of switching energy scaling with SET latching also includes effects on the window of vulnerability itself (as opposed to simply the ratio of a fixed WOV to a shrinking clock period) and pulse shaping effects (as opposed to a fixed SET pulse width). Energy scaled circuits tend to possess more responsive latch front-ends, with reduced setup and hold times over older counterparts – a counter influence on DSET fault scaling with frequency scaling. However, as seen in this section, energy scaling enhances the propensity for circuits to successfully propagate DSET pulses to latch elements – a compounding influence on DSET fault scaling with frequency scaling. The resultant impact of CMOS technology scaling on DSET temporal masking is not straightforward, with the literature demonstrating non-linear relationships between technology node, clock frequency, and DSET combinational logic faults.

For the extreme environment design technologist, careful consideration of both electrical and temporal masking effects in modern CMOS platforms is warranted. This consideration usually involves Monte Carlo EDA simulations (across vulnerable nodes and single event arrival time) of DSET propagation using full parasitic extraction and non-linear electrical analysis of typical data paths.

*Analog / Mixed-Signal*
To this point we have discussed the interaction of single event charge collection currents with the operation of digital circuits, both static and dynamic. However, as this author is fond of telling his circuits students, “digital is just a special case of analog.” In other words, all circuits operate on continuous signals (analog), it is just that digital circuits attempt to force the signal into two semi-stable extremes (high or low). Single events can impact analog circuitry just as readily as digital, although analog effects have enjoyed a less comprehensive long-term study in the literature.

The manifestation of a single ionizing event in analog circuitry is the generation of a signal pulse – a transient – that can be indistinguishable from legitimate signal modulation. Such a transient is called an analog single event transient or ASET. As might be expected in the continuous-time domain, analog transients appear in an unlimited variety of signal forms; nevertheless most observed transients fall into three categories that describe their basic voltage signature: positive transients, negative transients, and bipolar transients (see Fig. 4.2.10) – all appear in and are important to analog CMOS platforms.

The first confirmed malfunction on a spacecraft attributed to radiation-induced ASETs occurred shortly after the launch of NASA’s TOPEX/Poseidon satellite in 1992 – an anomaly caused by single event transients in an OP-15 operational amplifier. Over the years, various problems in space-based systems have been attributed to ASETs. The circuit effects of ASETs can range from a spurious glitch in an analog signal stream (not particularly significant due to inherent noise tolerance) to long-duration, rail-to-rail perturbations (very significant as these events can trigger widespread data loss or even system shutdown.)

Because space systems contain many different types of linear circuits – such as operational amplifiers, voltage comparators, voltage references, pulse-width modulators, voltage
controlled oscillators, DC/DC converters – ASETs pose a serious systems threat in extreme environments.

Analog SET propagation and fault generation is much more heavily dependent on the interpretation of the signal than in digital (where an incorrect logical state is relatively unambiguously defined). As such, there exists no universal ‘upset metric’ for analog single event effects – the ‘result’ or severity of the phenomenon is often not known until down-stream signal processing has completed. However, application-specific metrics for ASETs have been proposed, and these provide design information for analog/mixed-signal platform users.

Analog Upset Metric:

Because most analog topologies are bandwidth limited, because analog circuits typically employ noise cancelling features, and because analog circuits commonly exhibit continuous voltage as the relevant state variable, both the frequency characteristic and the absolute magnitude of an ASET pulse are important to the severity of the system response. For example, a very short pulse in the time domain (high frequency) with a large voltage magnitude may be filtered and of less import to the system response than a wider pulse (lower frequency) of lesser voltage magnitude. Likewise, a very long pulse of small voltage magnitude may not exceed the noise tolerance of the downstream circuitry and be of little concern to the system.

A generally-applicable ASET representation is the ‘amplitude / pulse-width’ scatter plot, as shown in Fig. 4.2.11. This plot displays the signal modulation at a particular critical or vulnerable node in terms of peak amplitude and full-width half-max pulse width – both of which are important extrema to analog designers. A system-imposed safe operating area defined by failure coordinates of magnitude and pulse width superimposed on the scatter plot of Fig. 4.2.11
reveals the vulnerability of a particular design to measured ASET data for the technology platform. This boundary establishes an ‘upset’ threshold for analog SETs – from this metric, upset rates and confidence limits can be quantified.

4.2.3.4 Design Implications of Energy Scaling

Critical charge and propagation delay have both decreased with Moore’s Law scaling of technologies. Because of the concomitant scaling of supply voltage and capacitance (switching energy scaling shown in Fig. 4.2.12) which both appear in Eq. 4.2.1, critical charge tends to scale geometrically with technology node. Propagation delay scales less dramatically, but also depends on energy scaling. As a result, Si CMOS platforms exhibit several dynamic sensitivities to the environment:

As $Q_{crit}$ decreases, circuits become less robust to charge perturbation and a larger percentage of design topologies become sensitive to radiation with each technology node. Soft errors are not limited to memory elements, but also appear in other subsystems found in Si CMOS platforms: static and dynamic logic, analog signal processing, reference and biasing circuitry, phase-locked-loop and delay-locked-loop clocking systems, clock distribution circuitry, and even input/output drivers.

Also, with $Q_{crit}$ values falling below 1fC, Si circuits are sensitive to an expanding sampling of the environment, ever increasing with each technology node – that is, more of the environment appears ‘extreme’ to the platform. Not only are modern CMOS platforms sensitive to highly-energetic galactic cosmic ions and terrestrial neutron reaction products, but also to particles once considered irrelevant: low-energy alphas, directly-ionizing protons, even subatomic muons. Many of the latter particles are extremely abundant in various environments, both space and terrestrial, and because of their abundance (i.e. flux), pose a threat of
overwhelming architectural mitigation schemes such as EDAC.

Energy scaling has also exacerbated both analog and digital SET effects. Single event transients propagate freely in modern CMOS platforms and, in many cases, are indistinguishable from legitimate signals. Mitigation techniques based on discrimination, such as temporal filtering and gating lose effectiveness under these conditions.

Redundancy – either internal (DICE, DCC), local (TMR, DMR) or architectural (ECC, EDAC, etc) – remains the most accepted soft error mitigation approach to combat energy scaling in Si CMOS platforms destined for extreme environments. Local charge dissipation techniques are resource costly, and violate the fundamentals of energy scaling. Charge collection mitigation through the fabrication process (e.g. epitaxial layers, buried layers, or isolation regions) has helped contain the problem, but these variants are extremely costly, and are often overwhelmed by non-radiation performance constraints.

There are, of course, caveats to each of these generalities. In particular, caveats to the effectiveness of redundancy, of which the Si CMOS platform user should be aware, are discussed in Section 4.2.5.
4.2.4 CMOS for Radiation Environments – Layout Geometry [48-55]

As CMOS technology evolved into the deep submicron regime, single event response effects were observed that could not be explained by the classical charge deposition/collection mechanisms described in Section 4.2.3. Early evidence that “something else” was going on in scaled CMOS, something that could not be explained by first-order charge collection models, was the observation of multibit errors due to single ion strikes. These anomalies began to appear in earnest as CMOS technology passed the 250nm technology node – concurrently in a BAE 4 Mbit 10-T radiation-hardened SRAM fabricated in a dual-well, thin epitaxial 250 nm CMOS technology and in a commercial SRAM fabricated in a 130nm bulk CMOS technology.

Further complicating evidence appeared at technology nodes below 100nm. Digital circuitry specifically hardened against soft errors via redundancy – that were ‘known’ to be resilient to single-node charge collection – began to exhibit single event errors.

4.2.4.1 Spatial Response – The Region of Influence

It turns out that dimensional scaling is the culprit for many of the aforementioned single event response anomalies. The cause is the expanding region of influence of single event charge deposition relative to the feature size of platform circuitry. The ionization profile of an energetic ion is fixed by the attributes of the ion, not by the platform (to first order, ignoring some modulation by the material parameters, e.g. doping). However, feature size – and more importantly, packing density – scales with each technology node. The ‘relative’ size of an ionization track imposed on a sub-100nm layout is quite large. The impact of the energy deposition by way of perturbed fields and potential modulation is larger still. On a modern CMOS technology platform, it is not unusual for a single ion to simultaneously affect many
circuit nodes within a circuit primitive, and even multiple circuit modules within a complex topology.

Easily-observable multiple-bit errors in dense memory arrays have shown an exponential increase with technology scaling – these effects have been controlled via error detecting and correcting (EDAC) codes and interleaving of memory cells to achieve substantial physical spacing between same-digital-word bits. However, the more insidious ‘region of influence’ problem in CMOS platforms is the communication of single event charge collection among disparate nodes within a complex circuit. The mechanisms driving the ‘region of influence’ phenomenon follow:

*Charge sharing:*

A critically-important ‘region of influence’ effect for modern CMOS technology platforms is ‘**charge sharing**.’ Single event charge sharing\(^5\) is the collection of charge on multiple circuit nodes that is deposited by, or induced by, a single particle strike, as shown in Fig. 4.2.13. Another way to describe charge sharing is ‘single strike event, multiple charge collection events.’ Charge sharing appeared in CMOS technologies at the 250nm node, and its effect has intensified with each process technology transition.

The most troubling attribute of charge sharing is that it can thwart conventional soft error hardening schemes employed in many CMOS platforms. Any design technique that relies on spatial information redundancy is susceptible to charge sharing, for redundancy always assumes unperturbed information is ‘hidden’ from the affected information, usually via one or more

\(^5\) ‘Single event charge sharing’ is used here to distinguish this phenomenon from ‘electrical charge sharing’ associated with the normal clocking operation of MOS digital circuits.
circuit nodes separated from the ‘protected’ node.

The dual interlocking storage cell (DICE) is an example of a hardened CMOS design that is extremely vulnerable to charge sharing. The concept of the DICE is the use of four interconnected inverters, rather than two, to create a memory storage element. Instead of node voltages representing high and low binary values, the DICE design incorporates two high values and two low values (dual redundancy on each) across the four inverters. The inverters are interconnected in such a way that a state change on any one node is corrected by the unperturbed value via positive voltage feedback. DICE (and several similar variants employing local voltage redundancy) have been stalwart digital memory cells for extreme environment design platforms for decades.

Single event charge sharing was shown to cause DICE upsets in 2005. This was a watershed event because (1) it demonstrated that any spatially redundant technique for radiation resiliency is suspect in modern CMOS technologies and (2) it heralded a host of failure mechanisms, some yet to be observed, in dimensionally-scaled CMOS technologies as a direct result of the many-device-encompassing ‘region of influence’ of single particle strikes.

Charge sharing must be an integral consideration for all radiation-hardened designs in current and future CMOS technology platforms.

**Parasitic Bipolar Conduction:**

The region of influence for a single event in modern CMOS platforms is not limited to those transistors enclosed within the ‘charge cloud’ of the ion; the influence can be far-reaching by way of induced parasitic effects. The most important of these in modern CMOS is the induction of *parasitic bipolar action* due to single-event-induced well collapse, as shown in Fig.
4.2.14.

Relative to the operational charge of modern CMOS, a single event deposits a massive amount of charge deep into the silicon substrate. While some of this deposited charge may be collected at surface devices, a large portion will transport to substrate and well contacts (taps). In 2004, at the 130nm technology node, it was observed that the transport of single-event charge within a well region (toward the well tap) could debias the body region underneath the surface channel region, induce charge injection, and trigger parasitic bipolar conduction between the source and drain of one or more surface transistors. From a circuit-response perspective, the induced bipolar current can modulate a circuit node just as would a direct single event strike, and upset a cell just as readily as direct charge collection from the ion.

Well Bias Collapse:

An ancillary mechanism to parasitic bipolar conduction is well bias collapse, as shown in Fig. 4.2.15. The charge deposition process of a single event ion, when within a confined region such as a well, will perturb the bulk potential of the region – in modern sub-100nm CMOS for several micrometers surrounding the strike location – triggering charge injection and bipolar conduction as described in the preceding subsection some distance from the strike. At the 90nm technology node, the effect has been shown to cause simultaneous multi-bit upsets of tens of cells – and the effect worsens with dimensional scaling. Mitigation may be attempted by high

---

6 While this was deemed a ‘new’ observation for modern bulk technologies in 2004, the mechanism had been known since the late 1980’s. Similar debiasing of the substrate (body) region of MOS devices has been observed in body-tied SOI technologies exposed to single event radiation for decades. In the SOI case, the diffusion of excess charge through a highly-confined body region toward the body tie would debias the body, create charge
spatial frequency of well taps; however, this requirement is resource expensive and only serves to mitigate, not eliminate, the problem.

4.2.4.2 Pulse Quenching

A DSET mechanism that is a direct result of single event charge sharing is pulse quenching. Discovered in 2009 at the 130nm technology node, pulse quenching is the name given to the phenomenon of single event pulse width reduction due to delayed charge collection (via charge sharing) as the pulse is en route through the data path. Pulse quenching appears only when the signal propagation along an electrical path occurs on the same time scale as charge sharing among adjacent devices (a condition that did not exist prior to the 130nm technology node). It is the interaction of these coincident events that modulates the propagating pulse.

Fig. 4.2.16 shows a comparison of conventional DSET creation (top) and the relevant steps involved in a pulse quenching event (bottom). The mechanism involves a race condition – here the PMOS gate control signal transient and the diffusion of holes to device P2. If the SET gate signal arrives prior to the diffusive charge collection, P2 is turned off, the drain voltage changes state (H to L) due to the expected inverter response, and the P2 drain is receptive to charge collection. Upon the delayed arrival of the diffusive charge, the drain of P2 collects charge and the voltage state is modulated once again (L to H) – back to the initial state. Thus, the single event initiates a double state change at the drain of P2 and an abbreviated (or quenched) voltage pulse width is observed at the output of the inverter.

Pulse quenching is important because it directly affects electrical and temporal masking injection, and induce parasitic bipolar current flow between source and drain. The effect was predicted for bulk CMOS in the 1990's.
in CMOS platforms – the pulse width is a key failure parameter and impacts resilient design choices. Pulse quenching has also explained a counter-intuitive weak dependence of DSET pulse widths on incident particle energy in sub-100nm CMOS technologies and it has been associated with an observed saturation of DSET rates with dimensional scaling.

4.2.4.3 SOI

Silicon-on-insulator (SOI) CMOS technologies have found favor for both extreme and benign environment applications. From a single-event perspective, the constrained collection volumes and inter-device isolation offered by SOI are appealing. Because the active silicon volume surrounding each MOS device is geometrically constrained, the volumetric extent of radiation ionization is limited – SOI devices typically display significantly reduced single-event-induced charge collection over bulk CMOS counterparts. In addition, because of the inter-device isolation provided by the bounding insulating regions, charge sharing and well collapse might be considered ‘problem solved.’

However, as is the case in so many engineering situations, the solution to one problem introduces others. As a design platform, SOI is not immune to single event phenomena. SOI CMOS devices suffer from parasitic bipolar conduction when the body region bias is modulated – precisely the result of a single-event particle ionizing charge in a highly constrained region. Source-drain bipolar conduction induces voltage modulation on the device terminals in the same way as direct charge collection – contributing to single event transients and upsets. Floating body SOI devices are acutely prone to this type of single event effect because there exists no direct conduction path to remove the excess body charge; however, even body-tied SOI devices are not immune because of the dynamic delay in charge conduction through the body tie path.
Likewise, while the inter-device isolation offered by SOI might seem a final solution to charge sharing effects, modern SOI technology platforms continue to display charge sharing and multiple-bit upsets. The culprit is dimensional scaling and the region of influence described in Section 4.2.4.1. In some cases several devices may share a single isolation region. In other cases, tightly-packed devices may share the same ion ‘region of influence’ (this is particularly relevant when the ion impinges at an oblique angle of incidence). In both cases, charge sharing occurs.

CMOS SOI controls single event charge collection and charge sharing, but the platform is not a panacea for extreme environment design. Moore’s law scaling, in fact, continues to erode the inherent radiation tolerance advantage of SOI over bulk CMOS as a radiation-hardened platform.

4.2.4.4 Design Implications of Dimensional Scaling

Feature size and critical feature pitch have both decreased with Moore’s Law scaling of technologies (dimensional scaling), as shown in Fig. 4.2.17. As a result, Si CMOS platforms exhibit several sensitivities to the environment:

Inter-device and intra-device leakage due to exposure to total ionizing dose is a significant issue for Si CMOS platforms. In particular, the leakage associated with isolation regions.

Scaling of inter-device pitch has exacerbated the problems associated with single event charge sharing and well collapse. Particular attention must be directed to layout topology, especially if local redundancy, such as TMR or DMR, is essential to successful operation of the platform in the environment.
Radiation-aware layout remains the best mitigation approach to combat dimensional scaling in modern Si CMOS platforms. In fact, layout techniques can be exploited to control sensitivity to ‘region of influence’ effects, as discussed in the next section.
4.2.5 Exploiting the Unique Attributes of the Modern CMOS Design Platform for Extreme Environments [56-60]

One of the most intriguing aspects of sub-100nm CMOS technologies as a platform for extreme environment design is the complex interaction of sometimes-conflicting features that emerge from Moore’s law scaling and affect transient radiation response. For example, dimensional scaling leads to smaller single event charge collection volumes with each technology node – a development that would seem beneficial to soft error resiliency; additionally, the cross-sectional area presented to the environment is reduced. However, concomitant with shrinking collection volumes is the reduction of transistor current drive, capacitance, and in turn, switching energy. In most cases, the latter trumps the former, leading to the increase in sensitivity of CMOS technologies to single event effects with scaling as discussed in the previous sections of this chapter.

Most hardening strategies associated with energy scaling adopt a philosophy of (1) charge control through process modifications such as isolation regions, buried layers, or charge sinks, or (2) photocurrent management through charge resupply, active feedback, well/substrate taps, or temporal filtering, or (3) event management through informational redundancy. There is very little that modern Si CMOS platform scaling offers in aid to these endeavors, other than sheer increase in available functionality and circuit complexity in which a designer may implement the strategies listed above.

However, the situation is different with dimensional scaling. As we have seen, this trend is often detrimental to single event resiliency, leading to the failure of classical RHBD, especially if unanticipated in a design. Yet, there exist methods to exploit region of influence
effects in scaled technologies to benefit radiation tolerance – that is, lemonade from lemons.

While typical soft error mitigation techniques for dimensionally-scaled platforms adopt a philosophy of physical separation to eliminate charge sharing and/or multi-bit upsets; there are recent examples of a contrarian approach – the exploitation of dimensional scaling and charge sharing in modern Si CMOS platforms – for soft error resiliency. These techniques employ a strategy of (1) maximizing charge sharing among critical nodes in order to induce digital-logic pulse quenching (to reduce DSET pulse widths) and thereby enhance natural electrical/temporal masking or (2) maximizing charge sharing in differential-signal analog circuitry (to transform a single event signal to a common-mode signal) and thereby enhance natural common-mode rejection. For further information, the reader is directed to the bibliography of this section.
4.2.6 Conclusions [61]

Working with Si CMOS design platforms in extreme radiation environments requires attention to many issues and relevant considerations:

- Sensitivity to front-gate threshold voltage shifts due to total ionizing dose has been mitigated by technology scaling and does not pose a serious threat to modern CMOS platforms.

- Subthreshold leakage current, both inter-device and intra-device, due to TID-induced trapped charge in isolation insulators, such as trench or buried oxides, is a significant issue in Si CMOS platforms.

- Back-gate threshold voltage shift coupling to the front-gate channel in very thin device structures, such as SOI or multi-gate transistors (MUGFETs), is an emerging issue in Si CMOS platforms.

- Switching energy and noise margin levels in modern CMOS platforms impose $Q_{\text{crit}}$ values below 1fC. With each technology generation there is an expanding spectrum of particles and particle energies capable of causing informational upset. Galactic heavy ions, solar protons, trapped protons, alphas, neutrons, and nuclear reaction products from back-end-of-line material stacks must all be considered.

- Switching energy levels below 1fJ implies that circuits are extremely sensitive to single events. Soft errors are not limited to memory elements – low switching energy enables DSETs and ASETs to appear and propagate freely in logic, analog, and mixed-signal circuitry, and are capable of causing operational failure. Inherent electrical and temporal maskings diminish as switching energy levels decrease.
• Dimensional scaling of CMOS platforms has generated rival response mechanisms to transient radiation. The projected surface area of cells or circuit nodes presented to the environment has been reduced with each technology node – per-bit error cross-sections are controlled. However, the relative ‘region of influence’ expands with shrinking feature pitch – multi-bit and multi-node effects proliferate. Region of influence effects are critically important in modern CMOS platforms.

• Of the ‘region of influence’ effects, charge sharing and parasitic bipolar modulation are extremely problematic in modern CMOS platforms. Internal and local redundancy mitigation methods, such as DICE or local TMR, are suspect unless singular care is exercised through radiation-aware layout of the circuit topology.

• Well bias collapse can contribute to multi-bit and multi-cell charge collection from a single event on dimensional scales much larger than the technology feature size. Radiation effects considerations in the placement of well and substrate taps are essential for extreme environment design.

• The phenomenon of pulse quenching and differential charge cancellation can be exploited for soft error and single event transient hardening with increasing effectiveness from dimensional scaling.
4.2.7 References

Comprehensive discussions of the role of radiation in silicon CMOS are contained in the seminal works listed below. These publications contain overview material, review papers, and short courses on the topics of this chapter, and provide an excellent resource to the interested reader.


4.2.1 Typical I-V currents for an NMOS devices exposed to total ionizing dose radiation. These data are representative of a 90nm bulk technology over dose levels of 0 to 10 MRad. Annotations show the three primary circuit response mechanisms.

4.2.2 Primary inter-device and intra-device CMOS leakage paths induced by total ionizing dose. Figure courtesy Prof. Hugh Barnaby, Arizona State University; used by permission.

4.2.3 Representation of a CMOS inverter under the influence of a single event strike near the NMOS device drain.

4.2.4 CMOS inverter characteristic rise time, $\tau_R$, normalized to the intrinsic time constant of Eq. 4.2.1 for $\beta=1$ (NMOS and PMOS devices sized for equal drive). Because of symmetry, the fall time, $\tau_F$, can be determined from these curves by interchanging $\alpha_n$ and $\alpha_p$ [45].

4.2.5 Single event transient charge collection current at the drain of a sub-100nm NMOS device configured as the pull-down device in a CMOS inverter. The curves represent ion linear energy transfer (LET) values of 1 to 40 MeV·cm$^2$/mg. The plateau, characteristic of sub-100nm CMOS circuitry, is caused by the collapse of the device depletion region and the balance of single event current and resupply current [46].

4.2.6 DSET pulse propagation in a string of CMOS inverters showing attenuation (filtering) and eventual electrical masking of the transient [45].

4.2.7 Gate-level diagram of a 4-bit ALU showing logical masking of single event transient propagation at three points within the structure [31].

4.2.8 The DSET window of vulnerability (WOV). Temporal masking occurs for DSET signals outside the WOV limit [34]. Figure courtesy Dave Mavis, MicroRDC; used by permission.

4.2.9 Complex circuitry, such as mixed-signal topologies, present time-varying sensitivity/masking of DSET transients. Here, results for a pipelined analog-to-digital converter show the temporal vulnerability curve, annotated by the subcircuit responsible for errors, against one complete conversion cycle of the ADC. Figure courtesy Jeff Kauppila, Vanderbilt Institute for Space and Defense Electronics; used by permission.

4.2.10 Typical ASET signatures seen in analog/mixed-signal CMOS platforms – positive, negative, and bipolar voltage transients. Figure courtesy Ray Blaine, Vanderbilt University; used by permission.

4.2.11 Full-width, half-maximum (FWHM) pulse width versus peak voltage magnitude scatter plot for an analog/mixed-signal circuit. Each data point represents an output voltage transient given a temporally and spatially random single event strike within the interior of the circuit topology. System-imposed limits on acceptable magnitude-width ASET characteristics (box...
overlay on the figure) allow ‘errors’ to be defined and confidence limits to be assessed [38].

4.2.12 Switching energy scaling of the Si CMOS platform as predicted by the ITRS roadmap for semiconductors. Vulnerability to single events is intimately tied to the declining switching energy of modern memory circuits through the critical charge for upset [61].

4.2.13 Representation of the single event charge sharing mechanism. The curves show the single-event-induced current on the active (struck) and passive (proximal) nodes for a 90nm bulk CMOS technology. The ion in this example: normal incidence on the active node drain with an LET of 60 MeV-cm²/mg. While the short-time profiles are dissimilar due to drift-assisted charge collection at the struck node only, the long-time profiles are similar for both nodes as diffusion-assisted charge collection from deep in the substrate reach both nodes simultaneously [50]. Simulations courtesy Dennis Ball, Vanderbilt Institute for Space and Defense Electronics; used by permission.

4.2.14 Representation of the single event parasitic bipolar mechanism [49].

4.2.15 Representation of the single event well collapse effect. In this 90nm technology, the n-well is debiased over a length far exceeding a single device, creating the conditions for parasitic bipolar action across many devices [52].

4.2.16 Schematic description of the single event pulse quenching mechanism, as described in the text [54]. The top diagram shows conventional single-node charge collection with no pulse quenching. The bottom diagram shows a single event transient pulse (DSET) that is quenched by the delayed arrival of diffusive collected charge at the drain of P₂.

4.2.17 Scaling of CMOS physical feature size (gate length) and density (half pitch) based on the ITRS roadmap for semiconductors. In terms relative to the ‘charge cloud’ generated by single event ionization, many devices fall under the ‘influence’ of a single ion at sub-100nm technology nodes [61].
Figures for Massengill Chapter 4.2

![Graph showing I-V curves for increasing total ionizing dose level with markings for subthreshold leakage shift, transconductance shift, and threshold voltage shift.](image)

Fig. 4.2.1 (above)
CMOS inverters

1. NMOS Drain-to-Source
2. NMOS D/S to NMOS S/D
3. NMOS D/S to NWELL
4. NWELL to NWELL (assume separate bias)

Fig. 4.2.2 (above)
Fig. 4.2.3 (above)
Fig. 4.2.4 (above)
Fig. 4.2.5 (above)
Fig. 4.2.6 (above)
Fig. 4.2.7 (above)
Fig. 4.2.8 (above)
Fig. 4.2.9 (above)
Fig. 4.2.10 (above)
Fig. 4.2.11 (above)
Fig. 4.2.12 (above)
Fig. 4.2.13 (above)
Fig. 4.2.14 (above)
Fig. 4.2.15 (above)
Fig. 4.2.16 (above)
Fig. 4.2.17 (above)