

The Pennsylvania State University
The Graduate School
Department of Computer Science and Engineering

SOFT ERRORS: MODELING AND INTERACTIONS WITH
POWER OPTIMIZATIONS

A Thesis in
Computer Science and Engineering
by

Vijay S. R Degalahal

© 2005 Vijay S. R Degalahal

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2005

The thesis of Vijay Degalahal was reviewed and approved* by the following:

Vijaykrishnan Narayanan
Associate Professor of Computer Science and Engineering
Thesis Adviser
Chair of Committee

Mary J. Irwin
Professor of Computer Science and Engineering
A. Robert Noll Chair of Engineering

Kenan Unlu
Professor of Nuclear Engineering
Associate Director of the Radiation Science and Engineering Center

Yuan Xie
Assistant Professor of Computer Science and Engineering

Raj Acharya
Professor of Professor of Computer Science and Engineering
Head of the Department of Computer Science and Engineering

*The signatures are on file in the Graduate School

Abstract

Soft errors are radiation induced ionization events that cause transient errors in circuits. Thus, the circuit always recovers from such errors as they do not damage the underlying semiconductor material. However, computation from the erroneous state of the circuit may corrupt the data, which is hard to correct or recover. In nanometer technologies, the reduced nodal capacitances and supply voltages coupled with more dense and larger chips are increasing the rate of soft errors, thus making them an important design constraint. This thesis models the phenomena of soft errors at the device level. This model is then used to create a transient current library which is used for circuit level soft error estimation. The library contains the transient current response to various factors such as ion energies, operating voltage, substrate bias, and angle and location of impact.

Excessive power consumption by CMOS devices is a major design limiter for current and next generation devices. As designers aggressively address the power problem, they need to be aware of the impact of power optimizations on the soft error rates(SER). Hence, this thesis also examines the influence of some popular power optimization schemes on SER. Finally, this thesis explores the effect on SER due to supply voltage scaling, one of the popular power optimization scheme, with accelerated neutron testing on a commercial memory chip. In summary, this thesis contributes towards understanding soft errors and their interactions with power optimizations by contributing in three significant directions namely modeling soft errors, understanding the implication

of power optimization schemes on soft errors, and studying the SER of SRAMs using accelerated neutron tests.

Table of Contents

List of Tables	viii
List of Figures	ix
Acknowledgments	xi
Chapter 1. Overview	1
1.1 Introduction	1
Chapter 2. Single event upset characterization for	
circuit error estimation	6
2.1 Introduction	6
2.2 Related work	8
2.3 SEAT-DA Tool	9
2.3.1 Charge deposition by neutron induced soft errors	10
2.3.2 Charge collection	12
2.4 Generating current pulse using SEAT-DA	15
2.4.1 Effect of neutron energy on the transient current pulse	16
2.4.2 Effect of impact location on the transient current pulse	17
2.4.3 Effect of technology	20
2.4.4 Effect of operation conditions	23
2.4.5 Current to Voltage Transformation	26

2.5	Case Study: Designing time redundant system	28
2.5.1	Time Redundancy	28
2.6	Conclusion	30
Chapter 3.	Influence of power optimizations on soft errors	33
3.1	Introduction	33
3.2	Soft Errors: Background And Related Work	34
3.2.1	Soft error mitigation schemes	37
3.3	Power and Soft Errors	37
3.3.1	Impact of supply voltage scaling on soft error rate	38
3.3.2	Impact of high threshold voltage devices on SER	39
3.3.2.1	Charge creation under high threshold voltages	40
3.3.2.2	Logic attenuation due to high threshold voltage de- vice	40
3.4	Methodology for circuit level analysis of soft errors	42
3.5	Impact of supply voltage scaling on soft error rate	44
3.5.1	Clustered voltage design	45
3.5.2	Drowsy SRAM design	46
3.5.3	Gated-Gnd SRAM Design	48
3.5.4	Quasi Static SRAM Design	49
3.6	Impact of high threshold voltage devices on SER	51
3.6.0.1	Effect of V_t on SER of SRAM and Flip-flops	51
3.6.0.2	Effect of V_t on Combinational Logic	55

3.6.1	Effect of delay balance using high V_t devices on soft errors . .	56
3.7	Conclusion	57
Chapter 4.	Testing of neutron induced soft errors	60
4.1	Test Facility	61
4.2	Experimental Setup for Soft Error Rate Measurements	63
4.2.1	Devices Under Test	63
4.3	Analysis of the results	65
Chapter 5.	Conclusion	68
References	73

List of Tables

2.1	Process parameters	15
2.2	Evaluating Time Redundancy	29
3.1	$Q_{critical}$ for different SRAM designs	50
3.2	$Q_{critical}$ and leakage power of SRAM and ASRAM with different V_t . Nominal V_t was 0.22V	51
3.3	$Q_{critical}$ of different flip-flops. Nominal V_t was 0.22V	53
3.4	$Q_{critical}$ and leakage power of various designs with different V_t . A high V_t TGFF was used at the output of the logic chain	58

List of Figures

2.1	SEAT-DA Simulation Tool Flow	13
2.2	Funneling	13
2.3	The distributions of the different reaction products	18
2.4	Transient current generated by various reaction products at 10MeV . . .	18
2.5	Transient current generated by various reaction products at 100MeV . .	19
2.6	Effect of distance of impact on charge collection	21
2.7	Effect of location of impact on charge collection	21
2.8	Effect of angle of impact on charge collection	22
2.9	Charge collection for different technologies	23
2.10	Drift current due to the collected charge when operating under different voltage	24
2.11	Comparison between PMOS and NMOS	24
2.12	Change in drain voltage due to the transient current	26
2.13	Change in drain voltage due to the transient current for different load capacitance	27
2.14	Logic Errors and Time Redundancy	30
2.15	Multi-Cycle Upsets	31
3.1	Circuit level evaluation of soft errors in logic circuit	42
3.2	Clustered voltage design	44
3.3	$Q_{critical}$ vs the supply voltage for different level converters	46

	x
3.4 Different SRAM designs	47
3.5 $Q_{critical}$ Vs the supply voltage for drowsy cell	48
3.6 Flip-Flops evaluated for SER	52
3.7 Asymmetric SRAM: Optimized for 0	52
3.8 Increase in $Q_{critical}$ of different designs with respect to operating nominal V_t of 0.22V	55
3.9 $Q_{critical}$ of typical inverter chain with either high V_t or low V_t flip-flop at the output of the chain. Nominal V_t was 0.22V	58
3.10 Delay Balancing	59
3.11 Effect of delay balancing	59
4.1 Thermal neutron energy distribution at the exit of a horizontal neutron beam port at PSBR	62
4.2 Testing the SRAM memories at the outer beam opening	63
4.3 Effect of supply voltage on soft errors	65
4.4 Effect of neutron Flux on soft errors	67

Acknowledgments

Graduate school at Penn State has taught me several things, and I will indebt to several people for it. It will be impossible to acknowledge all but I will mention just a handful of people.

I will always be greatly indebt to my advisor, Prof. N. Vijaykrishnan for providing timely support and encouragement. His insights and suggestions always helped me. In addition to sharing the first name, I will always aspire to share his enthusiasm and spirit. I will always be specially grateful to Prof. Mary Jane Irwin. Prof. Irwin is a great mentor who exposed the nuisances of the research and academic life which completed my overall education. Professors. Kenan Unlu, Yuan Xie and Mahmut Kandemir were always very supportive and helpful. It was a pleasure working with Tim Tuan at Xilinx. Tim was a great mentor and friend who taught me many things. I will be indebt to Sacit Cetiner for helping with nuclear engineering parts of this research. I also thank Gigascale Systems Research Center, Department of Energy, and Department of Computer Science and Engineering at Penn State for funding me at various stages. Special thanks to the Dean of College of Engineering at Penn State, for supporting my graduate studies with fellowship for three years.

The highlight of these four years has been the company of the best and brightest students from all parts of globe. In particular I would like to thank Victor, Theo, Rajaram, Greg, Yuhfang, Aman, Suresh, Nachiket, Lin, Gulin, Soontea, Fei, Charles, Kevin, In-Chao, Chrys, Ismail, Priya and Jie. Their company at MDL made the place

special. I would specially acknowledge Samsi, Batra, Ankur, Avanti, Ananth, Krishna, Angshu, Lav, Rama, Aparna and Angela. They were my family at State College. Special thanks to Monish for tolerating my friendship for so many years.

PhD is the culmination of a long process. A process whose foundations were laid by my parents and my teachers. My parents, my brother and my vadina have always been my support. I dedicate this thesis to them. In addition, I also dedicate this thesis to Dey Uncle and Shubendu Bhaiya. Their efforts in teaching me the basics of science and maths has shaped my entire academic career.

Chapter 1

Overview

1.1 Introduction

As technology scales, reducing power consumption and improving reliability of the CMOS devices is important as well as challenging. Power consumption is considered to be a major design constraint towards building more powerful chips and systems. As devices shrink, lower capacitance enables smaller and faster chips. Further, supply voltage is scaled down commensurate to the device geometry in the smaller technologies. To maintain performance, the threshold voltage is scaled down along with the supply voltage. However, leakage currents increase exponentially with reduction in threshold voltage. Lower supply voltage should ideally reduce dynamic power consumption; however, the quest for more powerful designs has increased the number of transistors on a chip. The dense integration of the transistors along with increased leakage currents, makes power density an important concern in newer technologies. Consequently, various optimizations for reducing power consumption have been proposed. On the other hand, as lower supply voltages and smaller nodal capacitances reduce energy consumption, they increase soft errors rates(SER). Hence, this thesis will focus on modeling soft error and investigate the influence of power reduction techniques on soft errors.

Soft errors are caused by natural radiation or decaying radioactive impurities present in packaging. These sources can be further classified as alpha particles from packaging materials, high energy neutrons from cosmic radiations, and residue by-products from the absorption reaction of cosmic ray thermal neutron with Boron-10 [11]. In simple physical terms, the soft error phenomenon can be described as a process of creation of excess electron-hole pairs due to the energy imparted, directly or indirectly, by an external radiation event. In the presence of an electric field, the electron-hole pairs do not recombine, but generate a current affecting a nearby device. If the collected charge due to ionization exceeds a certain critical threshold value, a transient pulse is generated that may change the data stored at the node. This critical threshold value is called $Q_{critical}$, and is an accepted metric to express susceptibility of a given circuit to soft errors. Current techniques that perform SER estimation at the circuit or the gate-level netlists, model charge collection as a simple current pulse or a glitch in the system [21, 53, 54, 44]. However, accuracy of glitch-based analysis is limited by the availability of accurate glitch models. In the case of neutron induced soft errors, the soft error induced current pulse is dependent on the ions generated by n-Si interaction, the device structure, and other physical properties. Hence, a charge collection model is essential for accurate circuit or gate level estimation of SER.

Analyzing charge collected due to a radiation-induced ionization calls for an interdisciplinary approach involving both nuclear and device physics. The phenomenon has its origin in nuclear and device physics, but most solutions exist in either the circuit or architecture level of abstraction. There are numerous approaches at modeling the phenomenon from a nuclear physics perspective. A good understanding of nuclear

physics is needed to model the reactions and calculate the energy of the byproducts. In addition to nuclear modeling, proper modeling of device behavior is also essential for accurate modeling of the charge collection phenomena. Most existing approaches focus on either the nuclear modeling aspects [47, 48, 51] or the device physics aspects of the phenomena [17, 37]. In this thesis the phenomena of a soft error is modeled by including both nuclear and device-modeling aspects. The methodology results in a simple but accurate charge collection model. Based on this methodology, a tool is developed to study the charge collection of a neutron induced soft error. Chapter 2 presents this model.

Next Chapter, the impact of power optimization on soft errors is analyzed. This thesis will explore the influence of two commonly used power reduction techniques namely supply voltage scaling and increasing threshold voltages, on the soft error rates.

- *Impact of supply voltage scaling on soft error rate.*

Voltage scaling is a very common technique to reduce dynamic power consumption. Dynamic power of the circuit is proportional to the square of the supply voltage. Hence, supply voltage is decreased to reduce the power consumption of the circuit. To maximize the gains from this technique, it is a common practice to employ clustered voltage design. In clustered voltage design, parts of a circuit operate at a lower voltage. Voltage level convertors are used to move from one voltage cluster to another. To examine the effect of such a design technique, the soft error susceptibility of six level convertors was analyzed. $Q_{critical}$ of these level convertors, was found to be linearly dependent on the supply voltage.

Supply voltage scaling is also employed to reduce the leakage energy. It is a common practice to reduce the supply voltage of the circuit when the circuit is not active and

the overheads do not facilitate turning off the supply. For example, in caches, when a cache line is not in use, the supply to the cache line can be reduced while still ensuring that the line retains the values [25]. The cache line is powered up before it is accessed. A custom designed cell, designed to operate at a voltage of 1V, could retain the values when the supply was reduced to 300mV. At 300mV the leakage was reduced by 70% but $Q_{critical}$ was also reduced by 65% [21]. Based on the above results it can be seen that even though voltage scaling reduces the dynamic and static power, there is a corresponding loss of immunity to soft errors.

- *Impact of the threshold voltage scaling on soft error rate.*

Higher threshold voltage devices are employed to reduce leakage. Various circuits were studied to quantify the effect of threshold voltage on soft errors. For all the circuits, an increase in the threshold voltage reduced the leakage energy drastically. On examining combinational logic circuits like NANDs and inverters, it was found that their $Q_{critical}$ reduces with higher V_t . In contrast, for circuits like transmission gate based flip-flops(TGFF) and asymmetric SRAMs[8], the $Q_{critical}$ increased with higher V_t . Thus, these circuits were less susceptible to soft errors. There are two distinct reasons that cause this anomaly. First, due to the physical properties of high V_t silicon, it requires higher energy to create electron-hole pairs in the substrate and hence high V_t device are more immune to soft errors. Second, higher V_t increases the gain and delay of circuits. This affects attenuation of the transient pulse. Hence, in the case of combinational circuits the use of high V_t devices increases the soft error susceptibility. Whereas in the case of TGFF and ASRAM, high V_t made the circuits

more robust. On examining the implications of these results in the context of *delay balancing* using dual V_t design libraries, use of high V_t flip flops with high V_t logic chains is recommend.

Chapter 4 presents the test setup for experimental evaluation of soft error rates in SRAM based memories. To verify the various observations deduced from the simulation studies, soft error testing infrastructure for neutron induced soft errors was built at Breazeale Nuclear Reactor(BNR) Facility. The experimental setup consists of a custom memory board interfaced to a computer through a GPIB card. The reactor has a maximum energy rating of 1MW. The beam has an intensity of 10^7 particles/s. The energy is distributed across a wide cross section. Tests from two off-the-shelf chips indicate that reducing the voltage increases the soft error rates. The chip works properly at these voltage levels without any errors before and after exposing to the neutron beam. Current tests clearly show the influence of voltage on soft errors.

Chapter 2

Single event upset characterization for circuit error estimation

2.1 Introduction

As the technology scales, transistors are getting smaller and faster, enabling smaller pipelines, faster clocks and dense memory structures. Smaller transistors have lower capacitance, and lower operating voltage, thus making them susceptible to soft errors. Soft errors are radiation-induced errors that cause a circuit error while the circuit itself is not damaged. Soft errors can occur at both logic and memory circuits. Recently, soft errors have been identified as a dominant source of failure in commercial designs [54, 13, 10, 4]. Soft errors are responsible for various failures, ranging from errors in memories used in large servers and aircrafts, to failures in implantable medical devices like cardiac defibrillators [15, 23].

Alpha particles, high-energy cosmic ray particles and neutron induced Boron- ^{10}B fission are the most significant sources of soft errors. Of these, in the absence ^{10}B due to elimination of Boro-Phosphorus-Silicate-Glass(BPSG), soft errors caused by high-energy cosmic ray induced neutrons are most dominant [10]. Cosmic rays of galactic origins interact with earth's atmosphere to produce a cascade of energetic particles. These particles include neutrons, protons, electrons, muons, pions, and gamma rays; however at sea level the flux is dominated by neutrons [55]. These energetic particles

can interact with silicon either directly or indirectly producing electron-hole pairs. Under the right conditions, these electron-hole pairs get collected, and create a pulse, which if significantly large, can cause error. This form of an upset is also called single event upset (SEU). As it is possible to recover from these errors, these are also called soft errors. These errors are distinguished from hard errors that are caused when the energetic ions cause permanent damage to the circuit, and hence the circuit is not recoverable.

The phenomenon of soft errors has its origin in nuclear and device physics, but most solutions exist in either the circuit or architecture level of abstraction. Hence, there is a need to create a soft error estimation toolset at these different hierarchies. To address this issue, Soft Error Analysis Toolset(SEAT) toolset is being built. SEAT, is a hierarchical toolflow, that models soft error across device, circuit and architectural levels. In this chapter SEAT-DA(Device Analysis) is presented. SEAT-DA is the first level in the SEAT toolset. As it is important to create a level of abstraction at each hierarchy, SEAT-DA abstracts the soft error impact as a transient pulse waveform. For a circuit designer, it is important to abstract the soft error impact as a transient current pulse. SEAT-DA models soft error using nuclear and device physics tools with an aim to create a transient current waveform library that captures different process and operating conditions that impact the soft error rate(SER).

Current techniques that perform SER estimation at circuit or gate level model charge collection as a simple current pulse or a glitch [21, 53, 54, 44]. This radiation induced current pulse is abstracted in Equation 2.1 as follows [44]:

$$I(t) \propto \frac{Q}{T} * \sqrt{\frac{t}{T}} * \exp\frac{-t}{T} \quad (2.1)$$

where Q refers to the amount of charge collected due to the particle strike. The parameter T is the time constant for the charge collection process and is a property of the CMOS process used. In the case of neutron induced soft errors, calculating Q and T is not trivial, as it is dependent on the ions generated by the n-Si interaction, the device structure, and other physical properties. Hence, SEAT-DA tool can be used to create a library of current pulses, and it compliments these methods by providing a way to accurately estimate Q and T . This method is akin to cell or device level characterization used to estimate various circuit and device level properties like delay, power and capacitance.

The importance of the SEAT-DA methodology is shown by examining a simple time-redundant system. It is found that in multiple giga-hertz processors regimes, there is a likely hood of an single neutron strike causing multi-cycle upsets. Its shown that an accurate modeling of the current glitch caused by the incident neutron strike is critical for optimal design of time-redundancy solutions.

The chapter is presented as following: Section 2.2 presents the related work. Section 2.3 describes the SEAT-DA tool. Section 2.4 explores the effect of different conditions on the transient current generated by a SEU. Section 2.5 presents a simple analysis of a time redundant circuit. Section 2.6 presents the conclusions from our study.

2.2 Related work

Soft error have been modeled at different levels of hierarchy. While [53, 54, 44] model soft errors at circuit and gate level, [56, 17, 37, 47, 45] model them at the device level. In [53, 54, 44], a combination of probabilistic and analytical approaches is used

to calculate soft error rate. These models assume the availability of accurate device level information. Burst Generation Rate (BGR) model was proposed in [56] and refined in [47]. The BGR model proposes that if the energy collected in the sensitive volume exceeds a critical charge, an error is said to have occurred. The energy of the ions is calculated by using nuclear codes. Such an approach is also used in [52, 20]. In contrast [17, 37] uses 3D device simulation to study and model SEUs in simple devices and structures like SRAMs. In [45], NUSPA (Monte Carlo based nuclear codes) are used to model the reaction products from the n-Si interactions, and FEM based methods are used to model the device behavior. While these do model accurate nuclear and device physics information, they do not provide a easy abstraction for the circuit level error estimation. Hence, a comprehensive methodology is necessary. Such a methodology should accurately model both the nuclear reaction and charge collection, and abstract the information in form useful for the circuit designer. In this work, such a methodology is presented. In the methodology, MCNP codes are used for modeling n-Si interaction, TRIM is used for estimating the charge deposition, and a commercial 3D device simulator (Synopsys Davinci) is used for capturing the device effects on charge collection. In addition, this methodology is technology independent (though the results based on a 130nm process), and extensive evaluation of various processes, technology and operating parameters is provided.

2.3 SEAT-DA Tool

Soft error induced transient pulse generation is dependent on exact charge deposited by the neutron-Si interaction and its subsequent collection. SEAT-DA is a

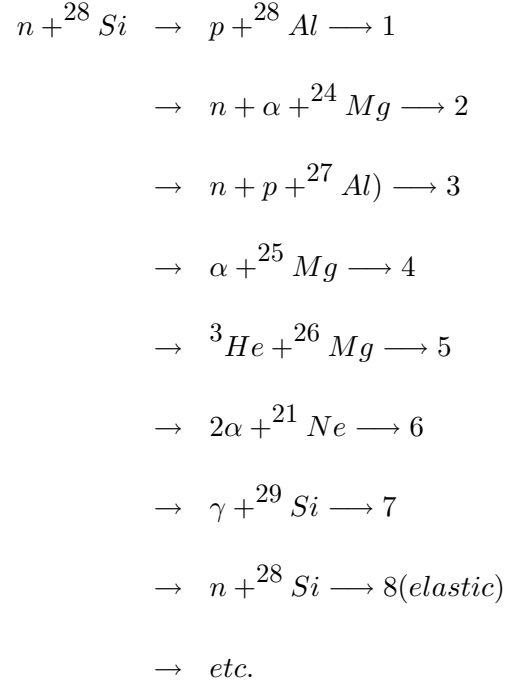
toolflow built on top of three different tools as shown in Figure 2.1. It models both charge deposition and charge collection as described in the following subsections.

2.3.1 Charge deposition by neutron induced soft errors

To study n-Si interactions, the Monte Carlo N-Particle (MCNP) toolset is used [2]. Input to MCNP includes a model of silicon substrate and the description of the neutron flux. MCNP can be made to run with the different reaction codes and neutron data files to model various reactions [2]. This feature is particularly useful as the neutron flux is dependent on the location and altitude, one may setup MCNP with the exact distribution of neutron flux at a given place to calculate the exact n-Si interaction. Customized scripts are created that parse the MCNP output to identify the different reactions and their outputs. MCNP is used for studying neutron, photon, electron, or coupled neutron/photon/electron transport. This tool has been traditionally used in nuclear engineering applications such as, reactor designs, radionuclide based imaging, and others.

Neutron-Si reactions can be classified into two main groups: elastic and inelastic. MCNP can model both elastic and inelastic scattering. Elastic scattering, due to the low mass of the neutrons, does not produce large ionizations but is the dominant form of the collision. In contrast, inelastic reactions occur when the neutron enters the nucleus and the unstable nucleus disintegrates to smaller particles. Many reactions are possible and various particles may be emitted (Please see Equation 2.2; these will be referred to by the numbers given below). Since these byproducts are heavier than the original

neutrons, they deposit more charge as they travel in silicon. The MCNP outputs are parsed for the both for elastic and inelastic reactions



(2.2)

Once the different reactions products are obtained, Transport of Ions in Matter (TRIM) is used to calculate the charge deposited by these ions. Interfacing MCNP and TRIM together enables an accurate analysis of the charge creation. TRIM is used to calculate the stopping power of ions [3]. TRIM identifies the range and the charge these ions are capable of depositing. Once the ion distribution resulting from a particle strike is known, its range and charge generation rate is calculated using TRIM. This generation

rate is fed to a 3D device simulator to calculate the charge collected in a given region of the device.

2.3.2 Charge collection

After the reaction products of n-Si interactions deposit charge, this charge may either recombine or get collected on the device terminal to generate current. For modeling charge collection, Synopsys TCAD Davinci 3D device simulator is used.

Davinci uses the physical model and equation interface (PMEI) to perform simulations that incorporate user-defined physical models and equations. The input to the 3D simulator includes the device structure, device parameters and device level equations. The charge may be collected in the device terminals by either drift or diffusion processes.

When the ion track is sufficiently far from the space charge zone of the drain junction, the carriers generated in the track mainly move by diffusion. However, for charge collection the most sensitive regions are reverse-biased p/n junctions of the transistor. The high field present in a reverse-biased junction depletion region can collect the charge generated by the ion tracks through drift processes, leading to a transient current at the junction. An important phenomenon associated with the charge collection is called field funnel. Charge generated along the ion track can locally collapse the junction electric field due to the highly conductive nature of the charge track and separation of charge by the depletion region. Figure 2.2 shows the field in a device after the field has collapsed. The funneling effect can increase charge collection at the struck node by extending the junction electric field away from the junction and deep into the substrate such that

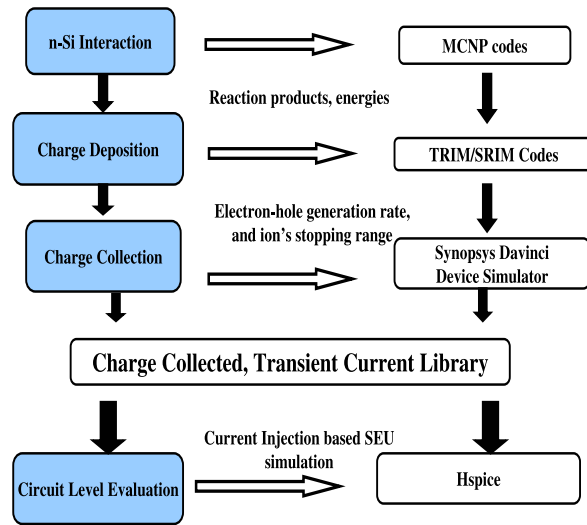


Fig. 2.1. SEAT-DA Simulation Tool Flow

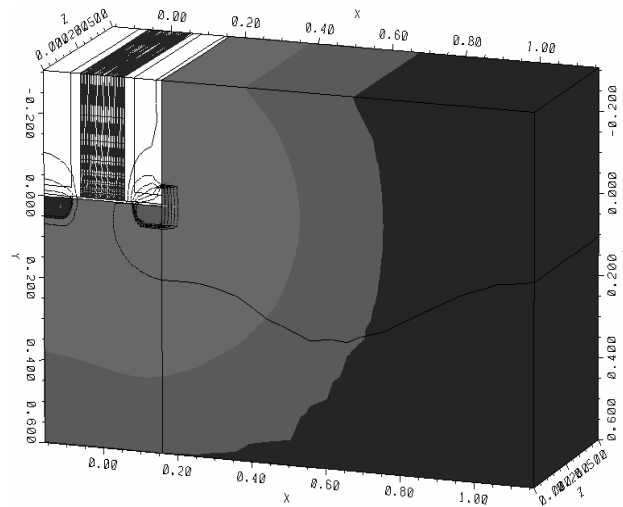


Fig. 2.2. Funneling

charge deposited some distance from the junction can be collected through efficient drift process.

In deep-sub-micron technology, another phenomenon termed as alpha-particle source-drain penetration effect (ALPEN) also contributes to the phenomenon of charge collection [23]. Due to ALPEN, if a particle strike passes through both the source and the drain at near-grazing incidence, a significant but short-lived source-drain conduction current that mimics the “on” state of the transistor, is generated. However, in sub-100nm devices, when electron-hole pairs are generated there is a high probability that such a generation spans a region greater than the gate length. Hence, the definition of ALPEN is expanded to include these effects. In addition, the processes of funneling and ALPEN will be referred to as drift processes.

The simulator was setup to use the physical models that include standard drift-diffusion laws and classical physical models. These models include: Carrier-carrier scattering mobility model (CCSMOB) to account for the large carrier concentrations present in the charge column. CCSMOB also includes effects of doping and temperature on mobility. Field-dependent mobility model (FLDMOB), to account for reverse biased junction, and high electric fields in the depletion region. Shockley-Read-Hall and Auger recombination models to account for recombination of the carriers. Band-gap-narrowing (BGN) model is used to model the pn-junction as a bipolar device. The device was loaded with lumped resistance and capacitance models to ensure realistic conditions.

The electron-hole pairs are introduced in the simulation as a charge column. As in [37], the charge column is assumed to have a Gaussian profile. The charge is generated over a period of about 6 picoseconds using a Gaussian waveform. The structure was setup

to solve time-dependent solution lasting up to 5ns. This is sufficient to resolve the drift and diffusion component of the charge collection process. However, the diffusion charge collection may continue for a longer period, but its contribution to total charge collection is negligible. The output from the 3D simulation is used to generate current profiles for the different particle strikes. The current is integrated over the time to calculate the charge collected by the soft error.

Hence, a typical transient current generated by a soft error has a high drift component, which lasts for a few picoseconds and after the collapsed field is re-established, the charge collections is predominantly due to diffusion. For glitch based circuit level analysis, its important to model both drift and diffusion component accurately as the drift process is responsible for the peak, and the diffusion process is responsible long tail of the fast-rising slow decaying current pulse.

2.4 Generating current pulse using SEAT-DA

Using SEAT-DA, the transient current information for various different conditions namely; energy, location and angle of the neutrons, operating voltages, effect of load capacitance, and substrate bias, was charaterized.

Table 2.1. Process parameters

	NMOS	PMOS
Epitaxial layer doping/ cm^3	3^{15}	3^{17}
Substrate doping/ cm^3	5^{17}	3^{19}
Channel junction depth/ μm	0.05	0.05
Channel peak S/D Doping/ cm^3	2^{20}	2^{20}
Power supply voltage/V	1	1

2.4.1 Effect of neutron energy on the transient current pulse

To investigate the effect of neutron energy on the n-Si interaction, the MCNP scripts were setup with different input neutrons energies and reaction modes. For each reaction and energy point, 5 million neutrons were simulated. A 130nm PMOS transistor was used as a test design to the device simulator. The details of the process of this transistor are given in Table 2.1.

Neutrons are highly penetrating. Figure 2.3 presents the distribution of different reactions at different energy. Referring to Figure 2.3, one finds that as the incident neutron energy increases, the number of byproducts also increases. Note that, at a low energy one or two reactions dominate, while at a higher energy more reactions start to occur. This particular inform is very useful as the current profile created by each of these reaction products differs, and in the library, the current waveforms from the most dominant reactions will be included.

For example, for neutrons under the energies of 20MeV, notice that reactions 1 and 2 (please refer Equation 2.2), dominate as they are about 60% and 40% of the total reactions. Hence, for doing current pulse based circuit level analysis, one will have to just use the pulse from the ions generated from these reactions namely; ^{28}Al , ^{25}Mg , proton and alpha particle. Figure 2.4, shows the transient current pulses of ions produced by these reactions. Also, as alpha particle and proton have comparatively lower sizes, the transient pulse generated by them is smaller in magnitude (Less than few hundreds of nAmps).

Similarly, at energy range greater than 100 MeV, it can be seen from Figure 2.3 that reactions 1, 2, and 3 contribute to more than 97% of reaction products. Hence, the current pulse information from only these ions is needed. Figure 2.5 presents the current waveforms at these energies. Note that, ^{15}N is only shown in both Figure 2.4 and 2.5 for comparison, and is not dominant in these energy ranges. Based on these figures, notice that typical current pulse height is about 1.8 milli-Amperes for the dominant ions, and the charge collection time is about 0.8ns. Another interesting point that can be observed from these figures is that the magnitude of the current pulse is dependent on the energy and the size of the ion. Larger ions create larger pulses, and collect more charge. The waveform of the lighter ions is more dependent on energy of the particle than that of heavier ion. Also note that the total charge collected is simply an integral of this current pulse.

Based on these examples, the library characterization method is optimized, by just including the current pulses from the dominant reaction products. In addition, since the terrestrial neutron flux is dependent on the latitude and altitude of the location, the current library for the neutron flux distribution at a given location can be characterized accordingly.

2.4.2 Effect of impact location on the transient current pulse

To examine the effect of impact and ion generation location, the charge column was induced in 5 different locations as shown in Figure 2.6(a). The locations were separated by $0.25\mu\text{m}$. The output current waveforms are shown in Figure 2.6(b). From Figure 2.6(b), one can see that as the point of impact moves away from the transistor

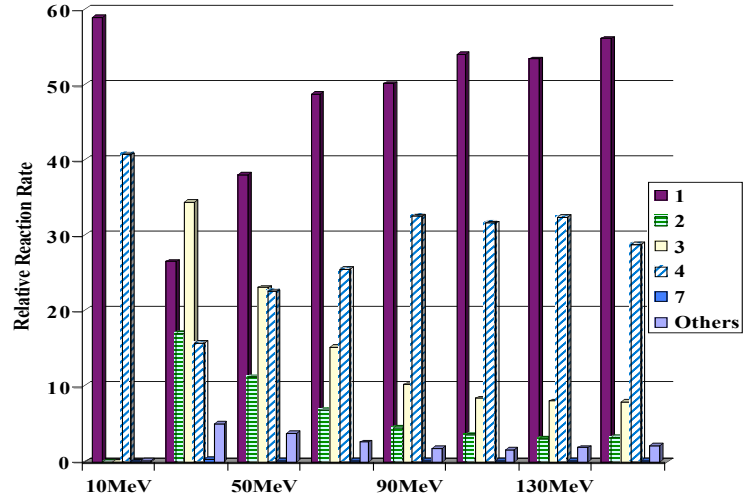


Fig. 2.3. The distributions of the different reaction products

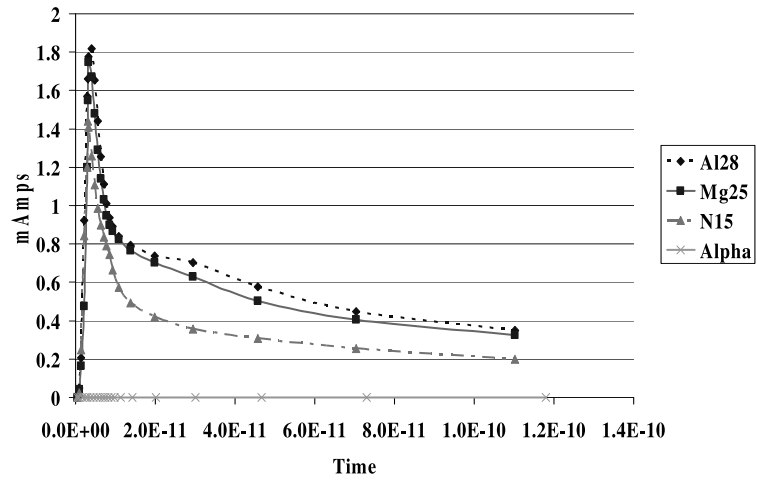


Fig. 2.4. Transient current generated by various reaction products at 10MeV

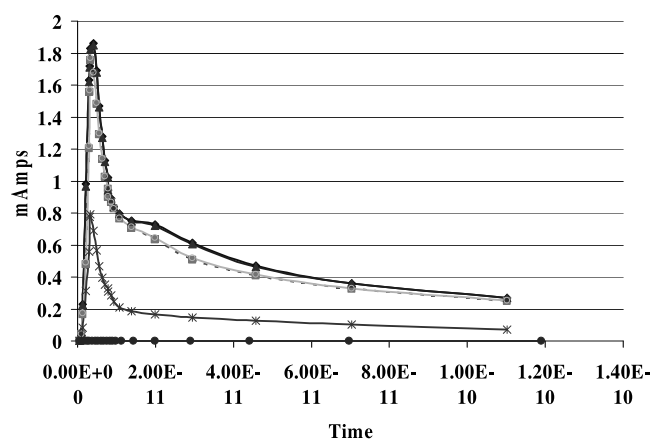


Fig. 2.5. Transient current generated by various reaction products at 100MeV

(refer to Figure 2.6(a)), the sharp peaks in the current waveforms reduce in magnitude. This indicates the reduction in contributions of drift processes like ALPEN and funneling. This was also evident from the noting the field contours in the simulator. At a distance of 0.5 microns, the charge collected was highest, as both drift and diffusion processes contribute to the charge collection. For distances greater than 0.5 microns the diffusion dominates the charge collection phenomenon. From the simulations, note that for diffusion dominated charge collection currents extend for more 4 ns, while the magnitude was only in the order of 10^{-4} amps. Hence, diffusion dominated current pulses have a long decay time resulting in long tail and very low peak value. Overall, based on the location of impact the peak magnitude of the current changed from few hundreds of nano-amperes for diffusion dominated strikes, to 1.4 milli-Amperes for a strike close to the drain terminal. However, the tail of these strikes was about 1.4ns.

Similarly the angle of impact of these ions was changed. The path traversed by all the different angles crossed the device either in the beginning or at the end of the track. This ensured that drift processes dominate the charge collection, and in turn, the impact of these angles on charge collection and the current waveform was negligible.

2.4.3 Effect of technology

To examine the effect of the technology scaling, devices were designed in 180nm and 250nm. To compare the results a similar charge column was introduced at the drain of a PMOS designed in each of the three technologies. The results are presented in Figure 2.9. Note that as the charge deposited across the 3 different technologies differs. As the larger device has larger cross section, the charge collected by the larger device

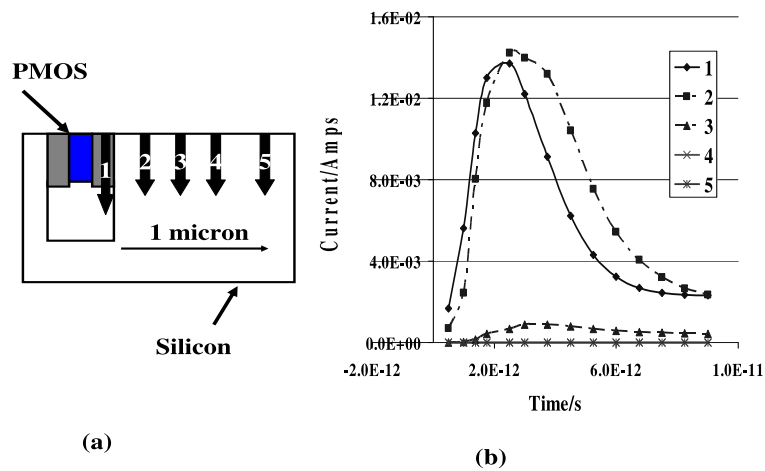


Fig. 2.6. Effect of distance of impact on charge collection

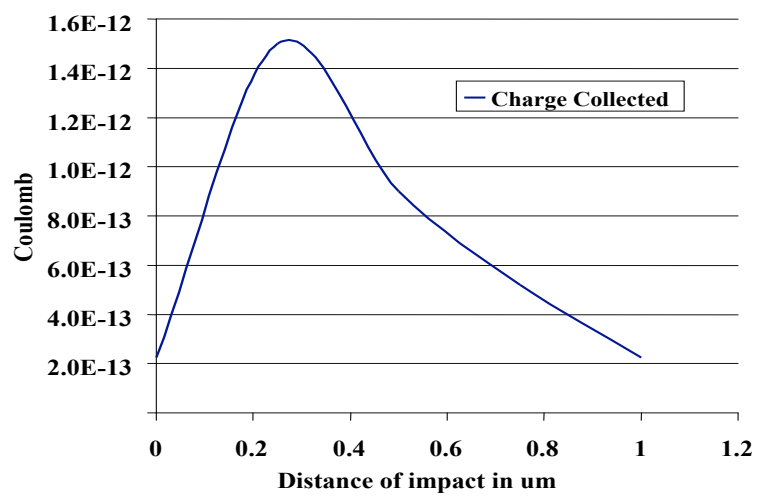


Fig. 2.7. Effect of location of impact on charge collection

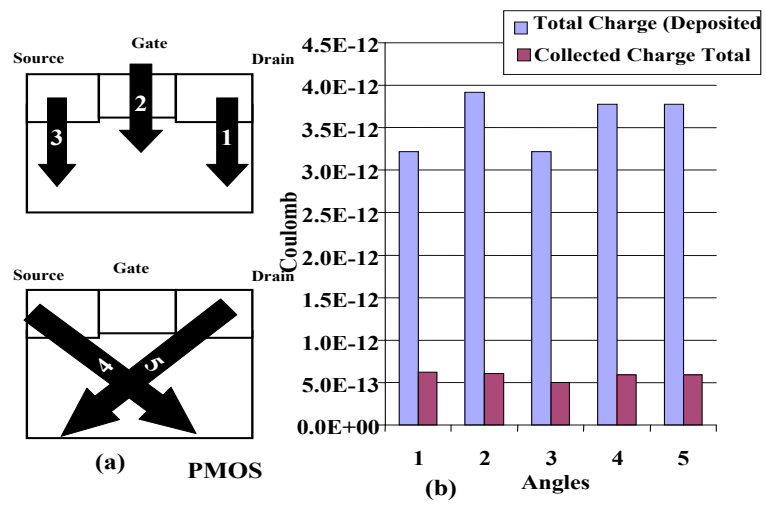


Fig. 2.8. Effect of angle of impact on charge collection

is greater than that of a smaller device. However, the charge collected by the node is higher for the smaller devices. Due to smaller device dimension, the charge collection due to drift processes is more efficient in smaller devices.

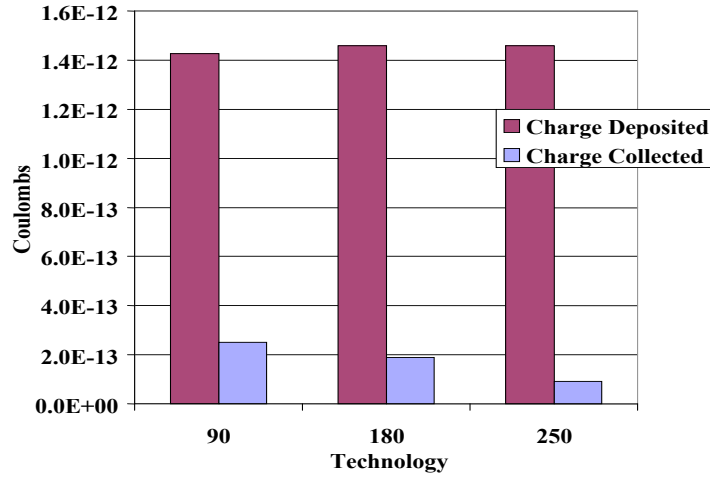


Fig. 2.9. Charge collection for different technologies

2.4.4 Effect of operation conditions

Next, SEAT-DA is used to characterize the transient current pulse response of operation the device at different operating voltages, substrate bias and sensitivity of PMOS in comparison to NMOS.

For evaluating the effect of different operating voltages, the device was operated between 0.5V to 2V. The resultant current transient that was generated differed with

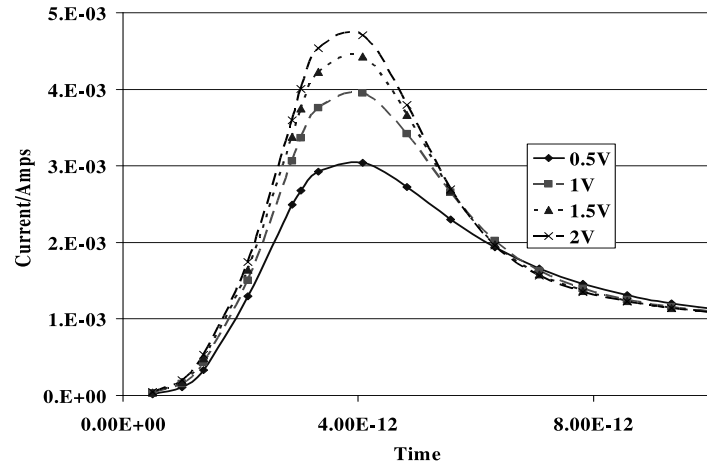


Fig. 2.10. Drift current due to the collected charge when operating under different voltage

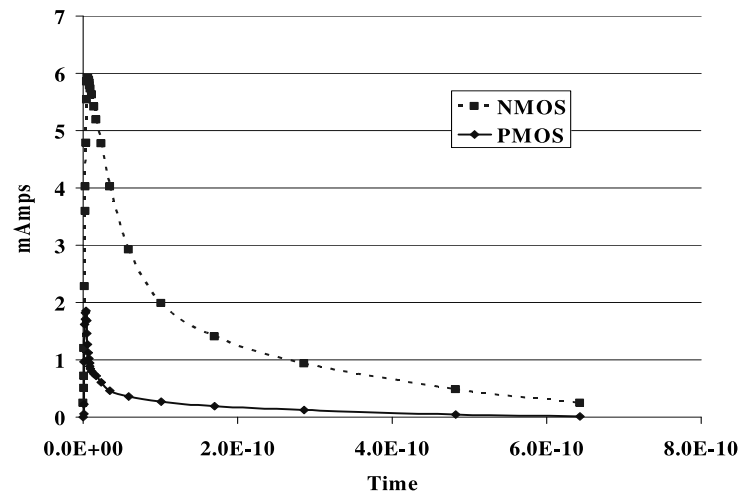


Fig. 2.11. Comparison between PMOS and NMOS

the operating voltage. Figure 2.10 shows the drift current due to the collected charge for the first 10ps. Based on the current profile we can see that, at lower voltages, the contribution of the drift current to the charge collection is low. As the fields are low the charge takes longer time to recombine, and hence diffusion current increases, thus contributing to similar charge collection. However, it is important to note that the transient current pulse has a lower peak magnitude value. At 2V, the pulse peaked to 4.7 milli-Amperes, while at 0.5V the pulse peak was only 3 milli-Amperers for a ^{26}Al ion. Hence, while estimating the soft error error susceptibility of circuits working at multiple operating voltages, a suitable (scaled) transient current pulse should be used.

Next, the effect of the substrate bias on charge collection was investigated. For the experiments, the substrate potential was increased by a maximum of -0.2V, for the device itself was operating at 1V. If the substrate was biased with negative or positive voltages, electrical fields in the device in the changes. This affects the charge collection. However, in the simulation, it notice that the funneling and the drift current dominate the charge collection. Hence, marginal increase in the substrate bias will not change the charge collection phenomenon, hence no significant difference in the transient current pulse shape was noticed. Similarly, no change in the transient current (and hence, in charge collection process) was noticed when the load capacitance was changed. But the load capacitance, changed the voltage output of the node. The effect of load capacitance will be further elaborated in the next subsection.

Lastly the difference in sensitivity between NMOS and PMOS was compared. NMOS was found to collect more charge and hence was more susceptible. This is attributed to the difference in the doping concentration of the n and p-type regions given

in Table 2.1. Figure 2.11 presents the results from this experiment. Notice that for the current pulse generated by a ^{28}Al at 100 MeV, the peak charge collected by NMOS is almost twice in magnitude than the one generated by the PMOS.

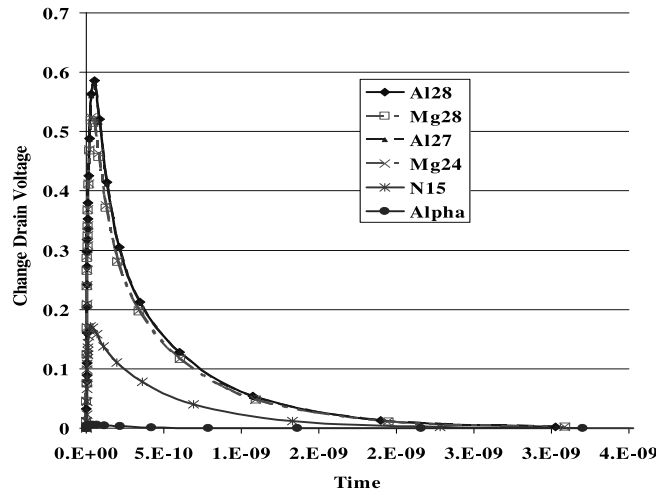


Fig. 2.12. Change in drain voltage due to the transient current

2.4.5 Current to Voltage Transformation

While the single event upset generates a transient current pulse, the operation of the circuit is actually effected by the voltage pulse that gets effect by the collected current. Figure 2.12 presents the voltage waveform for the transients presented in Figure 2.5. Notice that even though the current transients settle in about 1.4 ns, the voltage

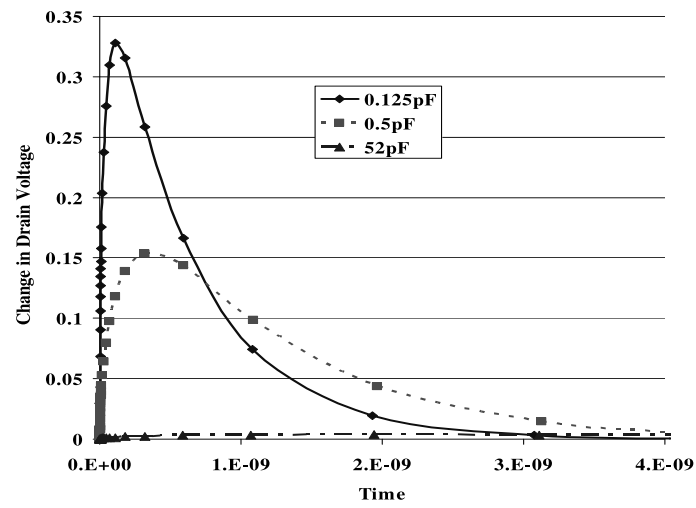


Fig. 2.13. Change in drain voltage due to the transient current for different load capacitance

transients last for longer time. Also it is interesting to check the effect of the load capacitance on the node voltage. Figure 2.13, presents the effect of load capacitance on the nodal voltage. As noted in the previous subsection, on changing the load capacitance no change in the charge collection process is observed. However, from Figure 2.13, notice that on increasing the nodal capacitance, the magnitude of the voltage transient reduces. Infact, for an increase in the capacitance by 5 times, the magnitude of the pulse reduces by half. For about 500 times increase in capacitance, the voltage transient is almost negligible.

2.5 Case Study: Designing time redundant system

Next, a simple study of a time redundant system is presented.

2.5.1 Time Redundancy

Soft errors in logic circuits are caused when the transient pulse generate by a neutron strike is latched by a latch or a register (as shown in the Figure 2.14(a)). These pulses are also some times refereed to as the single event transients (SETs). The circuit is made robust against such SETs by sampling data at two different time. The two sampled data are compared to check for errors. As the transients get attenuated in time, the probability of an error occurring at two different sampling time windows is very low if the window is larger than the typical pulse. Referring to Figure 2.14(b), if the value of the logic is sampled again after delay of δ , then the error transient may not be latched if width of the transient pulse is lesser than δ . If δ is less than the expected pulse width, than there is still a probability of soft error. If δ is too large than the pulse width, it

will just slow up the circuit and degrade the performance. Thus, the effectiveness of these kinds of circuits techniques is dependent on the ability to predict the transient pulse width. Using the SEAT-DA methodology, the transient pulse width can be measure very accurately and hence, help in designing optimal time redundant solutions.

For evaluating the time redundancy, the circuit presented in Figure 2.14(b) was designed. The simulations were simplified by assuming just a chain of inverters. The circuit was running at 1 GigaHertz. Based on the simulations of SEAT-DA, a current pulse was injected to mimic the soft error. The magnitude of the current pulse was varied from a very low value to mimic diffusion dominated current pulse. To a very high magnitude current pulse that was caused by a heavier ions like Al. Hence, notice that for a larger pulse the error pulse is latched on for multiple cycles. Table 2.2 presents the different pulse widths caused by different ions and their corresponding error in terms of cycles. Hence, for designing the circuit for time redundant solution, the δ delay in the clock and delay clock should be larger than the most dominant pulse. However, one may further try to optimize the solution. Based on Figure 2.3, note that their are only a few reactions possible at neutron energies less than 20MeV. Based on [55], neutrons of low

Table 2.2. Evaluating Time Redundancy

Pulse width ns	Pulse Height mamps	Typical Energy Range MeV	No. of cycles of errors
0.6	4	Above 150MeV, ^{26}Al	2
0.4	3.6	100MeV, ^{26}Al	2
0.2	2.4	10MeV, ^{26}Al	1

energy occur more frequently than neutrons of high energy. Hence, one can potentially design for the most common case of n-Si interaction.

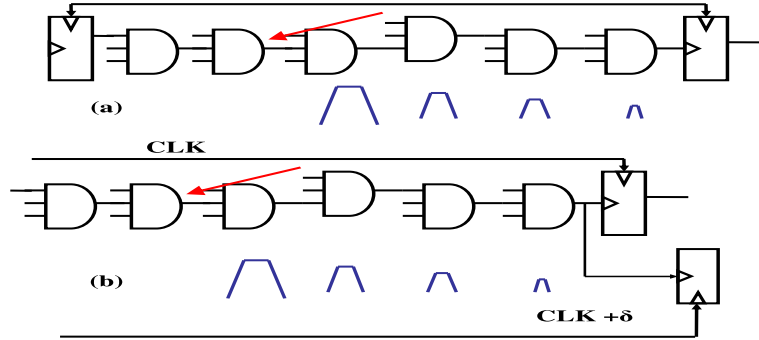


Fig. 2.14. Logic Errors and Time Redundancy

2.6 Conclusion

This chapter presents the SEAT-DA tool. SEAT-DA is used for characterizing the transient current pulse generated by a neutron induced soft errors. SEAT-DA is used to characterize and study the effect of various factors on the transient current pulse, and subsequently on soft errors. Based on n-Si interaction, it was found that the charge collected in the silicon depends on the reaction products and their energy. It was also found that the charge collection is dominated by drift processes near the device. For an impact at a distance the charge collection is through the diffusion process. It was also

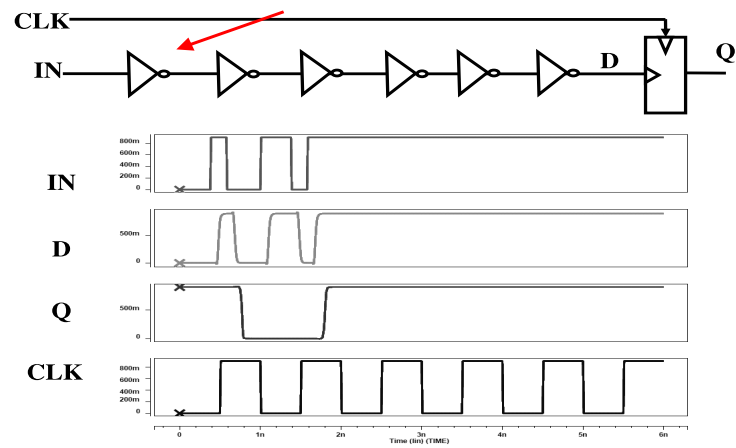


Fig. 2.15. Multi-Cycle Upsets

found that the charge collection is weakly dependent on voltage, substrate bias and angle.

SEAT-DA tool is very useful in modeling the soft error induced current pulse.

Chapter 3

Influence of power optimizations on soft errors

3.1 Introduction

As silicon industry enters the nanometer regime, it is facing new challenges on several fronts. In the past, aggressive technology scaling has improved performance, reduced power consumption, and helped the industry obey Moore's law. In the sub-130nm regime, supply voltage is also scaled down to reduce the power consumption. To compensate for the lower supply voltage, the threshold voltage of the device is also reduced. This increases the subthreshold leakage [14]. In addition, the ultra thin gate oxides increase the tunneling probability of the electrons, thus increasing the gate leakage. Furthermore, the dense integration of the transistors along with increased leakage currents makes power density an important concern in newer technologies. Hence power, by many, is regarded as the most significant road block in realizing the benefits of scaling for next generation. Consequently various optimizations for reducing power consumption have been proposed [50, 42]. This chapter will mainly examine the impact of higher threshold voltage devices and supply voltage scaling on soft error rates(SER).

The direct consequence of the lower supply voltage is lower signal to noise ratio (SNR). This results in increased susceptibility of circuits to noise sources like soft errors. In contrast, the effect of the higher V_t devices is not straight forward. As explained in the previous chapter, soft errors are transient circuit errors that are induced

by external radiations. Conventional ways of reducing the soft error rates include adding redundancy, increasing nodal capacitance and using error correcting codes. This work analyzes the effect of increasing the V_t of the device and supply voltage scaling on soft errors in standard memory elements like SRAMs and flip-flops and, on combinational circuits like inverters, nand gates and adders, which represent the most common CMOS logic styles. Such an analysis is very important because it helps in making intelligent design choices that reduce leakage power consumption and improve the reliability of the next generation circuits.

The chapter is organized as follows: Section 3.2 presents the background for soft errors, correcting schemes and related work. Section 3.3 also presents a relatively short back ground on power consumption in CMOS focusing on the impact of supply voltage scaling and high V_t devices on SER. Section 3.4 presents the experimental methodology that is used to examine soft errors on circuits. The following sections discusses the detailed results of the experimental analysis of SER on different circuits, and Section 2.6 presents the conclusions.

3.2 Soft Errors: Background And Related Work

When energy particles hit a silicon substrate the kinetic energy of the particle generates electron hole pairs as they pass through p-n junctions. Some of the deposited charge will recombine to form a very short duration current pulse which causes a soft error. In memory elements, these can cause bit flips, but in combinational circuits they can cause a temporary change in the output. In combinational logic such a pulse is

naturally attenuated, but if a transient pulse is latched, it corrupts the logic state of the circuit [12, 16].

According to [27], the soft error rate of memory element is expressed as Equation 3.1:

$$SER \propto N_{flux} * CS * \exp \frac{-Q_{critical}}{Q_s} \quad (3.1)$$

where

N_{flux} is the intensity of neutron flux,

CS is the area of cross section of the node,

Q_s is the charge collection efficiency,

$Q_{critical}$ is the charge that is stored at the node and hence is equal to

$VDD * C_{node}$, where VDD is the supply voltage and

C_{node} is the nodal capacitance.

Hence any reduction in supply voltage or nodal capacitance increases the soft error rate. The soft error rate is also proportional to the area of the node CS . Smaller devices have lesser area and hence are less susceptible for an upset. But lower $Q_{critical}$ coupled with higher density of the devices in larger dies ensures an increase in soft errors for each new generation [44, 43, 27].

In combinational circuits soft errors are dependent on many factors. A transient glitch due a radiation event at a node alters the execution by generating a computation

error only if the resultant glitch causes a change in the output which is then latched by a register. These factors derate the soft errors rates in logic. Hence in logic circuits, the

$$SER \propto N_{flux} * CS * Prob_G * Prob_P * Prob_L \quad (3.2)$$

where

$Prob_G$ is the probability that an transient pulse is generated for a particle strike at that node

$Prob_P$ is the probability that the generated transient pulse propagates through the logic network

$Prob_L$ is the probability that the transient pulse is latched at the output

For static CMOS logic, all the factors except $Prob_G$ are dependent on the circuit structure, inputs to the circuit, operating voltage and technology. Hence for studying the effect of frequency on SER, we find SER inversely proportional to clock period. Hence, shorter pipeline lengths, and higher frequencies are contributing to higher error rates in logics [43, 10]. In this chapter, we will not be elaborating on the effects of frequency scaling on SER.

3.2.1 Soft error mitigation schemes

The basic soft error mitigation techniques involve information redundancy, space redundancy and time redundancy. These techniques have been applied at different granularity. Recently, researchers have proposed techniques to make use of inherent hardware redundancies of multi-threaded and on-chip multiprocessor architectures in concurrent error detection [41, 40]. In memory structures, the information redundancy can be reduced by clever use of codes and scrubbing techniques. There are also many modifications one can do at circuit level to make the circuit robust. For example, circuits can be designed with higher capacitance or designed using partial or complete redundancy measures at the different nodes [6]. Since the logic state of a circuit at a node is stored as the charge stored at that node ($Q = CV$), one can increase the nodal capacitance of the gate and thus make the node more robust [31]. The $Q_{critical}$ at a node will decrease as voltage or nodal capacitance decreases. The nodal capacitance is strongly dependent on the layout. Some designs offer better immunity against SER than others. All these techniques cost a lot in terms of complexity, area, power and performance. This chapter's focus is on examining the tradeoff involved between susceptibility to a soft error, and power saved by various power reduction schemes.

3.3 Power and Soft Errors

Power consumption is a major design constraint towards building more powerful chips and systems. The total power consumed in any circuit is composed of both dynamic power consumption and leakage power. To achieve higher density and performance and

lower power consumption, CMOS devices have been scaled for more than 30 years. Supply voltage has been scaled down in order to keep the power consumption under control. Hence, the transistor threshold voltage also has been commensurately scaled to maintain a high drive current and achieve performance improvement. However, the threshold voltage scaling results in the substantial increase of the subthreshold leakage current [14]. Hence, as technology is scaled, the leakage power consumed in circuits is increasing. On the other hand, even though the operating voltage is reduced the dynamic power is increasing due to higher operating frequency of the new generation circuits. Subsequently, there have been several efforts spanning from the circuit level to the architectural level at reducing the energy consumption (both dynamic and leakage). Circuit mechanisms include adaptive substrate biasing, dynamic supply scaling, dynamic frequency scaling, and supply gating [50, 42]. Many of the circuit techniques have been exploited at the architectural level to control leakage at the cache bank and cache line granularities. [50, 42] provide a comprehensive overview of these techniques. These optimizations influence the susceptibility of the circuits to soft errors. The subsequent sections will present the effect of two of the most dominant power reduction techniques namely, reducing supply voltage and increasing the threshold voltage, on soft error rates.

3.3.1 Impact of supply voltage scaling on soft error rate

Voltage scaling is a very common technique to reduce dynamic and leakage power consumption. Dynamic power of the circuit is proportional to the square of the supply voltage. Hence, supply voltage is decreased to reduce the power consumption of the circuit. To maximize the gains from this technique, it is a common practice to employ

clustered voltage design. In clustered voltage design, parts of a circuit operate at a lower voltage. Voltage level convertors are used to move from one voltage cluster to another [32]. Based on Equation 3.1 we know that SER increases exponential with reduction in $Q_{critical}$. $Q_{critical}$ is proportional the supply voltage. Hence, the SER increases exponentially with lower supply voltages. To examine the effect of supply voltage scaling on SER, we analyzed the soft error susceptibility of six level converters, and examine the impact of the voltage scaling on SER.

Supply voltage scaling is also employed to reduce the leakage energy. It is a common practice to reduce the supply voltage of the circuit when the circuit is not active and the overheads do not facilitate turning off the supply. For example in caches, when a cache line is not in use, the supply to the cache line can be reduced while still ensuring that the line retains the values [25]. The cache line is powered up before the values are accessed. A more detailed study of the effect of the supply voltage scaling in caches on soft errors is presented in [21, 34].

3.3.2 Impact of high threshold voltage devices on SER

There are two distinct factors that affect soft error rates due to increase in threshold voltages. First, due to the physical properties of high V_t silicon, we require higher energy to create electron-hole pairs in the substrate. This effect can potentially reduce SER. Second, higher V_t increases the gain and delay of circuits. This affects attenuation of the transient pulse.

3.3.2.1 Charge creation under high threshold voltages

This section gives a simplified theory of the semiconductors and we use this analysis to explain the phenomenon of charge creation under high V_t . Equation 3.3 represents the factors on which the threshold voltage depends.

$$V_t = V_{fb} + V_b + V_{ox} \quad (3.3)$$

where,

V_t is the threshold voltage of the MOS device

V_{fb} is the flat band voltage

1in V_b is the voltage drop across the depletion region at inversion

1in V_{ox} stands for potential drop across the gate oxide

When we change the threshold voltage of a device we change the flat band voltage(V_{fb}) of the device. Flat band voltage is the built in voltage offset across the MOS device [18]. It is the workfunction difference θ_{ms} which exists between polygate and silicon. By increasing the threshold voltage, we increase the energy required to push the electrons up the valence band. This is the same reason for which the device slows down. So when we increase the threshold voltage, the charge creation and collection characteristics change.

3.3.2.2 Logic attenuation due to high threshold voltage device

As mentioned in the earlier section in pass transistors and transmission gates the transient pulses attenuate due to V_t drop across the devices. But static CMOS sees

different trends. In static CMOS, the gain of the circuit is positive. The gain of an inverter is given by Equation 3.4

$$GainG = \frac{1 + r}{(V_m - V_t - V_{dsat}/2)(\lambda_n - \lambda_p)} \quad (3.4)$$

where

r is a ratio which compares the relative driving strength of NMOS transistor in comparison with PMOS transistor, V_M is the switching threshold (usually V_m is half of the supply voltage), V_{dsat} is drain saturation current, and λ_n , λ_p are channel length modulation factors for an n-channel and p-channel respectively. We can see that due to higher gain, a transient pulse will propagate in a system for a longer time and travels more logic stages. Another important fact to be considered is the delay. High V_t causes the device to slow down. Now in a simple logic network, under normal V_t , a particle strike will manifest as a bit flip only if the pulse is latched under a certain window. Any pulse occurring earlier or later will not be latched and hence will not result in a logic error. Assuming the stage takes t time units and the window of vulnerability is (t_v) , any error in the time interval $t - t_v$ will either be attenuated or will not be latched at the output. In case of high V_t , owing to the slower pulse and higher magnitude, t_v is longer, thus making the logic chain more susceptible to errors.

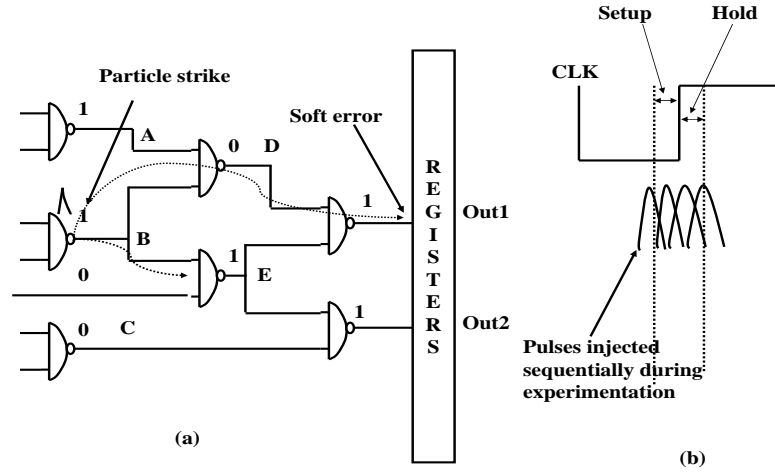


Fig. 3.1. Circuit level evaluation of soft errors in logic circuit

3.4 Methodology for circuit level analysis of soft errors

For a soft error to occur at a specific node in a circuit, the collected charge Q at that particular node should be greater than $Q_{critical}$. $Q_{critical}$ is defined as the minimum charge collected due to a particle strike that can cause a change in the output. If the charge generated by a particle strike at a node is more than $Q_{critical}$, the generated pulse is latched, resulting in a bit flip. This concept of critical charge is generally used to estimate the sensitivity of SER. The value of $Q_{critical}$ can be found by measuring the current required to flip a memory cell and derived using Equation 3.5. The particle strike itself is modeled as a piece wise linear current waveform where the waveform's peak accounts for funneling charge collection and the waveform's tail accounts for diffusion charge collection. To find the minimum height for which a wrong value is latch by the

memory element, the peak of the waveform is appropriately scaled. A similar approach has been used in prior work [43].

The actual magnitude of the charge is given by Equation 3.5.

$$Q_{critical} = \int_0^{T_f} I_d dt \quad (3.5)$$

Where, I_d is the drain current induced by the charged particle. T_f is the flipping time and in memory circuits it can be defined as the point in time when the feedback mechanism of the back-to-back inverter will take over from the incident ion's current. For logic circuits, T_f is simply the duration of the pulse. A pulse injected such that it reaches the input of the register within the latching window and this procedure is repeated such that the pulse sweeps the entire latching window, as shown in Figure 3.1(b). Among these pulses, the pulse which can be injected closest to the hold time and still cause an error, is chosen. Next, the magnitude of this pulse is changed to determine the minimum value of the pulse that can cause a error. The $Q_{critical}$ of the pulse is determined using the formulation provided by Equation 3.5. In this study, $Q_{critical}$ is the primary metric used to compare the SER of the designs, since the other parameters, such as charge collection efficiency are quite similar across designs.[21, 39] have characterized SER of different SRAM and flip-flop designs using similar procedure. In this study, two types of designs were used; memory elements which include 6T-SRAM, asymmetric SRAMs(ASRAM), flip-flops, and logic elements which include 6-inverter chain, 4-FO4 nand chain, 1-bit transmission gate (TG) based adders. All the circuits were custom designed using 70nm

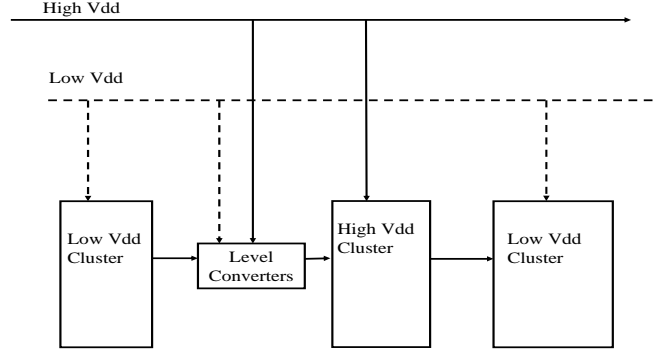


Fig. 3.2. Clustered voltage design

Berkeley predictive technology [1] and the netlists were extracted. The netlists were simulated using Hspice. The normal V_t of these devices is 0.22V, and the supply voltage of 1V is used. V_t is changed using *delvto* option of Hspice [7]. *Delvto* changes the V_t of the transistors by the amount specified.¹ All circuits were analyzed by changing V_t by 0.1V and 0.2V for both PMOS and NMOS.

3.5 Impact of supply voltage scaling on soft error rate

This section will detail three different ways of using supply voltage scaling to reduce the power consumption.

¹A more accurate method to change the V_t would involve changing V_{ox} in the transistor model. However due to the ease of the use of the DELVTO option (also used in [8]), this approach is used here.

3.5.1 Clustered voltage design

Voltage scaling is a very common technique to reduce dynamic and leakage power consumption. Dynamic power of the circuit is proportional to the square of the supply voltage. Hence, supply voltage is decreased to reduce the power consumption of the circuit. To maximize the gains from this technique, it is a common practice to employ clustered voltage design. In clustered voltage design, parts of a circuit operate at a lower voltage. Figure 3.2, provides a schematic view of the clustered voltage design. Voltage level converters are used to move from one voltage cluster to another [32]. While no level converting logic is needed to move from a high voltage cluster to a low voltage cluster, a level converting circuit is needed to move from a low voltage cluster to a high voltage cluster. Level converters are needed in the second case because in this case low voltage based devices need to drive high voltage based devices. In clustered voltage designs, the error can be generated in either the clusters or the level converters. Based on Equation 3.1 we know that SER increases exponential with reduction in $Q_{critical}$. $Q_{critical}$ is proportional to the supply voltage. Hence, the SER is exponentially dependent on supply voltages.

To examine the effect of supply voltage scaling in clustered voltage designs on SER, the soft error susceptibility of six level converters was analyzed. These level converters are presented in [32]. The effect of voltage scaling on the soft error susceptibility of six level converters is presented in Figure 3.3. Notice that the $Q_{critical}$ of these level converters is linearly dependent on the supply voltage.

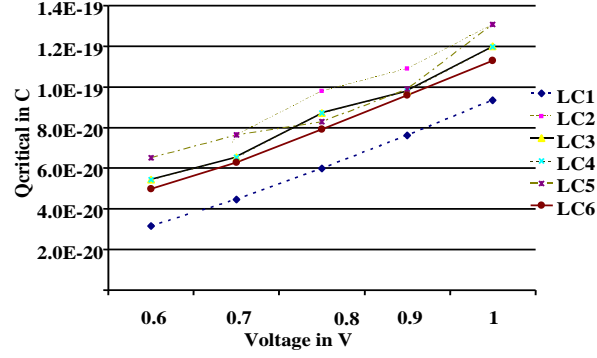
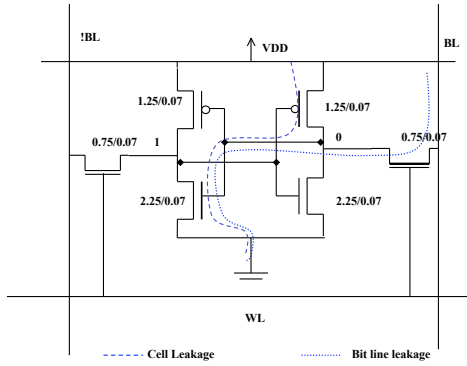


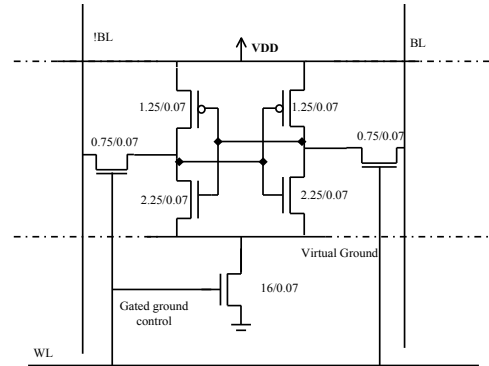
Fig. 3.3. $Q_{critical}$ vs the supply voltage for different level converters

3.5.2 Drowsy SRAM design

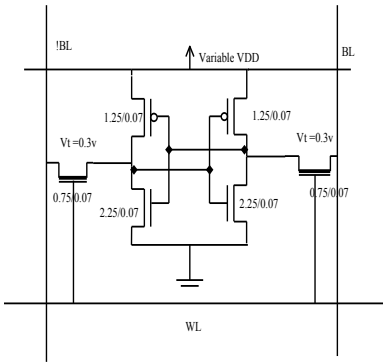
Supply voltage scaling is also employed to reduce the leakage energy. It is a common practice to reduce the supply voltage of the circuit when the circuit is not active and the overheads do not facilitate turning off the supply. For example in caches, when a cache line is not in use, the supply to the cache line can be reduced while still ensuring that the line retains the values, such a cell is called *drowsy cache* [25]. The drowsy cache cell due to its reduced supply voltage in the leakage control mode is clearly more susceptible to the soft errors as compared to the standard 6T SRAM. However, the use of DVS for controlling leakage offers an interesting opportunity for trade off between leakage reduction and soft error susceptibility. The effects of varying the supply voltages on leakage reduction and $Q_{critical}$ are plotted in Figure 3.5. $Q_{critical}$ has a linear relation with the supply voltage. It can be seen that even though there is a large reduction in leakage energy, there is a corresponding loss of immunity to soft errors.



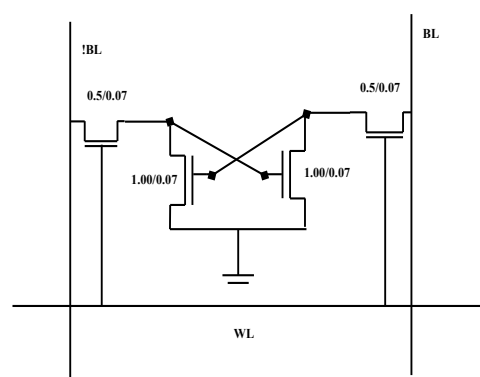
(a) Leakage in a 6T SRAM cell



(b) DRG Cache



(c) Drowsy Cache Cell



(d) 4T SRAM cell, without VDD

Fig. 3.4. Different SRAM designs

As it is the same circuit, all parameters influencing SER are the same except $Q_{critical}$. Hence, there is an exponential increase in the SER for reduction in supply voltage [please refer Equation 3.1]. Thus, it is important to appropriately choose supply voltages such that the leakage energy savings can be balanced with concerns of reliability.

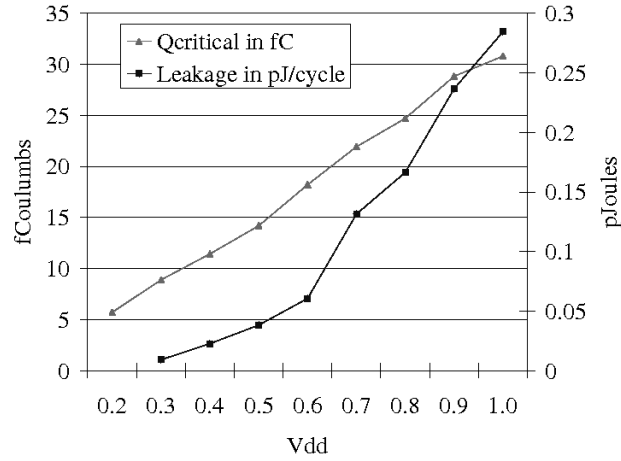


Fig. 3.5. $Q_{critical}$ Vs the supply voltage for drowsy cell

3.5.3 Gated-Gnd SRAM Design

Our first low power SRAM design is based on the gated-GND design proposed in [5, 33](see Figure 3.4(b)). In this method, an NMOS transistor is inserted in the path to ground. This NMOS acts as a switch to shut off the path to ground whenever the memory cell is accessed, and drastically reduces the leakage. The extra NMOS

is turned on only when the bit in the cell is read or written. Gated-GND technique results in less leakage due to the stacking effect of two NMOS transistors in the path to ground. This technique helps to reduce both cell and bitline leakage. Further, careful sizing of the NMOS transistor can enable the data value to be retained when leakage is controlled [5]. Our focus is on evaluating the soft error susceptibility when the cell is in low leakage mode (i.e., the cell is gated-GND). The creation of a virtual ground during the low leakage mode operation can potentially make this circuit more vulnerable to soft errors. In the DRG-Cache, when the array is shut off from the ground using the gated ground control, the virtual ground node does not stay at 0V and charges up to a higher voltage (0.4V). This makes this design more vulnerable to a 0 to 1 transition as a smaller induced charge is sufficient to trigger the flip. Table 3.1 presents the results. Referring to the values of $Q_{critical}$ from Table 3.1. The table also presents the $Q_{critical}$ values of a regular SRAM cell.

3.5.4 Quasi Static SRAM Design

All the above designs considered are commonly used in caches where it is required that the data stored be persistent (for at least few hundreds of cycles). But SRAM cells are also used in other structures such as branch target buffer (BTB) and branch predictors. In these memory structures the data is constantly replaced and is very transient in nature. So this makes them ideal candidates for using 4T based SRAM designs. The 4T based SRAM design has no pull up transistor and consequently has a floating node when storing a value of one (see Figure 3.4(d)). Hence, the value stored in the cell decays over time. The use of these 4T cells offers about 60-80% reduction in

leakage and about 12 to 33 % advantage in area [29]. Due to the absence of a pull-up network, these cells are anticipated to be even more susceptible to soft errors. Table 3.1 presents the results. Referring to the values of $Q_{critical}$ from Table 3.1, we observe that the 4T (without VDD) SRAM design is the most susceptible to a $1 \rightarrow 0$ flip. Due to the absence of the pull circuit in 4T (without VDD) SRAM design, the internal node decays naturally with time. On being disturbed by a transient pulse, the node discharges faster and loses the stored data.

As the value of a “1” is already decaying to a value of “0”, the transient current pulse, mimicking the cosmic particle strike, only accelerates the decay. Thus, the $Q_{critical}$ of this design is fundamentally different from other designs considered. The absence of the restoring power of the back-to-back inverters) makes the 4T based design highly susceptible to soft errors. Another notable difference is the increased immunity of the 4T (without VDD) SRAM design to a 0 to 1 flip. In contrast, to the natural decay process of the 1 to 0 flip, this is flip is against the normal decay.

Table 3.1. $Q_{critical}$ for different SRAM designs

	$Q_{critical}$ / fC		$Q_{critical}$ / fC
4T with out VDD	0.000071	4T with out VDD	61.631
Drowsy Cache at (0.3V)	8.899	DRG-Cache	132.632
6T Standard SRAM	15.104	6T Standard SRAM	325.482

(a) 1 to 0 flips

(b) 0 to 1flips

3.6 Impact of high threshold voltage devices on SER

The following sub-sections present the detail analysis of the effect of using high threshold devices for different types of circuit elements.

	ΔV_t	$Q_{critical}/C$	$Leakage/W$
ASRAM	0	4.75e-14	2.20e-07
	0.1	6.58e-14	9.10e-09
	0.2	7.58e-14	3.42e-10
SRAM	0	4.75e-14	2.40e-07
	0.1	4.04e-14	9.66e-09
	0.2	3.82e-14	9.46e-10

Table 3.2. $Q_{critical}$ and leakage power of SRAM and ASRAM with different V_t . Nominal V_t was 0.22V

3.6.0.1 Effect of V_t on SER of SRAM and Flip-flops

Table 3.2 presents the $Q_{critical}$ of the SRAM and Asymmetric SRAM cell. From Table 3.2, observe that the threshold change does not affect $Q_{critical}$ of the standard 6T SRAM significantly. By increasing V_t by 0.2V, no significant change in $Q_{critical}$ can be observed. Because the threshold voltage of both PMOS and NMOS in the back-to-back inverter configuration was changed, the regenerative property of the circuit ensures that there is no loss of charge and hence relatively no gains in terms of $Q_{critical}$. However, on analysis an ASRAM [8] optimized for leakage while storing a preferred logic state, a

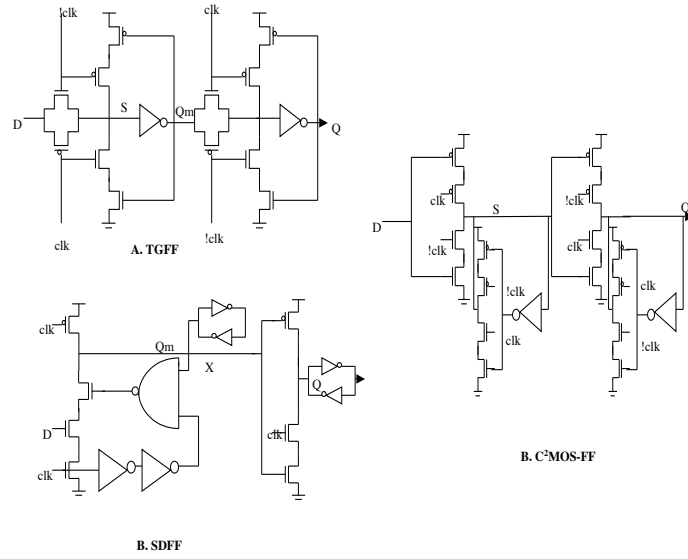


Fig. 3.6. Flip-Flops evaluated for SER

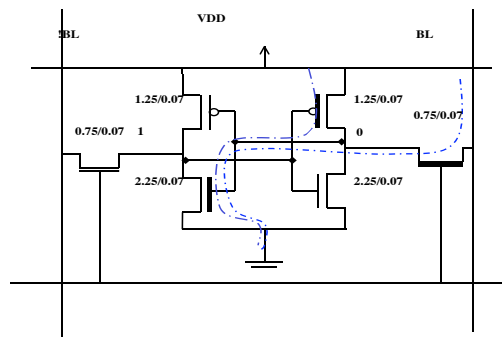


Fig. 3.7. Asymmetric SRAM: Optimized for 0

	ΔV_t	$Q_{critical}$ at input / C	$Q_{critical}$ at most susceptible node / C
SDFF	0	6.06e-21	1.24e-20
	0.1	5.08e-21	1.33e-20
	0.2	-	-
C^2MOS	0	3.69e-20	7.12e-21
	0.1	5.64e-20	7.12e-21
	0.2	1.68e-19	-
TGFF	0	1.99e-20	7.36e-21
	0.1	1.77e-19	7.36e-21
	0.2	3.87e-17	7.36e-21

Table 3.3. $Q_{critical}$ of different flip-flops. Nominal V_t was 0.22V

different trend is observed.

Figure 3.7 shows a circuit schematic of ASRAM optimized for reducing leakage when storing a 0. In ASRAM, the threshold voltage of transistors in the leaky path of circuit is increased to reduce leakage. For a stored value of 0, the transistors in the leaky path are shown. The V_t of these transistors is increased to reduce the leakage. The $Q_{critical}$ of this SRAM in its preferred state (i.e, when storing a 0) increases significantly, however for the non preferred state it remains the same. Specifically, when V_t is increased by 0.2V, $Q_{critical}$ increases by 59%. This is due to the fact that if one tries to charge the node to 1, the PMOS due to its high V_t will not be able to provide necessary feedback to quickly change the bit. However, if a value of 1 is stored, and one attempts to discharge it, then $Q_{critical}$ does not change as the NMOS is still at normal V_t . A similar behavior is also observed for an ASRAM designed for storing a preferred state of 1.

Three different flip-flops were characterized; transmission gate flip flop(TGFF), C^2MOS flip-flop(C^2MOS), and semi-dynamic flip-flop (SDFF). In each case, effect of increasing threshold voltages on $Q_{critical}$ was estimated . Figure 3.6 shows detailed schematics of these designs. The blank fields in the table represents the points where the flip-flop became unstable and could not latch the input data. There are two different aspects that should be investigated with respect to the effects of threshold voltages and susceptibility of soft errors on flip-flops. First, the soft error rate of the flip-flop itself could change. This is found by evaluating $Q_{critical}$ at the most susceptible node [39]. Second the ability of the flip-flop to latch onto an error at its input could change. This effect will be useful in analyzing its behavior in a datapath. Hence Table 3.3, lists the $Q_{critical}$ at both the nodes for all the flip-flops.

From Table 3.3, note that, for a TGFF, $Q_{critical}$ at the input node, when V_t is increased, while the $Q_{critical}$ at the node S is same. This trend is ascribed to the presence of to the transmission gate present at the input. On the other hand for the node S, the higher gain of the inverter cancels out the effect of the transmission gate at the slave stage and hence the $Q_{critical}$ remains almost constant. Similar testing was done on a C^2MOS flip-flop which also has master-slave stages similar to that of the transmission gate flip-flop. Since C^2MOS flip-flop does not have any transmission gate based structures it has a lower $Q_{critical}$ compared to the TGFF. One of the pulse triggered designs SDFF, was investigated for it's $Q_{critical}$. SDFF has few large sized devices in its feedback path thus resulting in a much higher $Q_{critical}$ at the most susceptible node (X) as compared to the other flip-flops considered. Since this node feeds back into a NAND gate, when

the threshold increases, due to the increase in delay of the NAND gate and 2 inverters, $Q_{critical}$ increases. Thus the flip-flop by itself has a higher $Q_{critical}$ as threshold voltage increases. At the input the larger overlap time helps pull down the voltage at node X and hence reduces the $Q_{critical}$.

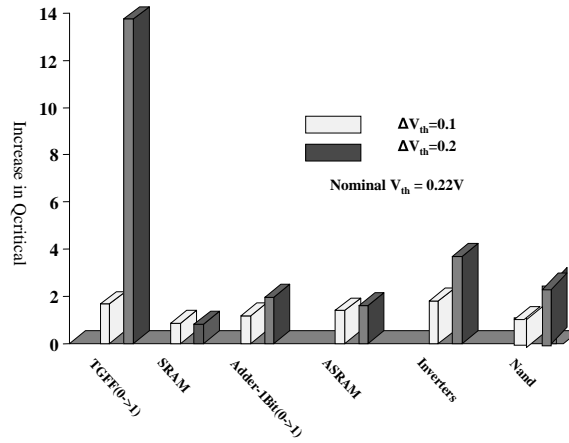


Fig. 3.8. Increase in $Q_{critical}$ of different designs with respect to operating nominal V_t of 0.22V

3.6.0.2 Effect of V_t on Combinational Logic

Three kinds of logic circuits were analyzed: chain of 6-inverters, chain of 4-nand gates and transmission gate based full adders. For all of these circuits we check for an error by latching the transient pulse at the end of the logic chain. A transmission gate flip-flop (TGFF) was used to latch the values. TGFF was chosen as it is one of the most

commonly used flip-flop. From Table 3.4, note that the $Q_{critical}$ of the circuit is increasing for increasing threshold voltages. For TG based adders, the threshold drop across transmission gates accounts for the increase in $Q_{critical}$ as the V_t increases. However, for static logic this is counter intuitive. Based on the pulse propagation characteristics, the $Q_{critical}$ of the circuits should be lower. This can be accounted for the robustness of flip-flops. In Figure 3.8, find that the $Q_{critical}$ increase for the flip-flop is many orders of magnitude higher than the others. To confirm the observations, the 6-inverter chain was simulated again with normal- V_t flip-flops at the output of the logic chain and it was found that as the V_t increased, the $Q_{critical}$ values decreased. The results are presented in Figure 3.9. Next section, shows how this fact can be leveraged to reduce power and increase robustness of the circuit.

3.6.1 Effect of delay balance using high V_t devices on soft errors

Figure 3.10 shows a typical pipeline. The logic between pipeline stages is distributed across slow and fast paths, with the slowest path determining the clock frequency. Thus, slow paths become critical paths and fast paths become non-critical paths. It is an accepted practice to use high V_t devices on non-critical paths. Since these are not delay sensitive, one can achieve high leakage power savings with minimal performance penalty. This is some times referred to as *delay balancing*. To examine the effect of delay balancing on $Q_{critical}$, two circuits were simulated; one with a 6-inverter chain which forms the critical path and the other with a 3-inverter chain. Figure 3.11, shows the $Q_{critical}$ of the 6-inverter chain as compared to the $Q_{critical}$ of the 3-inverter chain with

both low and high V_t TGFFs. On performing delay balancing on this logic with low V_t TGFF, and high V_t 3-inverter chain, one can observe the $Q_{critical}$ of 3-inverter chain reduces. Thus, it is evident that this path now becomes more vulnerable to soft errors. Consequently, if a high- V_t flip-flop is used for latching, the $Q_{critical}$ of the 3-inverter chain (relative to the 6-inverter chain) is still high. So, while performing delay balancing it is recommended to use high V_t flip-flops at the end of the logic chain to improve the immunity to SER.

3.7 Conclusion

This chapter estimates the effect of the high threshold voltages and voltage scaling on SER. We find that for certain designs like transmission gate based designs SER reduces while for static logic SER deteriorates as V_t is increased. The experiments show that, as in ASRAM, using high V_t cleverly can reduce both SER and leakage power. Finally, the study finds that the use of high V_t for delay balancing can potentially increase SER, but the reliability can be bought back by the use of high V_t flip-flops. In general, the study shows that use of high V_t devices not only reduces leakage but also affects the reliability of circuit. In contrast, voltage scaling almost always increases the susceptibility to SER.

	ΔV_t	$Q_{critical}/C$	$Leakage/W$
Nand	0	1.31e-20	22.56e-07
	0.1	2.26e-20	9.92e-09
	0.2	2.83e-20	4.90e-10
Inverters	0	1.28e-20	2.20e-07
	0.1	2.3e-20	4.90e-10
	0.2	4.73e-20	41.99e-11
TG Adder	0	4.60e-20	1.18e-07
	0.1	1.35e-19	3.42e-08
	0.2	5.87e-17	3.40e-08

Table 3.4. $Q_{critical}$ and leakage power of various designs with different V_t . A high V_t TGFF was used at the output of the logic chain

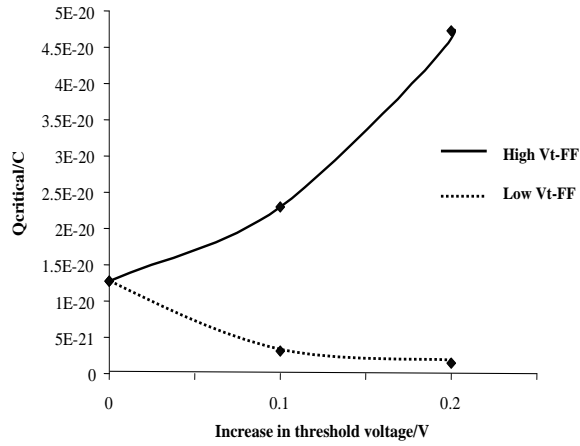


Fig. 3.9. $Q_{critical}$ of typical inverter chain with either high V_t or low V_t flip-flop at the output of the chain. Nominal V_t was 0.22V

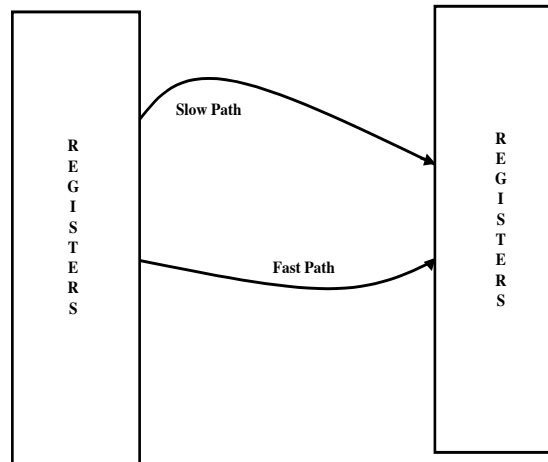


Fig. 3.10. Delay Balancing

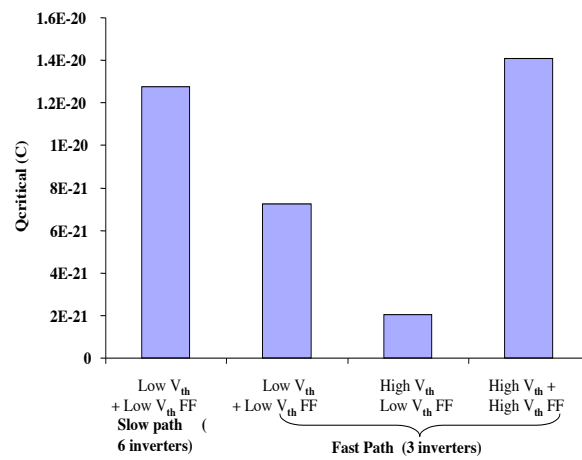


Fig. 3.11. Effect of delay balancing

Chapter 4

Testing of neutron induced soft errors

Soft error rate (SER) testing reported in literature has involved both neutron and alpha particles. Testing the effects of alpha particles on semiconductor devices is relatively simple. Alpha particle testing experiments are carried out either with alpha particles originating from ^{238}Th foil, or a similar Alpha emitting substance. In contrast to neutrons, alpha particles cannot penetrate materials thicker than a few microns. Consequently, most alpha particles that are of concern come from either processing material or packaging material. In contrast to alpha particles, neutrons cannot be shielded by even a few feet of concrete. Hence, accelerated neutron testing is used for our soft error testing experiments. Beam 30L of Weapon Neutron Research at the Los Alamos National Laboratory is a JEDEC prescribed test beam for soft errors [30], and is the only one of its kind. This beam is highly stable and it closely replicates the energy spectrum of terrestrial neutrons in the 2-800 MeV range while providing a very high neutron flux. Most SER testing reported in literature recently, were performed at this facility [27, 35, 26]. The following sections, describe the test facility built at Penn State's Nuclear engineering Facility for accelerated soft error testing. Two off-the-shelve memory cards where tested at this location, the results of which are presented in the subsequent section.

4.1 Test Facility

The Radiation Science and Engineering Center (RSEC) at The Pennsylvania State University serves as test facility for this study. It includes the Penn State Breazeale Nuclear Reactor (BNR). The BNR is a 1 MW, TRIGA (Training, Research, Isotopes, General Atomics) nuclear reactor with moveable core in a large pool and with pulsing capabilities. The core is located in a 24 ft deep pool with 71,000 gallons of demineralized water. A variety of dry tubes and fixtures are available in or near the core. When the reactor core is placed next to a D2O tank and the graphite reflector assembly near the beam port locations, thermal neutron beams become available from two of the seven existing beam ports. In steady state operation at 1 MW, the thermal neutron flux is 10^{13} n/cm²s at the edge of the core and 3×10^{13} n/cm²s at the central thimble. The PSBR can also pulse with the peak flux for maximum pulse 6×10^{16} n/cm²s with a pulse half width of 10 msec. The maximum rated power of the reactor is 1 MW in the continuous mode and can be increased to 2000 MW in the pulse mode. The reactor power can also be stepped from 10 W to 1 MW to observe the soft error rate dependence on neutron flux.

Neutrons are classified based on their energies. Neutrons with energy more than 0.1MeV are called fast neutrons and those with energy less than that are called epithermal and thermal neutrons. The atmospheric neutron flux has both of these components, though neutrons at lower energy are far numerous than higher energy neutrons. The neutron flux emitted from the reactor core has both fast and thermal components. Near the reactor core, the neutron flux is dominated by fast neutron. The fast neutron flux

near the core can be accessed through long tubes (standpipes) and will be used to test the effect of the fast neutrons on the devices. This fast neutron flux from the reactor core passes through a heavy water tank, so the resultant neutron flux at the output port is that of thermal neutrons. The average thermal flux at the exit of the beam port is about 3×10^7 neutrons/cm²s. Neutron spectrum of this beam port is measured with a slow neutron chopper and shown in 4.1. The spectrum measurement is compared with a corresponding Maxwell Boltzmann distribution. Since the generation of the neutrons is a statistical phenomenon, it is usually difficult to isolate the neutrons of a given energy. The fast neutron spectrum should be calibrated before drawing conclusions from the study. However, upper energy limit of fast neutron flux from the reactor is less than the atmosphere neutrons spectrum, but it corresponds to the dominant portion of the atmosphere neutron flux.

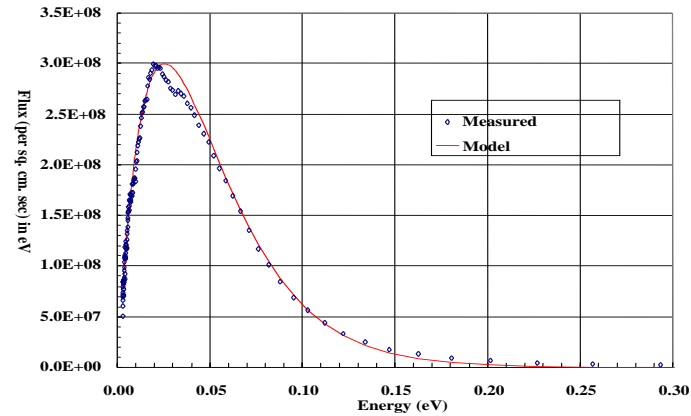
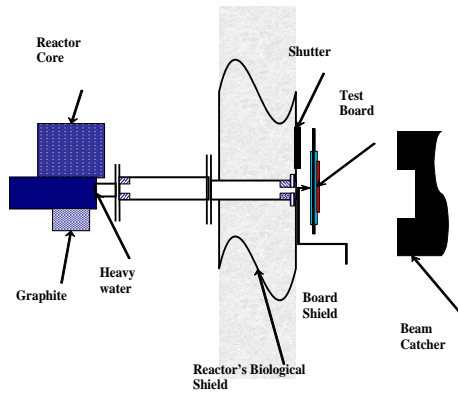
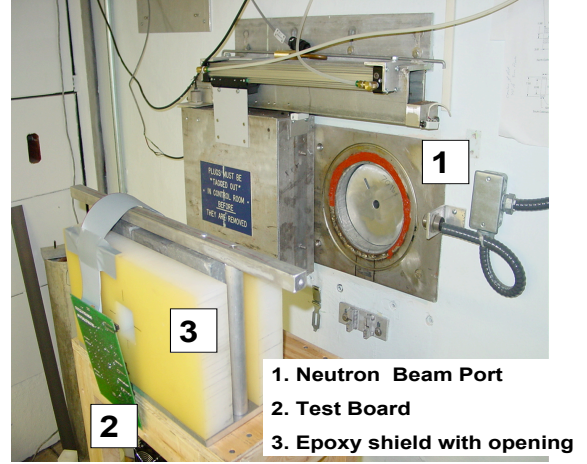


Fig. 4.1. Thermal neutron energy distribution at the exit of a horizontal neutron beam port at PSBR



(a) Schematics



(b) Test picture

Fig. 4.2. Testing the SRAM memories at the outer beam opening

4.2 Experimental Setup for Soft Error Rate Measurements

One set of experiments were performed using the neutron flux obtained at the beam ports after the heavy water tanks. This neutron beam has thermal neutrons as the dominant portion of the flux. Currently, two commercially available SRAM chips were tested at this port. The following sections explain the different test setups that will be used to test the memory chips.

4.2.1 Devices Under Test

The proposed accelerated soft error testing is on commercial SRAM based memories. Two commercially available memory chips from two different vendors were tested. The results of the tests performed using these devices are presented in Figures 4.3 and 4.4.

In these experiments, a fixed pattern of data was written into the memory chips and then compared after exposure to the neutron beam to identify soft errors. The experimental setup for soft error rate measurements consists of a custom board interfaced with a computer through a GPIB card (from National Instruments). The board is controlled through software interface built using Labview. The controlling application consists of simple routines to read and write a user specified value across the whole memory. During the readout, it compares the written value to the value in each address. The circuit board is placed in the neutron beam line, and connected to a PC outside using a 25-ft cable. Schematics of the experimental setup for soft error rate measurements at PSBR is shown in Figure 4.2(a) and the picture of the setup from the neutron beam cave is shown in Figure 4.2(b). This configuration allows for continuous read-write and changing the operating conditions without interrupting the experiment. The board was tested online multiple times in the actual setup before the reactor is started. The board was exposed to neutron beam after the reactor reached the desired power level. Tests results from a Cypress Chip (CY7C128A-20PC, 16k bits) and Toshiba chip (TC554001AFT7L, 4Mbits) are presented in Figures 4.3(b) and 4.3(a). The principle motivation of the experiment was to understand the interactions between power and SER. Reducing operating voltage is a popular method to reduce power consumption. The Cypress and Toshiba chips were found to be operational at 3V and 4V respectively. They are rated to operate at 5V. Next to understand the effect of neutron flux on soft error the reactor was operated at different power levels. The intensity of neutrons is directly related to the reactor power level.

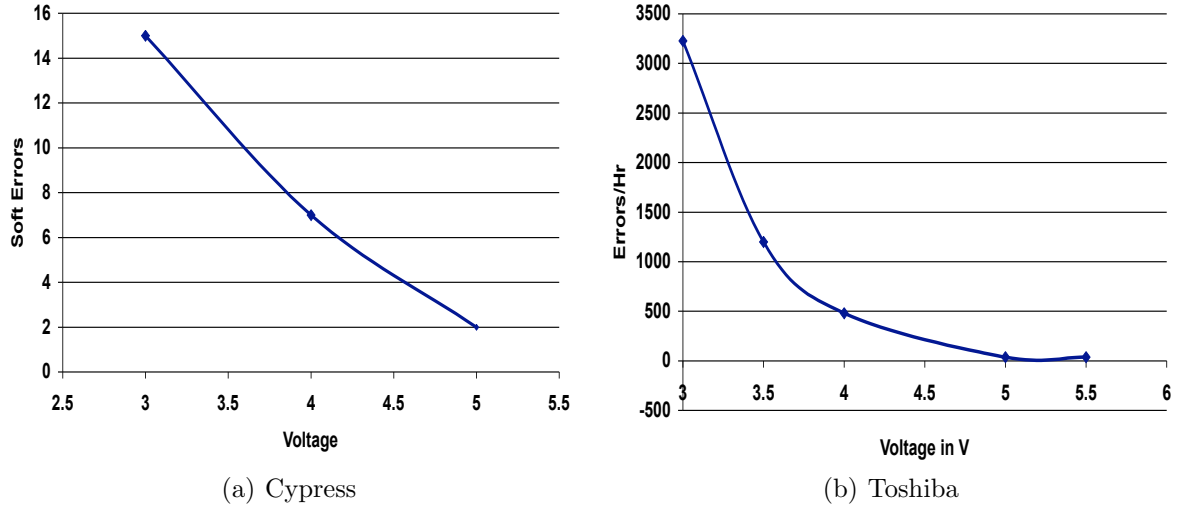


Fig. 4.3. Effect of supply voltage on soft errors

4.3 Analysis of the results

Figures 4.3(b) and 4.3(a) present the effect of supply voltage variation on the soft error rate. As presented in Equation 3.1, repeated here for convenience, the soft error rate is dependent on the $Q_{critical}$ and hence on the operating voltage.

$$SER \propto N_{flux} * CS * \exp \frac{-Q_{critical}}{Q_s} \quad (4.1)$$

where

N_{flux} is the intensity of neutron flux,

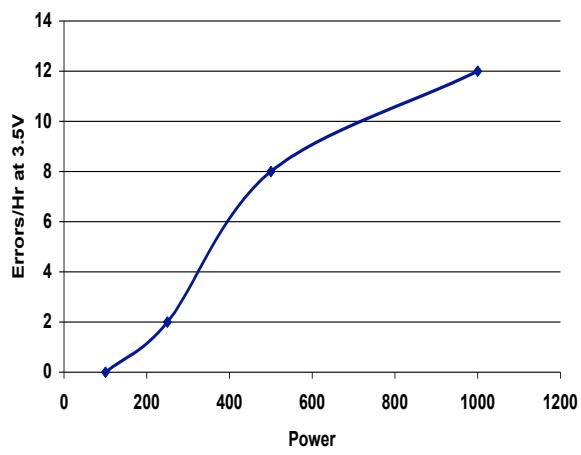
CS is the area of cross section of the node,

Q_s is the charge collection efficiency,

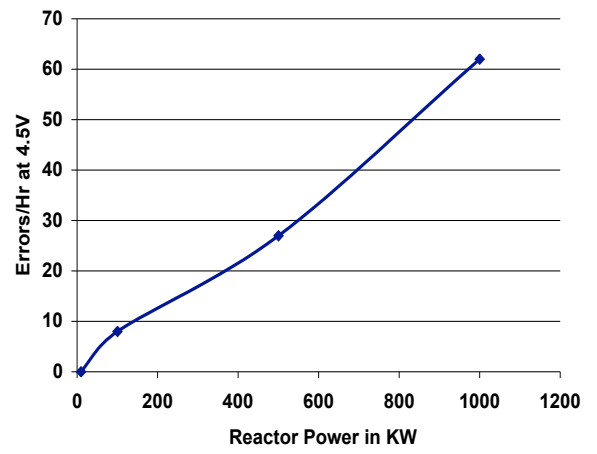
$Q_{critical}$ is the charge that is stored at the node and hence is equal to $VDD * C_{node}$, where VDD is the supply voltage and C_{node} is the nodal capacitance.

While employing voltage scaling for power reduction, there is a reduction in the $Q_{critical}$ of the cell. This was shown using circuit simulation in Section 3.5. Now, if all the other factors remained same there should be a super-linear increase in the SER. However, based on the Figures 4.3 we see a linear increase in the SER for a corresponding decrease in the voltage. This is attributed to the results presented in Section 2.4.4. In Section 2.4.4 it was shown that for a change in supply voltage, the resultant current transient also changed. As the supply voltage was reduced, the magnitude of the current changed. This affects the regenerative feedback of the SRAM cell. Thus, though the error rates increase with decrease in the operating voltage. The increase in SER is proportional to the decrease in voltage.

For examining the statistical accuracy of the accelerated tests we performed the tests at various reactor power levels. As mentioned previously, the reactor power level is directly dependent on intensity of the neutron flux. Hence, we can control the neutron flux, by controlling the reactor power level. The results are presented in Figures 4.4. We can notice that SER is directly proportional reactor power. While the SER for the Cypress chip is not as linear as the Toshiba chip, this is attributed to the relatively small size of the Cypress Chip. Hence, it is suggested that for statistical accuracy of accelerated SER measurement be performed using larger memories.



(a) Cypress



(b) Toshiba

Fig. 4.4. Effect of neutron Flux on soft errors

Chapter 5

Conclusion

As supply voltages reduce and feature sizes become smaller in future technologies, soft error tolerance is considered a significant challenge for designing future electronic systems. For example, a 1GB memory system based on 64Mbit DRAMs has a combined error rate of 3435 FIT (failure in 10⁹ hours of operation) when using single error correction and double error detection [22]. An even higher soft error rate of 4000 FIT was reported for a typical processor [36], with approximately half of the errors affecting the processor core and the rest affecting the cache [19]. Such errors also affect the fast growing FPGA segment. Recently, Xilinx investigated soft errors by designing three circuit boards, each containing 100 Virtex II devices. These were placed at sea level and 3,600m and monitored 24 hours a day. At sea-level there were four soft errors in 262 days, while at 3,600m there were eight upsets in 28 days [9]. The problem is expected to compound in future technology generations as critical charge per node will reduce while the number of nodes per chip is expected to increase. Consequently, a smaller charge is sufficient to upset the logical value of a node and more nodes can be affected. The first effect is reflected by measurements reported in [24] that showed soft error rates increased from 2500 FIT/Mb to 3000 FIT/Mb when moving from 130nm to 90nm feature sizes. In addition, transient glitches created in combinational logic by particle strikes are more

likely to be latched with increase in clock frequencies [10]. Consequently, both sophisticated servers (due to the numerous electronic components they contain) and embedded systems (due to the high volume of units sold) require soft error-aware design.

The issue of SEE was first studied in the context of scaling trends of microelectronics in 1962 [49]. Interestingly, the forecast from this study that the lower limit on supply voltage reduction will be imposed by SEE is shared by a recent work from researchers at Intel [14]. However, most works on radiation effects, since the work in 1962, focused on space applications rather than terrestrial applications. As earth's atmosphere shields most cosmic ray particles from reaching the ground and charge per circuit node used to be large, SEE on terrestrial devices has not been important until recently. The galactic flux of primary cosmic rays (mainly consisting of protons) is very large, about 100,000 particles/ m^2 as compared to the much lower final flux (mainly consisting of neutrons) at sea level of about 360 particles/ m^2 [55]. Only few of the galactic particles have adequate energy to penetrate the earth's atmosphere. However, with continued scaling of feature sizes and the use of more complex systems, soft errors in terrestrial applications are becoming an increasing concern and have drawn attention since late 1990s. There have been various documented failures due to soft errors ranging from memories used in large servers and aircrafts to implantable medical devices like cardiac defibrillators [23]. A widely cited soft error episode involves L2 caches with no error correction or protection that caused Sun Microsystem's flagship servers to crash suddenly and mysteriously [4]. This problem resulted in loss of various customers for Sun Microsystems. More ominous than this failure can be errors in embedded devices such as cardiac defibrillators that are becoming an integral part of our society. As computing systems become an integral part

of various critical applications ranging from medical implants to fly-by-wire aircrafts, immunity against soft errors becomes more critical for the society as a whole. While these problems may seem to be easily solved through techniques traditionally used to counter radiation effects in space applications, they are not suitable for adoption by commercial manufacturers of terrestrial devices as many of the solutions consume more power, reduce manufacturability and severely influence IC performance [23]. Even space applications are moving away from the use of radiation hardened process technology. They are using commercial off-the-shelf components that employ soft error protection techniques at software and architectural level for cost and performance reasons. As a result, many researchers have been focusing on devising new soft error countermeasures ranging from process to software levels. Advances in process technology such as adoption of SOI, elimination of Boron-10 (B^{10}) impurities are expected to mitigate the soft error problem to a certain extent. However, solutions at higher levels will still be necessary for reliable operation of the computing system

The work leading up to this thesis has underscored the importance of soft errors in the context of low power design. The thesis first proposed SEAT-DA toolset to model the soft errors. The lack of fault models that abstract the physical phenomena of soft errors accurately in a fashion that is accessible to computer engineers and the absence of tools that analyze the effectiveness of soft error countermeasures are affecting researchers in their quest for tackle the soft error problem. The proposed toolset SEAT will fill this critical void by modeling the nuclear reaction effects of particle strikes on semiconductor devices and by creating higher-level abstractions of these effects for analysis at the circuit and architectural level. This infrastructure will enable researchers working on circuit,

architectural and software countermeasures for soft errors to obtain a better perspective of the physical phenomena and help them tune their techniques accordingly. If the fault model used at architectural or circuit level fail to model the SEE accurately, the underlying value of solutions proposed at higher abstractions become meaningless.

Next the thesis has explored the effect of various power optimization schemes on soft errors. This work also explores the correlation for different circuits structures such as SRAMs, flip-flops and logic. Further, it has demonstrated the potential tradeoffs between power savings and susceptibility to soft error rates when using high V_t devices and reducing frequency. We believe the results from this work will help designers strike a good balance between reliability and power savings.

Finally, the work in the test setup at BNR has given an understanding of the impact of neutron induced soft errors on SRAM memory. Most commercial soft error testing in U.S.A. is performed at the Los Alamos test facility, access to which is expensive and cumbersome due to security clearances that are required. Consequently the expertise developed in setting up the current facilities will help the research community with alternate soft error testing facility. Preliminary testing in BNR has shown that that voltage scaling reduces the robustness of the circuit against soft errors.

This work has also contributed to several other important works that including [34, 28, 46]. In [34], the effects on voltage scaling on SER in memories are examined in terms of the cache power saving protocols, *cache decay* and *drowsy cache*. In [28], the effect of power savings and soft errors has been examined in the domain of VLIW processors. In [46], the finding of [21], are extended in the context of FPGA. In addition to these, the work potentially will impact other spheres of designs as well. For

example, the work can be easily extended to soft error-aware gate-level synthesis, and energy aware error correction.

References

- [1] Berkeley predictive model. <http://www-device.eecs.berkeley.edu>.
- [2] Mcnp available from <http://laws.lanl.gov/x5/mcnp/index.html>.
- [3] Trim. available at <http://www.srim.org>.
- [4] Sun microsystems- soft memory errors and their effect on sun fire systems, 2002.
- [5] Amit Agarwal, Hai Li, and Kaushik Roy. DRG-cache: a data retention gated-ground cache for low power. In *Proceedings of the 39th conference on Design automation (DAC-39)*, pages 473–478, 2002.
- [6] L. Anghel, D. Alexandrescu, and M. Nicolaidis. Evaluation of a soft error tolerance technique based on time and/or space redundancy. In *Proceedings of the 13th symposium on Integrated circuits and systems design*, page 237. IEEE Computer Society, 2000.
- [7] Avant! *Hspice User Manual*, 2003 edition.
- [8] N. Azizi, A. Moshovos, and F. N. Najm. Low-leakage asymmetric-cell sram. In *Proceedings of the 2002 International Symposium on Low Power Electronics and Design*, pages 48–51, 2002.
- [9] R. Ball. Xilinx Attacks Actel Stance Over Soft Error Rates in ICs. In *Electronics Weekly*, October 2003.

- [10] R. Baumann. The impact of technology scaling on soft error rate performance and limits to the efficacy of error correction. In *Digest. International Electron Devices Meeting, 2002. IEDM '02*, pages 329–332, 2002.
- [11] R. C. Baumann. Soft errors in advanced semiconductor devices-part I: the three radiation sources. *IEEE Transactions on Device and Materials Reliability*, 1(1):17–22, 2001.
- [12] M.P. Baze and S.P.Buchner. Attenuation of single event induced pulses in CMOS combinational logic. *Nuclear Science, IEEE Transactions on*, 44(1):2217–2223, December 1997.
- [13] S. Borkar, T. Karnik, and V. De. Design and reliability challenges in nanometer technologies. In *41st Design Automation Conference*, 2004.
- [14] Shekhar Borkar. Design challenges of technology scaling. *IEEE Micro*, 19(4):23–29, 1999.
- [15] P. D. Bradley and E. Normand. Single event upsets in implantable cardioverter defibrillators. *Single event upsets in Implantable cardioverter defibrillators*, 45:2929–2940, December 1998.
- [16] S. Buchner, M. Baze, D. Brown, D. McMorow, and J. Melinger. Comparison of error rates in combinational and sequential logic. *Nuclear Science, IEEE Transactions on*, 44(1):2209–2216, December 1997.

- [17] K. Castellani-Coulie, B. Sagnes, F. Saigne, J.-M. Palau, M.-C Calvet, P. E Dodd, and F. W Sexton. Comparison of nmos and pmos transistor sensitivity to seu in srams by device simulation. *IEEE Transactions on Nuclear Science**IEEE Transactions on Nuclear Science*, 50(6):2239–2244, December 2003.
- [18] J. Y. Chen. *CMOS Devices and Technology for VLSI*. Prentice-Hall, Englewood Cliffs, NJ, 1990.
- [19] Tandem Compaq Corporation. Data Integrity for Compaq NonStop Himalaya Server. In *Compaq White Paper*, 1999.
- [20] V. Degalahal, S. Cetiner, F. Alim, N. Vijaykrishnan, K. Unlu, and M. J. Irwin. Sese: Soft error simulation and estimation engine. In *2004 MAPLD International Conference*, 2004.
- [21] V. Degalahal, N. Vijaykrishnan, and M.J Irwin. Analyzing soft errors in leakage optimized sram design. In *Proceedings of 16th International Conference on VLSI Design*, pages 227 –233, January 2003.
- [22] T. J. Dell. A White Paper on the benefits of Chipkill - Correct ECC for PC Server Main Memory. In *IBM Microelectronics Division*, 1997.
- [23] P. E. Dodd and L. W. Massengill. Basic mechanisms and modeling of single-event upset in digital microelectronics. In *IEEE Transactions on Nuclear Science*, 50(3):583–602, June 2003.

- [24] E. Dupont. The nanometer world is not so dark. In *IRoC technologies Newsletter*, September 2003.
- [25] Krisztin Flautner, Nam Sung Kim, Steve Martin, David Blaauw, and Trevor Mudge. Drowsy caches: simple techniques for reducing leakage power. In *Proceedings of the 29th annual international symposium on Computer architecture (ISCA-29)*, pages 148–157, 2002.
- [26] P. Hazucha, T. Karnik, J. Maiz, S. Walstra, B. Bloechel, J. Tschanz, G. Dermer, S. Harelund, P. Armstrong, and S. Borkar. Neutron soft error rate measurements in a 90-nm CMOS process and scaling trends in SRAM from 0.25-micron to 90-nm generation. In *Digest. International Electron Devices Meeting, 2003. IEDM '03*, 2003.
- [27] P. Hazucha and C. Svensson. Impact of cmos technology scaling on the atmospheric neutron soft error rate. In *IEEE Transactions on Nuclear Science*, 47(6), 2000.
- [28] Jie Hu, Fei Li, Vijay Degalahal, Mahmut Kandemir, N. Vijaykrishnan, and Mary Jane Irwin. Compiler-directed instruction duplication for soft error detection. In *to appear in Proceedings of the 2005 DATE*, 2005.
- [29] Z. Hu, P. Juang, P. Diodato, S. Kaxiras, K. Skadron, M. Martonosi, and D. Clark. Managing leakage for transient data: decay and quasi-static 4t memory cells. In *the 2002 International Symposium on Low Power Electronics and Design*, pages 52–55, 2002.

- [30] JEDEC. Standard jesd89-measurement and reporting of alphaparticles and terrestrial cosmic ray-induced soft errors in semiconductor devices. (<http://www.jedec.org/download/default.cfm>).
- [31] T. Karnik, S. Vangal, V. Veeramachaneni, P. Hazucha, V. Erraguntla, and S. Borkar. Selective node engineering for chip-level soft error rate improvement [in cmos]. In *VLSI Circuits Digest of Technical Papers, 2002. Symposium on*, pages 204– 205, 2002.
- [32] S. H. Kulkarni and D. Sylvester. High performance level conversion for dual v/sub dd/ design. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 12(9):926– 936, September 2004.
- [33] L. Li, Ismail Kadayif, Yuh-Fang Tsai, Narayanan Vijaykrishnan, Mahmut T. Kandemir, Mary Jane Irwin, and Anand Sivasubramaniam. Leakage energy management in cache hierarchies. In *Proceedings of the 2002 International Conference on Parallel Architectures and Compilation Techniques*, pages 131–140. IEEE Computer Society, 2002.
- [34] Lin Li, Vijay Degalahal, N. Vijaykrishnan, Mahmut Kandemir, and Mary Jane Irwin. Soft error and energy consumption interactions: A data cache perspective. In *Proceedings of the 2004 international symposium on Low power electronics and design*, pages 132–137, 2004.

- [35] J. Maiz, S. Hareland, K. Zhang, and P. Armstrong. Characterization of multi-bit soft error events in advanced SRAMs. In *Digest. International Electron Devices Meeting, 2003. IEDM '03*, 2003.
- [36] A. Messer and et. al. Susceptibility of modern systems and software to soft errors. In *HP Labs Technical Report HPL-2001-43*, 2001.
- [37] J.M. Palau, G. Hubert, K. Coulie, B. Sagnes, M.C, Calvet, and S. Fourtine. Device simulation study of the seu sensitivity of srams to internal ion tracks generated by nuclear reactions. *IEEE Transactions on Nuclear Science*, 48(2):225–231, April 2001.
- [38] R. Rajaraman, N. Vijaykrishnan, Y. Xie, M. J. Irwin, and K. Bernstein. Soft errors in adder circuits. In *MAPLD*, 2004.
- [39] R. Ramanarayanan, V. Degalahal, N. Vijaykrishnan, M. J. Irwin, and D. Duarte. Analysis of soft error rate in flip-flops and scannable latches. In *SOC Conference, 2003. Proceedings. IEEE International [Systems-on-Chip]*, pages 231 – 234, September 2003.
- [40] Joydeep Ray, James C. Hoe, and Babak Falsafi. Dual use of superscalar datapath for transient-fault detection and recovery. In *Proceedings of the 34th annual ACM/IEEE international symposium on Microarchitecture*, pages 214–224. IEEE Computer Society, 2001.

- [41] Eric Rotenberg. Ar-smt: A microarchitectural approach to fault tolerance in microprocessors. In *Proceedings of the Twenty-Ninth Annual International Symposium on Fault-Tolerant Computing*, page 84. IEEE Computer Society, 1999.
- [42] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proceedings of the IEEE*, 91(2):305–327, Feb 2003.
- [43] N. Seifert, D. Moyer, N. Leland, and R. Hokinson. Historical trend in alpha-particle induced soft error rates of the *AlphaTM* microprocessor. In *39th Annual IEEE International Reliability Physics Symposium*, pages 259–265, 2001.
- [44] P. Shivakumar, M. Kistler, S. Keckler, D. Burger, and L. Alvisi. Modeling the effect of technology trends on the soft error rate of combinational logic. In *International Conference on Dependable Systems and Networks*, 2002.
- [45] G. R. Srinivasan, H. K. Tang, and P. C. Murley. Parameter-free, predictive modeling of single event upsets due to protons, neutrons, and pions in terrestrial cosmic rays. In *IEEE Trans. Nucl. Sci.*, 41:2063–2070, December 1994.
- [46] Suresh Srinivasan, Aman Gayasen, N. Vijaykrishnan, Mahmut Kandemir, Yuan Xie, and Mary J. Irwin. Improving soft-error tolerance of fpga configuration bits. In *Proceedings of the International Conference on Computer Aided Design (ICCAD)*, 2004.

- [47] Y. Tosaka, H. Kanata, T. Itakura, and S. Satoh. Simulation technologies for cosmic ray neutron-induced soft errors: Models and simulation systems. *IEEE Transactions on Nuclear Science*, 46(3):774–780, June 1999.
- [48] Y. Tosaka, H. Kanata, S. Satoh, and T. Itakura. Simple method for estimating neutron-induced soft error rates based on modified bgr model. *Electron Device Letters, IEEE*, 20(2):89–91, February 1999.
- [49] J. Wallmark and S. Marcus. Minimum size and maximum packaging density of non-redundant semiconductor devices. *In Proc. IRE*, 50:286–298, 1962.
- [50] Liqiong Wei, K. Roy, and V. K. De. Low voltage low power CMOS design techniques for deep submicron ICs. In *VLSI Design, 2000. Thirteenth International Conference on*, pages 24–29, 2000.
- [51] F. Wrobel, J.-M Palau, M.C Calvet, O. Bersillon, and H. Duarte. Incidence of multi-particle events on soft error rates caused by n-si nuclear reactions. *IEEE Transactions on Nuclear Science*, 47(6):2580 – 2585, December 2000.
- [52] F. Wrobel, J.M Palau, M.C Calvet, O. Bersillon, and H. Duarte. Simulation of nucleon-induced nuclear reactions in a simplified sram structure: scaling effects on seu and mbu cross sections. *IEEE Transactions on Nuclear Science*, 48(6):1946 – 1952, December 2001.
- [53] Ming Zhang and Naresh R. Shanbhag. A soft error rate analysis (sera) methodology. In *ICCAD 2004*, November 2004.

- [54] Chong Zhao, Xiaoliang Bai, and Sujit Dey. A scalable soft spot analysis methodology for compound noise effects in nano-meter circuits. In *Proceedings of the 41st annual conference on Design automation*, pages 894–899, June 2004.
- [55] J. Ziegler. Terrestrial cosmic ray intensities. *IBM Journal of Research and Development*, 40:19–39, 1996.
- [56] J. F. Ziegler and W. A. Lanford. Effect of cosmic rays on computer memories. *Science*, 206:776, 1979.

Vita

Vijay S. R. Degalahal, completed his undergraduate in Electrical Engineering with distinction from Jawaharlal Nehru Technological University, Hyderabad, India in 2000. He worked in Satyam Computers from 2000 to 2001 as a software engineer. He joined the PhD program of Pennsylvania State University's Department of Computer Science and Engineering in 2001. He was awarded the College Engineering Dean's fellowship for 3 years from 2001 to 2004. The fellowship is for a maximum of 3 years. He was also an summer intern at Xilinx Research labs for the summer of 2003. Currently, Vijay has joined Mobile Platforms Architecture and Development team of Intel Corporation in Bangalore, India.

Vijay Degalahal, is also a student member of IEEE, IEEE-CS, ACM and ACM's SIGDA. He has also reviewed papers for many conferences and journals including: CASES, DAC, DATE, ISCA, ISCAS, ISLPED, ISVLSI, GLVLSI, VLSI Design, TECS and TVLSI. He also published papers in various conferences and journals.